Dissecting the Purchase-to-Pay Process: An exercise in Process Mining
BPI Challenge 2019

C.J. van Dyk , Himalini Aklecha, Tom Kennes, and Elham Ramezani

KPMG Netherlands
{vanDyk.ChineadJustine, Aklecha.Himalini, Kennes.Tom, Ramezani.Elham}@kpmg.nl
https://home.kpmg/nl/en/home/services/advisory/management-consulting/convert-your-data-into-insight/process-mining.html

Abstract. The 2019 Business Process Intelligence (BPI) Challenge requires participants to analyse and understand the Purchase to Pay (PTP) process of a large multinational company operating in the area of coatings and paints within the Netherlands. The process owner aims to gain business insights through addressing a variety of business questions. We analysed the data using a variety of process mining and analytical tools. This report summarises our understanding of the event log, as well as our analytical approach, and the different techniques and steps undertaken to successfully answer the business questions. Wherever applicable, we also provided additional analysis and discussed the limitations associated with working with such data in the absence of a more comprehensive understanding of the underlying business context.

Keywords: BPI Challenge · Process Mining · Purchase to Pay · Event Logs

1 Introduction

The BPI challenge is an annual process mining competition, which implores participants to utilise various tools, techniques and materials to analyse real life event logs in order to provide assorted business insights. This year, the event log [2] refers to the execution data of a Purchase to Pay (PTP) process, and contains purchase order items with one or more line items which. For each line item, there are roughly four types of flows in the data.

Purchase to pay is one of the back office processes that companies (especially large corporations such as the owner of this year's data set), aim to standardise and streamline. This process is often handled and monitored in shared service centres for large corporations, and typically a large number of financial transactions take place within this process. Consequently, process efficiency and the corresponding compliance of this process is of great importance for the leadership of companies.

Small improvements in such a process can be materialised into millions of euros, given that the company is large enough. External stakeholders such as auditors also investigate this process in great detail as it has a direct impact on the financial statements of companies.

Within this report, we aim to leverage multiple techniques, including process mining, to visualise and reconstruct the corresponding PTP process as it has occurred within the system.

Although encouraged to report on a broader range of aspects, the process owners pose the following questions to be answered explicitly:

1. *Business Question 1*: Is there a collection of process models that can best describe process in the data?
2. *Business Question 2*: Which purchasing instances stand out? where are the deviations observed from the discovered process models? Are there any relation between contextual data of this process (such as invoice value) and detected deviations?
3. *Business Question 3*: What is the throughput time of the invoicing process as part of the overall PTP process?

In the remainder of this paper, we first give a *Management Summary* on the main findings of our analysis in Section 2. Section 3 gives an overview of the steps we took to reach our findings. The data overview is discussed in Section 4, with the assumptions made about the data stated in Section 5. Section 6 refers to *Business Question 1* and the utilisation of process mining to create a collection of descriptive process models which explain the process within the data.

In Section 7, we present descriptive statistics about different purchase orders, invoice values and reworks as well as leveraging process mining to discern any outliers or aberrations which may negatively impact business operations, thereby focusing on *Business Question 2*. Section 8 centralises around *Business Question 3* and the throughput time of the invoicing process. Additionally, the development of a technique used to match multiple invoices within an invoice line item will be presented. Lastly, we conclude the paper in Section 9 discussing the limitations of the data set, recommendations and possible follow up steps for the process owner.

2 Management Summary

In this section we provide an overview of the insights obtained for all the business questions previously mentioned in the Introduction (Section 1).

- For the given PTP event log, there are 251,734 purchase order items. This includes both complete (closed) and incomplete (open) cases. Complete cases are those which have the ending activity ‘Clear Invoice’ or ‘Delete Purchase Order Item’. Complete cases constitute 75% of all purchase order items (189,587 items), with a total net value of 12.6B euros.

- 63% of PO items are product related (PR) with a total net value of 3.49B euros, and 34% of PO items are non-product related (NPR) with a total net value of 8.92B euros.
- The average throughput time of the PTP process (189,587 completed order items) is 81 days.
- The only activities occurring prior to 2018 are 'Create Purchase Requisition Item', 'Vendor Creates Debit Memo' and 'Vendor Created Invoice'.
- The main process deviation observed within PO items that belong to the item category '2-way match', is that approximately 56% of the purchase order items begin with the activity 'Vendor Creates Invoice'. Within a PTP processes, such a starting activity can indicate the occurrence of maverick buying, whereby purchase orders are made without the approval or formal ordering of the purchase order item. Maverick buying constitutes a total of 2.97 million euros for this item category.
- From a risk perspective, it is necessary that a payment block is placed before the payment is completed whenever there is a mismatch in the recorded net values in different events such as the creation of an order and goods and invoice receipt message. This occurs in 20% of order items (49,246 order items). For 17% of them it takes an average of 22 days between the recording of the invoice receipt and the removal of the payment block. Due to this, the average throughput time increases. Blocking is a process that should be minimised as much as possible. Blocking of the purchase order item only occurs in the '3-way match' categories.
- Approximately 7.54% of complete order items (14,295 PO items) have rework activities such as the cancellation of goods and invoice receipts.
- In 77% of the cases, there is a 44 day delay between the payment of the invoice after the invoice receipt message is recorded. (For the invoicing process)
- We observed that the activities 'Record Goods Receipt' and 'Record Invoice Receipt' can occur in any order despite the specifications in the 3-way match item-category.
- In total, the complete cases have an automation rate of 8.47% and a change rate of 11.43%.
- After 'Change' activities occur, there is no 'Change approval for purchase order'. This is peculiar as once creating changes, the purchase order should be reevaluated and approved once more.

3 Analysis Approach and Techniques

Given the overall objective of this report and the corresponding business questions, we applied several process mining techniques [1] in conjunction with various other analytical techniques (such as deploying SQL and python queries) to analyse the event log and derive discerning insights. Through applying such techniques, we aim to dissect the process owners purchase to pay process and provide them with unique results and insights which can be adopted to streamline business processes and potentially even reduce costs.

This was achieved through leveraging process mining solutions from Celonis [4] and open source tools such as ProM [3] and Python [5]. Our overall approach is illustrated in Figure 1. This figure will assist in visualising the different analysis steps taken. These steps are summarised below:

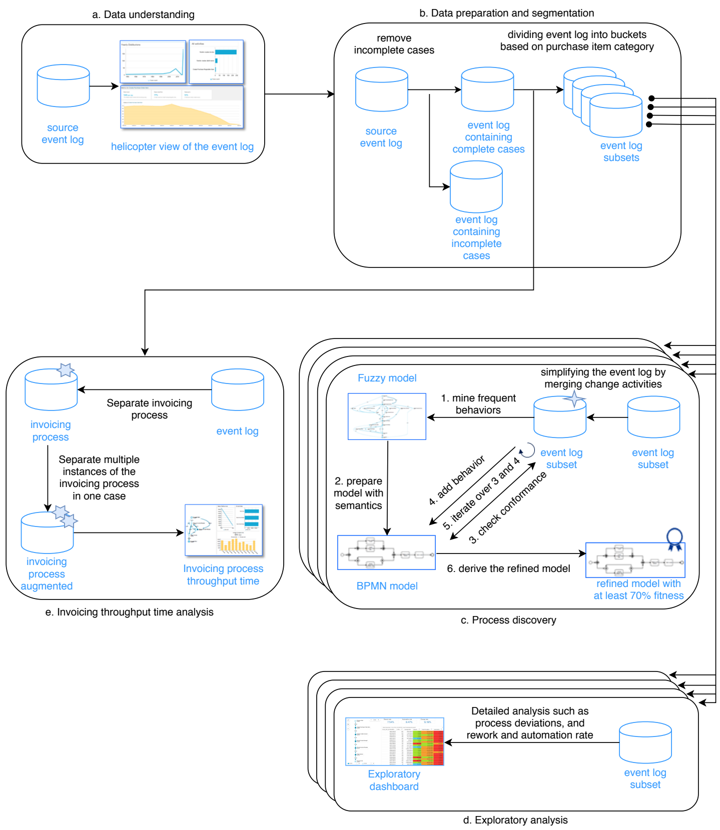


Fig. 1: Overview of the approach and steps taken to answer the business questions

- **Data understanding (Figure 1.a):** In this step we developed a helicopter view of the provided data set. The data and its corresponding characteristics will be discussed in detail in Section 4. Furthermore, we perform an exploratory analysis on the data and outline the assumptions that will be applied throughout the report.

- **Data preparation and segmentation (Figure 1.b)**: First, we prepared the data by removing incomplete instances of the process (i.e. incomplete order items) to ensure our findings are derived from a stable state of the process. Furthermore, we segmented the event log into four subsets based on the purchasing item category (as suggested by process owners). This is done to derive a collection of models that each provide a comprehensive description of the respective subsets process behaviour. Details of this step will be discussed in Section 4.
- **Process discovery (Figure 1.c)**: In order to learn a process model for each event log subset, we mine the most frequent behaviours in the data. From this we obtain four different process models that represent each event log subset. Additional behaviour is gradually added to each model in order to attain a well conforming model with respect to the event log. Furthermore, we analysed the differences between the collection of models discovered. The details of the steps taken are discussed in Section 6.
- **Exploratory Analysis (Figure 1.d)**: Here, the different processes are further evaluated in order to determine whether there are any distinctive purchasing documents which stand out from the rest. In conjunction, different attributes and data sets will also be investigated with the objective of discovering any interesting or discerning features. This step will be discussed in detail in Section 7.
- **Invoicing Process throughput time analysis (Figure1.e)**: We separated the events related to the invoicing sub-process and further analysed the throughput times of this sub-process. Moreover, we also performed a detailed root-cause analysis to detect patterns in the process context that can explain long throughput times in certain cases. Details of this analysis step are found in Section 8.

4 Data Overview

Before delving into profound business insights, this section focuses on introducing the data and its various attributes. In addition, we perform an exploratory analysis and derive assumptions about the provided PTP data which will be applied throughout the rest of the report.

4.1 General Statistics about the Event Log

The event log provided for this challenge, in its entirety, is comprised of 76,349 purchase documents containing 251,734 line items (cases). A combination of the purchasing document number and corresponding line item number (PO item) is considered to be the case identifier. Furthermore, there are a total of 1,595,923 events and 42 unique activities in this event log. This averages to 6.3 events per PO item, with the largest number of activities in a case being 990, and the smallest being 2. The data elements of this PTP event log are shown in Table 1, along with various data attributes that characterise the different purchase order items, hereby known as case level attributes, and their corresponding events, hereby known as event level attributes.

4.2 Data Understanding

From this, we can deduce key attributes that are vital when applying process mining. This includes Case ID, Activity and Timestamp which are all used as event log components. Among the listed attributes in Table 1, there are some attributes that are indicators. Firstly, *GR-Bases Inv. Verified* which is used to indicate if goods receipts based invoicing is required, and is either *true* or *false*. The same holds for *Goods receipt* which indicates whether a so called three-way match rule should be applied for the corresponding purchase order item or not.

In the given data there are important attributes such as *Item Category* that record control information about the purchase order items. They are described as follows:

- **Item Category: 3-way match, invoice before GR**
  For such PO items, the net value at the creation of the purchase order should match the net value at the time the goods are received (GR), furthermore, it should also match against the net value in the invoice (IR). Here, the GR-based IV flag is set to false and the Goods Receipt flags is set to true.
- **Item Category: 3-way match, invoice after GR**
  This follows a similar logic to the aforementioned case, however, the GR-based flag and Goods Receipt flag are both set to true.
- **Item Category: 2-way match**
  For this item category purchase orders do not require a GR in order to process an invoice. Additionally, the GR-based flag and the Goods Receipt flag are both set to false.
- **Item Category: Consignment**
  For these purchase orders, there are no invoices on a PO item level as this is handled entirely in a separate process. Here we see that the GR indicator is set to true, however the GR IV flag is set to false.

This attribute will be used further to divide the data to smaller subsets and subsequently, mine a model for each subset.

4.3 Exploratory Analysis of the Event Log

It is imperative to note the time span of data. Despite the data set pertaining to PO's submitted in 2018, from Figure 2 we observe that the time span of the entire process ranges from 1948 to 2020. When focusing on the specific activity 'Create Purchase Order Item' we see that PO items are only created in 2018 and 2019 (with 99.85% occurring in 2018), as seen in Figure 3. The activities occurring prior to 2018 are 'Create Purchase Requisition Item' and vendor side activities such as 'Vendor Creates Debit Memo' and 'Vendor Creates Invoice'. Albeit abnormal, it is assumed that this indicates that a purchasing process was

Table 1: Data Attributes Overview

Attributes	**Case level (C) Event level (E)**	**Detailed Interpretation**
case concept name	(C)	The Case ID for a given PTP process. It is the combination of the purchasing document number and the item number within each purchasing document. Hereafter known as the PO Item.
event concept name	(E)	It is the activity name and represents the events/steps that were executed in the PTP process.
timestamp	(E)	Timestamp of corresponding events. Represents when the activity of the specific PO Item was executed.
Company	(C)	Anonymised subsidiary from where the purchase request originated. There are 3 distinct anonymised company ID's.
Document type	(C)	High level type of a purchasing document. Categorises the type of document into either standard Purchase Order, framework order or EC purchase order.
Purchasing Document	(C)	Anonymized purchasing document identification. Anonymized due to privacy.
Purchasing document category name	(C)	Name of the category of the purchasing document.
Spend area text	(C)	High level description of the spending area for the purchase order item.
Sub-spend area text	(C)	Additional detailed description of the purchase item.
Vendor	(C)	Anonymised supplier
Item type	(C)	High level type of items purchased within purchase orders. Categorised as either consignment, limit, standard, service, subcontracting or third-party. Item types are defined according to the item templates supplied by SAP.
Item Category	(C)	Categorises the purchase orders into the different matching systems.
Spend classification text	(C)	Explanation of the class of the purchase item, either Non-product related (NPR), Product related (PR) or other.
GR-Based Inv.Verified	(C)	Indicator defining Item category. Flag indicating if GR-based (goods receipt) invoicing is required.
Goods receipt	(C)	Indicator defining Item category. A flag indicating if 3-way matching is required (true/false).
User	(E)	Indicates the type of user executing an activity (either 'batch', 'user' or 'none').
Cumulative net worth (EUR)	(E)	Cumulative net worth of the purchase order.

instigated many years ago (for example a contract being signed 10 years ago but still being used as the basis for purchases in 2018 and 2019). Further, the only activity pertaining to 2020 is ‘Vendor Creates Invoice’. This might relate to a contract that has already been set in motion, or it could simply be a data error.

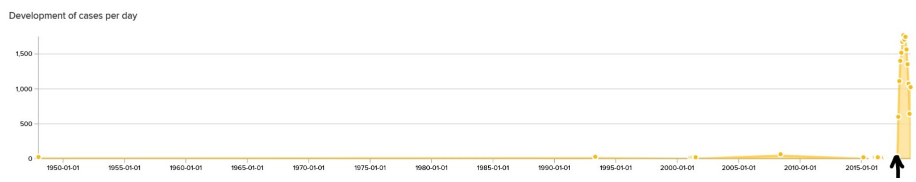


Fig. 2: Graph showing the time frame of events pertaining to the given PTP log. As observed, very few events occur prior to 2018 (marked with a pointer), and majority of the events occur in 2018 and 2019.

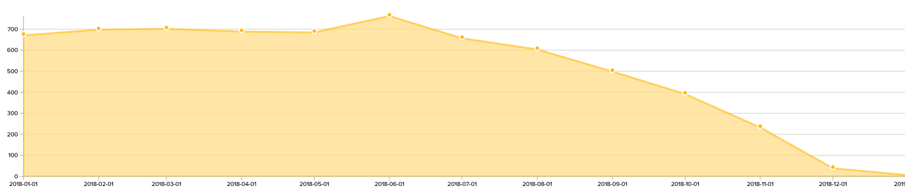


Fig. 3: Distribution of occurrences of activity ‘Create Purchase Order Item’ over time. All orders were created after 2018-01-01.

As previously mentioned, the data is further divided into four different categories where these categories correspond to the different matching systems, and hence, represent different process flows within the data. When visualising each process flow it is evident that they do not contain all activities, with some activities being unique to one event log subset (for example, ‘Block Purchase Order Item’ is unique to 3-way match, invoice before GR). An overall summary of the different event log subsets and their statistics can be found in Table 2. As can be seen in Table 2, 95% of the order items fall in the item category 3-way match, invoice before GR. However, the largest bucket with respect to monetary value (more than 8B euro) falls in the category 3-way match, invoice after GR.

4.4 User Analysis

Additionally, despite the different user types, there is no distinct activity that one user (or user type) performs thus voiding the notion that a particular user corresponds to a certain position (such as a manager who would typically be in control of approval activities). This was discovered through applying the ‘Social

Table 2: Item-Category Overview (for complete order items)

Item Category	**#PO Items**	**#Users**	**#Activities**	**Total Net Value (Euros)**	**% of cases**
2-way match	170	15	9	8.87M	<0.1%
3-way match, invoice before GR	179,429	557	39	3.96B	95%
3-way match, invoice after GR	9,578	245	36	8.63B	5%
Consignment	397	61	12	0	0.2%

Network Miner' in ProM [3]. However, there were some interesting observations made, such as:

- 'Change Approval for Purchase Order' is executed only by manual users.
- Item category 'Consignment' has no 'None' users.
- 'None' users only execute the following activities: 'Vendor creates invoice', 'Clear Invoice', 'Vendor creates debit memo' and 'Record service entry sheet'.
- 'Create Purchase Order' is executed by both Batch and Manual users.

5 Assumptions About the Data

In order to obtain a rough overview of possible subsets available in the data, we initially tried to apply several machine learning algorithms with respect to clustering and classification. However, the data did not seem to contain a combination of features that allowed for meaningful clustering. Therefore, there was no clear indication on how to sub-divide the event-log based on any other attribute asides from the item-category. The provided log file was then imported into ProM whereby the complete event-log was segmented into four smaller event-logs, based on the item category, namely '2-way match', '3-way match, invoice before GR', '3-way match, invoice after GR', and 'Consignment'. To gain an idea of each of their process flows, the event-logs were uploaded into Celonis. We used this tool to visualise individual process flows and identify (key) characteristics such that a well defined process model could be created.

As mentioned previously, we made certain assumptions about the data that will be adhered to throughout the report. The main assumptions and their reasons are stated as follows:

1. **SRM sub-process:** Activities relating to 'SRM' (Supplier Relationship Management) appeared to have their own distinct sub process as shown in Figure 6. In addition, SRM activities only occur in a total of 1,440 cases, given this, it was concluded that when constructing the relevant process models, any activity related to SRM would be excluded.

2. **Case status:** By analysing the process graph, we observe cases that do not end with typical PTP ending activities, and are therefore considered incomplete. To ensure for a full analysis, we make sure that the data is complete, i.e., have a **valid ending**. Thus by applying domain knowledge about PTP processes, we deemed that the activities ‘Delete Purchase Order Item’, and ‘Clear invoice’ were valid ending activities. Thereafter, we only considered PO items adhering to this as complete cases. This reduced the number of purchase order items from 251,734 to 189,587 (75% of the original data).

   Before moving on to further analysis, we provide insights about the incomplete cases. We have that all incomplete cases (25%) are only instigated in the years 2001 (5 PO items),2017 (32 PO items), 2018 (61,957 PO items), 2019 (15,384 PO items).

   Figure 4 shows the last activities of all open PO Items. This allows the business to drill down to the most frequent open activities and make a decision of whether the item is waiting due to inefficiencies in the business or due to the vendor. Furthermore, Figure 5 shows the PO items open since past few months. This is the elapsed time and is calculated as the number of months (rounded values) between the execution date of the last event, for an open item, and today. For example, 9,947 PO items are still open past 4 months. A point to note here is that the data given contains events that have a timestamp of 2020. We assume that this is an outlier in our data and to avoid the graph showing a negative elapsed time, we have filtered out these items from the graph.

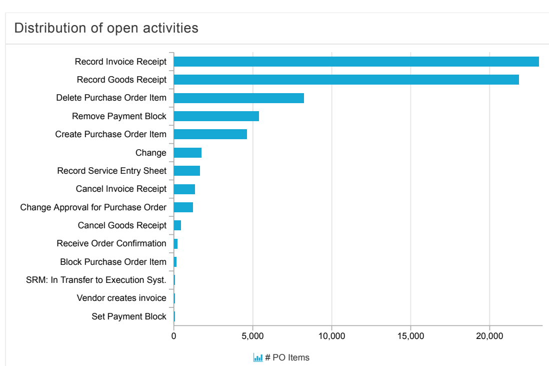


Fig. 4: Last executed activities for incomplete PO Items

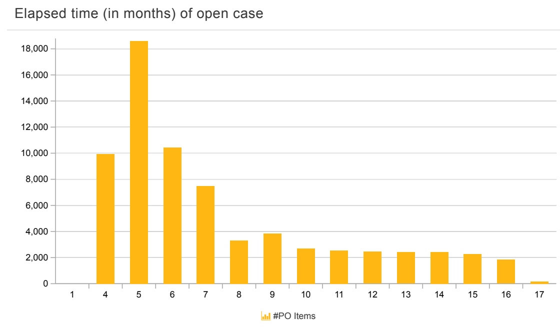


Fig. 5: Incomplete PO items and the elapsed time (in months) since they were last open

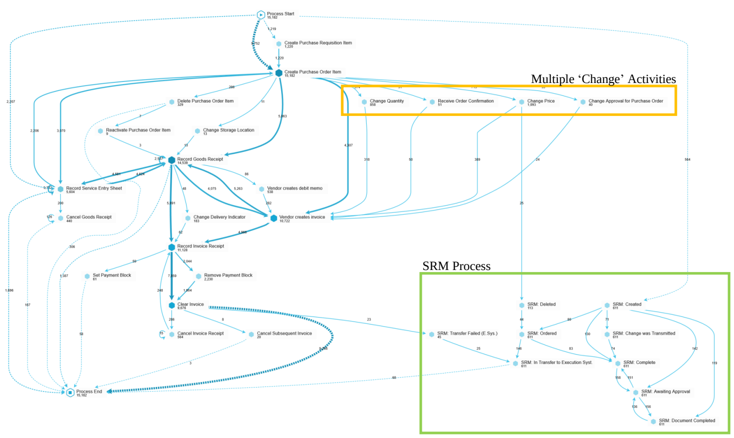


Fig. 6: Process discovery graph. Orange box shows the multiple 'Change' activities in parallel and the green box represents the SRM sub-process.

6 Creating Descriptive Process Models

The main aim of this section is to create a collection of process models that collectively describe the flow of purchase order items through the organisation's PTP process (Figure 1c). Formally, these process models are graphically represented in Business Process Model and Notation (BPMN). In this section we will determine the different process models using the mining techniques of Celonis, and incrementally edit the model using a BPMN tool [6]. The creation of a BPMN model can be challenging, especially given the spaghetti-like nature of the PTP process. Hence, we performed a data pre-processing step when generating the different process models. We identified that there were a myriad of 'Change' activities relating to different changes in the purchase order item, as seen in Figure 7.

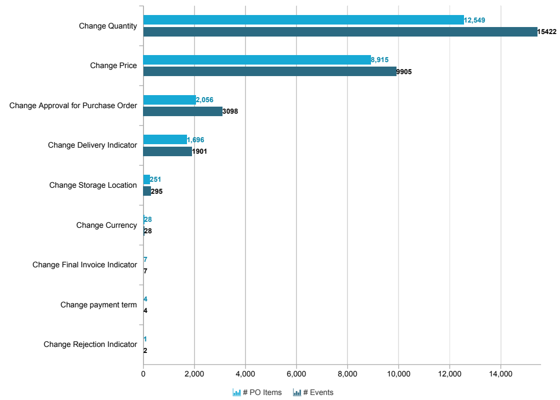


Fig. 7: Different 'Change' Activity with Case Frequency and Activities Frequency

Furthermore, as seen in the orange box in Figure 6, these activities often occurred parallel to each other, suggesting that they did not always have a well defined 'follows rule'. This results in a so called 'flower model' which does not help in understanding the actual flow of activities, as it implies that activities can occur in any order. Thus, in order to streamline the model a new activity called 'Change' was created which combined all the change activities except for 'Change Approval for Purchase Order', see Figure 8. This was due to the fact that this activity does not relate directly to changes in the PO, instead it appears to be a change within the process control.

Change Activities
Change Currency
Change Price
Change Payment Term
Change Quantity
Change Storage Location
Change Delivery Indicator
Change Rejection Indicator
Change Final Invoice Indicator

New Activity
Change

Fig. 8: Creating of new activity: 'Change'

A procedural approach was adopted in order to create each process model. The steps are as follows:

1. Divide the event log into subsets based on the four different item categories.
2. For each subset obtained from Step 1, select the most common variants from the as-is process and mine the same data set to obtain a BPMN model.
3. Based on the complexity of the previously obtained BPMN model, we further removed or incrementally added additional activities and edges to the model increasing its complexity. This ensured that we captured a higher number

of variants from the as-is process whilst still retaining the simplicity of the described model.

4. Perform conformance checking to ensure that the model from step 3 accurately represents and describes the population of data.

For each process model, the level of conformance with respect to their corresponding data population (percentage of cases which perfectly conform/align with the designed BPMN models) is stated below in Table 3.

Table 3: Process Model Conformance per Item Category

Data Model	Conformance Checking
2-way match	90 %
3-way match, invoice before GR	92 %
3-way match, invoice after GR	74 %
Consignment	92 %

Process Model: 2-way match

By following the steps stated above for all PO items within the '2-way match' category, we obtained the process model depicted in Figure 9. This category constitutes less than 0.1% of the total data and has a total net value of 8.8 million euros (given the valid ending assumption). As depicted by the parallel gate in the model, the activities 'Vendor creates invoice', 'Create Purchase Order Item' and 'Change Approval for Purchase Order' can occur in any order as the first activity, with 'Vendor creates invoice' occurring first the most (40% of the time). This is corroborated by the most frequent process path (Figure 10). However, such behaviour, within the PTP process, is considered as a maverick buying. This is because the vendor has created an invoice due to a purchase order made, however, without this order being logged. We have that 8% of the PO items start with 'Change Approval for Purchase Order' and another 7% starting with 'Create Purchase Order Item'. It is interesting to note the loop introduced for the activity 'Change Approval for Purchase Order'. This is due to the fact that it is repeated approximately 32% of the time, average 2.1 times per case. This high frequency of change may be of concern which may have follow on effects such as higher throughput times and hence increased costs. Lastly, we see that 90% of the PO items end with the process fragment 'Record Invoice Receipt' and 'Clear Invoice'. This model conforms with the selected population of data with a score of 90% (Table 3).

Process Model: 3-way match, invoice before GR

The process flow of PO items categorised as '3-way match, invoice before GR' is displayed in the BPMN model in Figure 11. With a conformance of 92% against

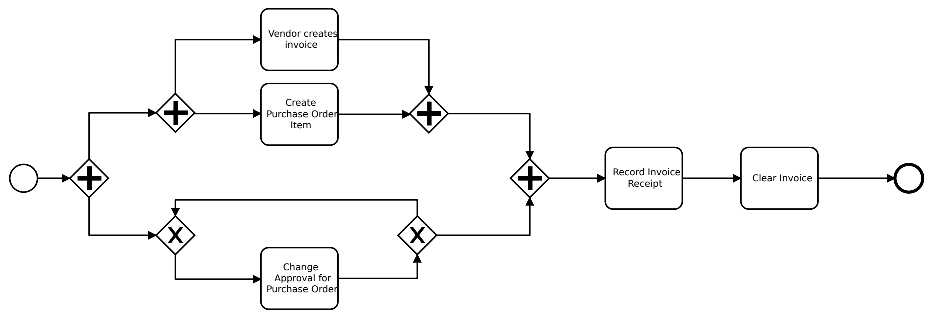


Fig. 9: BPMN model that describes the PO Items in the '2-way match' item category.

Fig. 10: Most frequent process path for PO Items pertaining to '2-way match' item category.

its data population, this bucket has a total net value of 3.96 billion euros and comprises the largest section of the of original data (95%). Here, 13% of the PO items begin with the activity 'Create Purchase Requisition Item' and in 99% of the cases are directly followed by the creation of the purchase order item. However, 85% of the PO items commence with the activity 'Create Purchase Order Item' hence the intuition behind allowing for both starting activities.

In contradiction to the namesake of this category, the recording of the invoice receipt occurs after the recording of the goods receipt, after which the corresponding payment is made (see most frequent process path, Figure 12). Moreover, there are multiple cases in which the two activities, corresponding to the GR and IR, are executed as a result of the activity 'Vendor create invoice'. This behaviour is captured by the parallel gate. The model allows for both of the 'valid' ending activities to occur, that is, 'Clear invoice' (96% of cases) which indicated the payment of the invoice, and the deletion of the PO item 'Delete Purchase Order Item' (4% of cases). It is important to note that the deletion of the purchase order item occurring directly after its creation is highly inefficient and may lead to complexities and increased costs.

Other activities such as 'Remove payment block' are added to the model, and are executed when there is a violation in the 3-way-match, such that no incorrect payments are completed (occurs in 26% of the cases for this item category). Lastly, activity 'Change' is added to allow for multiple change events. Unlike the '2-way match' process model, this model does not contain the activity 'Change Approval for Purchase Order' as it only occurs in 0.2% of the PO items. This means that there are a less changes within the process control allowing for a more streamline model. Additionally, 'Change' activities only occur in 2% of the cases. In general, it appears that '3-way match, invoice before GR' is most compliant with a PTP process compared to the other process models - that is it starts with 'Create Purchase Order Item' and ends with 'Clear Invoice'.

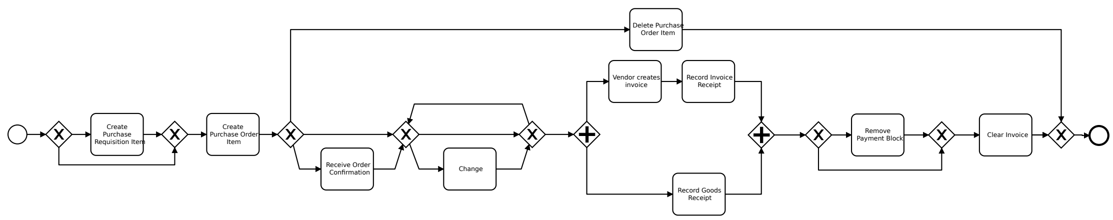


Fig. 11: BPMN model that describes the PO Items in the '3-way match, invoice before GR' item category

Fig. 12: Most frequent process for PO Items pertaining to '3-way match, invoice before GR' item category

Process Model: 3-way match, invoice after GR

The process model describing the process flow of PO items categorised under '3-way match, invoice after GR' is visualised in Figure 13, and has a conformance value of 74% (3). Representing 4% of the total data, it has a net value of 9.37 billion euros. Similar to the process model for '3-way match, invoice before GR', this model allows for items to begin with the activity 'Create Purchase Requisition Item' or 'Create Purchase Order Item'. In addition, given the multiple 'Change' activities, we introduced the a loop for the grouped 'Changes' activity (as explained in 3) This now allows for multiple changes in the PO item such as change in price, quantity and currency without restricting the different change activities to occur in a particular order.

Interestingly, this process model has the same frequently executed process path (Figure 14) as the '3-way match, invoice before GR' (12). Again, we see that 'Record Goods Receipt' directly followed by 'Record Invoice Receipt', however, in this instance the ordering corresponds to the item category name. The process model does allow for multiple record goods receipts (on average 4.4 times per case) and record invoice receipts (on average 1.5 times per case) to be repeated several times, as expressed by the loop in the process model. For this category, an interesting observation was made where activities corresponding to GR and IR are repeated in 1% of the cases. Although this is a small proportion of items, we introduce this behaviour into the model since it does not lead to complex loops and is considered to be a normal behaviour for a PTP process. Additionally, we see that the activities 'Vendor creates invoice' (in all PO items) and 'Record Service Entry Sheet' (in 24% of PO items) is also executed parallel to the GR and IR.

It is important to note the loop created for the activity 'Record Service Entry Sheet', which is unique to this item category and occurs on average 17.2 times per case. This is peculiar behaviour and may lead to additional costs, and therefore should be examined by the process owners. Further, the deletion of the purchase order item occurs both directly after the item creation, or after applying multiple changes. Once again this is inefficient and may lead to increased costs. Moreover, towards the end of the process model we observe that the invoice receipt is cancelled and then cleared, implying that it is paid. Although only occurring in 2% of the cases, this may pose a compliance issue. Once again, the process model ends with the 'valid' ending activities, 'Clear invoice' (97% of the PO items) and 'Delete Purchase Order Item' (3% of the PO items). Similar to the previous explanation, activity 'Remove payment block' is also added in the model.

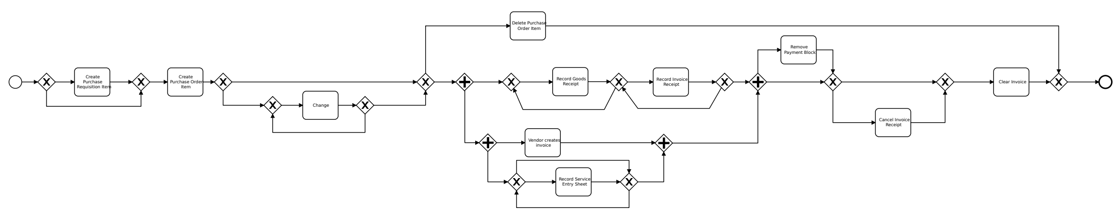


Fig. 13: BPMN model that describes PO Items in the '3-way match, invoice after GR' item category

Fig. 14: Most frequent process for PO Items pertaining to '3-way match, invoice after GR' category

Process Model: Consignment

The item category 'Consignment' has a very interesting process flow. Firstly, it has a total net value of 0 euros and constitutes 0.2% of the total data (for valid endings). Moreover, all of the PO items end with the activity ‘Delete Purchase Order Item’, meaning that they are deleted (as seen in Figure 16. In terms of cost savings and efficiency, this is highly ineffective. The process flow of this item category is displayed in Figure 15, and has a conformance rate of 92% (3). Within a consignment process, the vendor allows for goods to be stored in the organisation’s warehouse, and are then either consumed by the company (and hence deleted), or are given back to the vendor and as a result deleted from the item category.

In a similar fashion to both of the ‘3-way’ item categories. this process model also begins with either the creation of the purchase requisition item, or the purchase order item. Similarly, ‘Change’ activities are executed one or more number of times until the goods are received. Additional recording activities such as ‘Receive Order Confirmation’ and ‘Update Order Confirmation’ have been included in the model. For such a process, the purchasing company does not pay any amount to the vendor until the goods are used and a separate payment process is undertaken thereafter. Thus we do not see the activity ‘Record Invoice Receipt’ in the process model.

The model allows for additional activities such as ‘Cancel Goods Receipt’ (1.5% of PO items) and ‘Reactivate Purchase Order Item’, which occurs after the deletion of the purchase order item. This is rather interesting, as although it only occurs in 1% of the cases, why would a PO item be reactivated only to be immediately deleted again. Further, the activity ’Update Order Confirmation’ is unique to this item category.

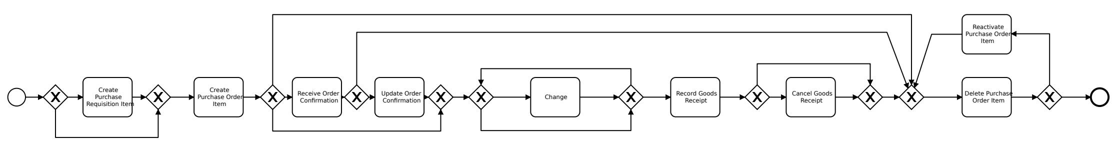


Fig. 15: BPMN model that describes the PO Items in the ‘Consignment’ item category

Fig. 16: Most frequent process for PO Items pertaining to ‘Consignment’ category

Collectively, these four separate process models provide an in depth analysis as to how the data is processed. Furthermore, process variants that deviated largely from the main process flow and cover a small proportion of cases were not included in the final BPMN models. This is due to the fact that such cases are outliers and including them in the process model would only serve to over-fit the model. The accuracy of such models are corroborated by their high conformance results. It is imperative to employ multiple models in order to fully grasp the behaviour of the data.

7 Further Analysis, Deviations and Anomalies

Within this section, the PTP event log is further analysed in order to determine whether there are any purchasing documents which are highlighted against the rest. For this, different attributes and data sets (each of four data sets, segregated by item category, as well as the entire event log) will be investigated with the objective of discovering any interesting or discerning features. This section relates to Figure 1.d in the data approach overview.

In addition to throughput time bottlenecks, rework activities also contribute to the inefficiencies within a process. Consequentially, this can lead to costs increases, additional workloads, and delays in completion time. Such activities are executed when the previous activities were not completed correctly or when there is data missing prior to the rework. Thus, in the PTP process we consider the following activities as rework activities: 'Cancel Goods Receipt', 'Cancel Invoice Receipt', 'Cancel Subsequent Invoice', 'Delete Purchase Order Item', 'Maintain GR/IR', and 'Update Order Confirmation'.

We analyse the rework rate of the process by drilling down on various attributes. Again, in order to simplify the analysis we perform rework analysis on each of the item categories separately. Here we define rework rate as the ration between the number of order items containing a rework activity compared to the total order items. Apart from rework, we also introduce the concept of automation. This is determined as the percentage of activities executed by digital agents such as batch users. Automating activities within processes helps to reduce costs by decreasing the amount of rework often caused by human error.

Table 4: Deeper analysis of the given PTP process per Item Category

Item Category	**#PO Items**	**Average Throughput Time (days)**	**Rework Rate**	**Change Rate**	**Automation Rate**
2-way match	170	76	4.12%	1.76%	0.29%
3-way match, invoice before GR	179,429	80	7.29%	6.64%	11.02%
3-way match, invoice after GR	9,578	90	8.73%	22.55%	19.44%
Consignment	397	18	100%	8.06%	2.27%

From Table 4, we conclude that item category '3-way match, invoice after GR' has the largest bottleneck in terms of rework rate, change rate, and automation rate.

In order to achieve a general overview of the process flow, we observed the as-is discovered process graph and dive deeper into the analysis we drill down on various categories in combination with other process attributes to highlight the main inefficiencies:

- **Process deviations**
  - The most common variant is 'Create Purchase Order Item' which occurs as the starting activity for 84% of the cases, where as 'Create Purchase Requisition Item' is a starting activity for only 12% of the cases.
  - For PO items in the '3-way match, invoice before GR' category, the goods receipt message is recorded prior to the invoice receipt message in approximately 88% of the items.
  - For item category '2-way match' the most common variant starts with activity ' Vendor creates invoice' followed by the creation of the purchase order item. This covers 45% of the PO items pertaining to this category. Such a process flow signifies that a purchase was made before a formal approval or order was made - known as maverick buying. This deviation is also seen in other process variants of this category as well as in other item categories.
  - In both '3-way match' categories, 3% of the purchase order items are created and the subsequently deleted.
- **Throughput times, rework rate and other inefficiencies**
  - When observing the PTP process in its entirety, we see that the rework and automation rates are rather low (7.53% and 8.47% respectively). item[−] 'Workforce Services' has the longest throughput time (275 days) whereby the sub spending area 'HR Services' and 'Third Party Labour' have significantly high average throughput times
  - For item category '2-way match', the largest throughout time is taken from 'Change Approval for Purchase Order' to 'Vendor Creates Invoice', i.e., average of 71 days. This is effected in 35% of the cases. This means that there is a long waiting time for the company from the vendor's side until the invoice has been created.
  - In item category '3-way match, invoice before GR', 'Record Invoice Receipt' followed by 'Clear Invoice' has a throughput time of 45 days, affecting 69% of cases. This fragment is a part of the invoicing process which has been discussed in Section 8. However, we would like to point that long throughput time for an invoice payment cannot be considered without comparing it to the vendor's/contract's payment terms. For example, if the payment term for a vendor is 60 days, then all invoices corresponding to that vendor should be paid within 60 days of receiving the invoice. Paying the invoice in lesser number of days can cause the company to reduce the cash currently in the company. On the other hand, paying the invoice late is also detrimental to the company (eg. loosing cash discounts) and their relationship with the vendor.

  - When drilling down on purchasing documents, the largest document (in terms of number of line items) '4507075965' has a rework rate of 14.29%. Upon further analysis, we observe that these purchasing documents relates to the spend area 'Real estate' and 'vendorID_1687'. Furthermore we observe that there is an automation rate of 0% for these line items. Through increasing the automation rate, rework could be reduced.
  - Purchase Document 4507006057, has 84 purchase order items which consist only of the two activities 'Create Purchase Order Item' followed by 'Delete Purchase Order Item', where each purchase order item has the same timestamp for each activity. Therefore, it has a 100% rework rate.
- **Additional observations**
  - For item category '2-way match' the activity 'Change Approval of Purchase Order' is executed 346 times in 166 PO items. This activity requires further investigation as it results in large throughput times. Furthermore, 32% of the cases start with this activity. Lastly, this activity is repeated multiple times (up to 12 times) in many PO items.
  - Spend area 'Chemicals & Intermediates' with only one sub spend area 'Catalysts' only has one purchase order item.
  - The Spending area of 'Logistics' can have multiple 'Record goods receipts'
  - The only document type that supports the SRM process is 'EC Purchase Order'.

8 Analysing Throughput Times of the Invoicing Process

A key area of interest for the process-owners is the invoicing process. This refers to the relationship between the *Goods Receipt, Invoice Receipt* and *Invoice Payment* (activity *Clear Invoice*). Thus in this section we provide a comprehensive analysis of the organisation's invoicing process. Furthermore, we also drill down on different factors and attributes which may influence the throughput times such as vendors and document types. This section refers to Figure 1.e in the analysis approach overview.

Within the given event log, activities corresponding to the invoice process occur only in the years 2018 and 2019, thus pertaining exclusively to the years in which the PO items were created. In order to analyse the invoicing process directly, a truncated data set including only these activities was created. However, within the data it is apparent that each line item may have multiple events and consequentially multiple goods receipts and multiple invoices within the line item. In essence, there can be many goods receipt messaged and corresponding invoices which are subsequently paid. Perhaps the simplest example to consider is paying rent, say yearly rent is \$1200. A purchasing document could have one item for paying the rent, but a total of 12 goods receipt messages with cleared invoices equalling $1/12^{th}$ (\$100) of the total amount.

Overall, for each line item, the amounts of the line item, the goods receipt messages (if applicable) and the invoices have to match for the process to be compliant. This is rather prevalent within logistical services where there may even be numerous goods receipts for one line item. Therefore, in order to combat this, a technique to match events within a line item (and determine how multiple goods receipts and invoices are related) is required.

8.1 Matching Events Within A Line Item

To match events within a line item, connections and correlations between event level attributes were first explored. However, within a line item, there were no distinctive features that allowed for a pattern or relationship to be identified. It was found that users executing the activity 'Record Invoice Receipt' and 'Record Goods Receipt' had no clear relationship, in terms of handover of work. Hence, given the nature of an invoicing process, irregardless as to ordering of the activities 'Record Invoice Receipt' or 'Record Goods Receipt', 'Clear Invoice' marks the end of the process. Therefore, through utilising this we can distinguish and group events with the line item, as depicted in Figure 17.

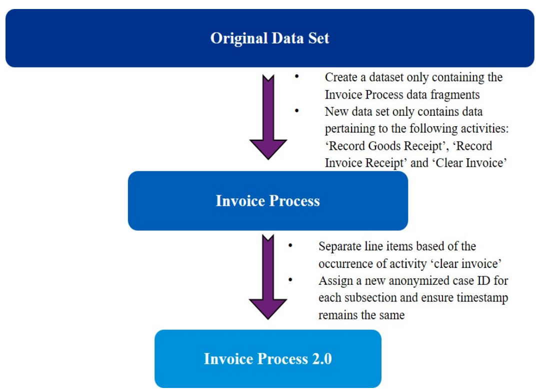


Fig. 17: Creation of new 'Invoice Process' data set

For each case, the invoicing process was considered complete if 'Clear Invoice' occurred after instances of both 'Record Invoice Receipt' and 'Record Goods Receipt' irrespective of their ordering. This would signal the start of a new sub-case, see Figure 18.

8.2 Insights

For the invoice process as a whole, the average throughput time from process start to process end is 64 days, with some being processed in less than 28 days and others taking up to one year. (Figure 19)

From the process graph (Figure 19, left), we observe that the majority of cases follow the path 'Record Goods Receipt', 'Record Invoice Receipt' then

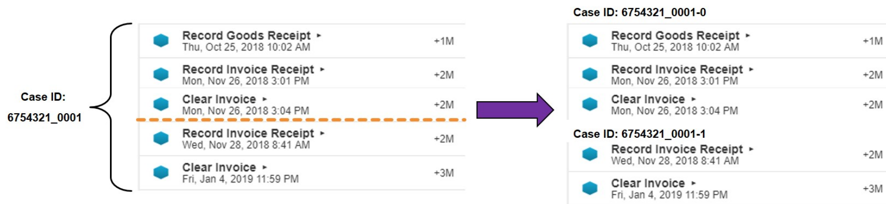


Fig. 18: Matching Events Within A Line Item

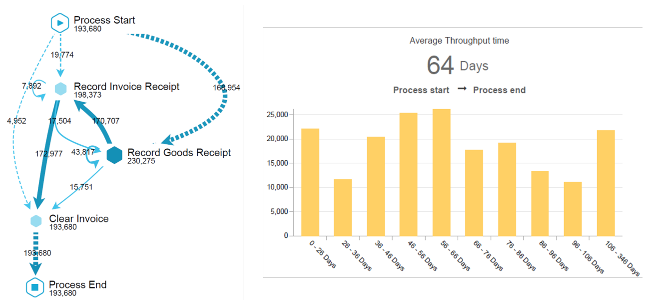


Fig. 19: Throughput Time of Invoicing Process

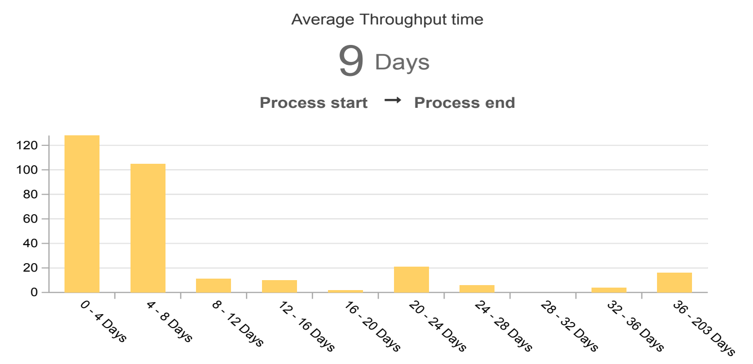
'Clear Invoice', however there are many repetitions for the activity 'Record Goods Receipt'. This is most likely a result of the logistical services which, as previously stated, can have multiple GR messages for one line item. We analysed the invoicing process for different item categories separately (see Figure 20). Here we see that item-category 'Consignment' in not included. This is due to the fact that it does not include the activity 'Record Goods Receipt'.

Both '3-way match' categories have significantly higher total throughput time (TPT) than '2-way match', which has the shortest average throughput time with majority of the invoices being completed within 8 days (Figure 20a) - this is because of the simplicity of the process model. The higher throughput times in '3-way match, invoice before GR' (Figure 20b) may result from the fact that the invoice does not actually occur before the goods receipt as evident in the process model created in Section 6.

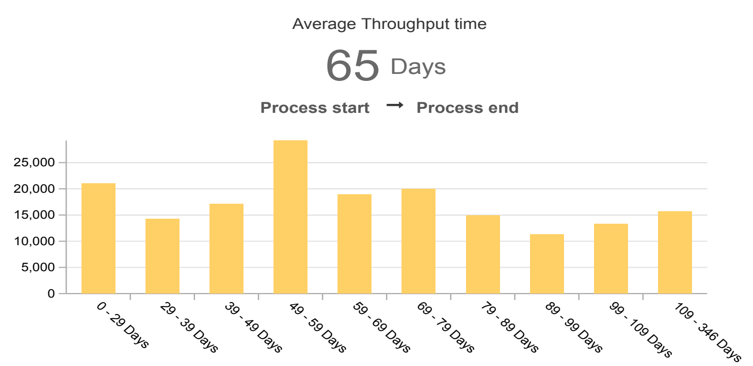
In order to gain a more comprehensive understanding of the total average throughput times (TPT), and what contributes to extended throughput times, we drill down on different attributes for complete cases.

- The document type 'Standard PO' has the longest average throughput time (60 days), and only occurs in both 3-way matching systems. This is due to a several reasons, firstly the 3-way matching systems contain all three invoicing process activities, therefore requiring additional time to process. Additionally, the vendors and spending areas associated with standard PO's have the longest throughput times.

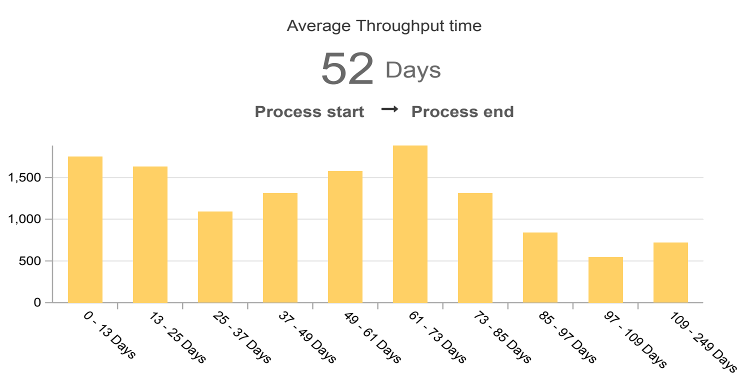
Fig. 20: Throughput Times of Item Categories



(a) Throughput Times 2-way match



(b) Throughput Times 3-way match, invoice before GR



(c) Throughput Times 3-way match, invoice after before GR

- The two vendors which have the longest throughput times (vendorID_0402, TPT:332 days and vendor_ID1413, TPT:210 days) relate each to only one case suggesting that perhaps the case is unique to the vendor, or one could change vendors to ensure a shorter TPT.
- companyID_0001 has the longest TPT (100 days), and only one purchasing document with a net value of $15,276. This raises concerns as to why this particular company is being used, and how come it only has one vendor.
- In contrast, companyID_0003 has the shortest average throughput time whilst utilising vendors with the shortest TPT.
- Top six most commonly used vendors all have an average throughput time greater than 66 days.

9 Limitations and Conclusions

In this section we will outline the limitations we faced for the given dataset. We will conclude our findings and give some suggestions for further analyses.

9.1 Limitations of the Data set

Overall, the process flow of the data set was rather complex and chaotic with approximately 25% of the cases failing to meet the valid ending requirements, not to mention the extensive time span of the data. Given the amount of incomplete (open) cases, and purchase order items pertaining to 50 years ago, it is recommended that the data be refined. Additionally, there are a lot of cases which are created, undergo different activities and are then deleted. This in turn creates large amounts of rework and increases throughput times, negatively impacting process efficiency and possibly even increasing expenses.

It was rather difficult to turn such observations into business insights, given that no background information was provided. For example, within the invoicing process there are no invoicing numbers meaning that it was unable link the different invoice receipts to the clearing of the invoice to form clear invoicing activities. Moreover, the invoicing process did not include any information about the payment term meaning that there was no way to determine late or early payments. Since we do not know the origins of the data, or what the various attributes extensively mean and to what extent they correlate to one another, we are unable to quantify the impact of process deviations and provide direct business advice.

9.2 Conclusion

Within this report, we have leveraged process mining tools such as ProM and Celonis in order to create four succinct models that collectively provide an accurate description of the PTP process at hand. Each model had a conformance rating of at least 74% (with the highest being 92%) whilst still maintaining simplicity. These models were subsequently used as a basis to perform further in depth analysis of the PTP process. We used the models to answer the subsequent business question, in particular relating to the throughput times of the invoicing process. In order to achieve this, an effective method used to match multiple events withing a line item was presented. These throughput times were thoroughly investigated, with additional insights provided in order to ensure a comprehensive analysis and understanding of different contributing factors was achieved. Finally, we highlighted any anomalies within the process, including bottlenecks, rework, automation rates and other general observations.

9.3 Future Work

Going forward, we would continue our work by conducting further predictive and prescriptive analysis and, through the combination of play-in, play-out and replay analysis attempt to construct an overarching process model which could explain the process in its entirety. Or conversely, determine alternative ways to segregate the event log to create more process models. In addition, given the high levels of rework in conjunction with relatively low automation rates, one area we would like to investigate the possibility, and business cases, of further automating the process. This can be achieved through identifying process fragments which are more suitable for automation and will lead to higher gain if they are automated.

Acknowledgements

We thank Ben Leijdekkers for his substantial contribution in preparing the business value of the insights found in this work.

Dashboard

Here you can view the dashboard (in Celnois [4]) created to derive insights and solve the business questions presented in this report:
`https://www.dropbox.com/s/lyihiw98j8uf5vr/BPI2019_Dasboard%20%28password%20-%20KPMG_NL_bpi2019%29.CTP?dl=0`

References

1. Wil M. P. van der Aalst: Process Mining: Discovery, Conformance, and Enhancement of Business Processes . 1st edn. Springer, Netherlands (2011)
2. DataSet Homepage, https://data.4tu.nl/repository/uuid:d06aff4b-79f0-45e6-8ec8-e19730c248f1.
3. ProM Homepage, http://www.promtools.org/doku.php.
4. Celonis Homepage, https://www.celonis.com/?gclid=EAIaIQobChMIqPzitdSO4gIVz-F3Ch1xyQtMEAAYASAAEgLqs_D_BwE.
5. Python Homepage, https://www.python.org/.
6. BPMN model designing tool, http://bpmn.io/.

Business Process Intelligence Challenge 2019 - Hierarchical process deviation analysis using evolutionary model discovery

Max Adaloudis[1], Veronika Cucorova[1], Koen van der Leij[1], and Koen Minartz[1]

Eindhoven University of Technology, Eindhoven, The Netherlands

Abstract. Process Mining nowadays is an extensively used field in process management to extract insights from event logs. The Business Process Intelligence Challenge of 2019 poses a problem and provides a real-life event log from a large multinational company in the area of coatings and paints. This paper uses process mining to find multiple measures of deviations within four purchase order types. Deviations are measured with respect to process flow, invoice matching and rework of purchase orders. To describe the four processes, we used evolutionary model discovery techniques and multi-level data analysis. For deviations regarding the process flow, we consider the three hierarchical levels event, case, and purchase order to analyze structural performance changes within the event logs. Our findings suggest that multivariate data analysis combining trace attributes with event data used to discover process deviations can provide insights to improve current process flow.

Keywords: Business Process · Intelligence · BPIC2019 · Process Mining · Evolutionary Model Discovery · Hierarchical Process Mining · Process Deviation Analysis · ProM

1 Introduction

A BPI challenge uses real life event-log and challenges participants to solve the process owner's questions or find unique insights into the process. There is no restriction on tools used by the participants. The participants will present their findings to a jury in the form of a paper, after which the winner is selected for each of the categories. The Business Process Intelligence Challenge 2019 (BPIC 2019) consists of two categories, student and non-student, where the student category is expected to answer one or multiple of the process owner's questions in depth and the non-student category are expected to focus on a broader range of aspects.

1.1 List of definitions

Purchase document	Also called **Purchase order**, represents a document that is issued to order one or more purchase items.
Purchase item	Also called **Line item**, represents one case (also named trace) within an event log.
Event	An activity occurring within a process logged within an event log.
Event log	Also called **Dataset**, represents the database containing process data composed of traces containing events.
ProM	Open source software platform allowing various process mining techniques.
Disco	Process mining tool from Fluxicon.
Petri net	Modelling language describing distributed systems using nodes called places to describe pre-conditions and post-conditions and transitions between places to describe activities.
Precision	Measure of how much of the behaviour present in the log is covered by the model. The higher value, the better, low value implies the model does not explain the data well.
Fitness	Measure of how much of the behaviour allowed by the model is actually present in the log. The higher, the better, low value implies that the model is too general and not many insights can be derived.

1.2 Data BPIC 2019

The 2019 BPIC uses real life data collected from a large multinational in the Netherlands. From now on this multinational will be referred to as the process owner. The challenge description specifies four process flows, into which the items can be partitioned according to the way their order is handled and which documents are required. A purchase document contains one or multiple line items. Moreover, multiple goods receipt messages and corresponding invoices for each item can exist. The process is defined to be compliant when for each line item the amounts of the line item, goods receipt messages and corresponding invoices are equal. The four flows for line items are following:3-way matching, invoice after goods receipt, 3-way matching, invoice before goods receipt, 2-way matching (no goods receipt needed), Consignment.

The provided dataset is anonymized and consists of over 1,5 million events for purchase orders submitted in 2018. This data contains the process of handling the purchase orders and information about the item categories, vendors and documents.

- The resources are split between fully anonymized batch users and normal users. The batch users are automated processes executed by different systems. The normal users refer to human actors in the process.
- The values of each event are fully anonymized from the original data using a linear translation respecting 0, i.e. addition of multiple invoices for a single item should still lead to the original item worth (although there may be small rounding errors for numerical reasons).
- Company, vendor, system and document names and IDs are again fully anonymized in a consistent way throughout the log. The process owner has the anonymization key, so any result can be translated by them to business insights about real customers and real purchase documents.

The event-log is in .XES format and is IEEE compliant. A total of 76,349 purchase documents containing 251,734 items are in the event-log. These items or cases consist of 1,595,923 events relating to 42 activities performed by 607 human users and 20 batch users. If the user field is empty or NONE, this indicates there is no user recorded in the system. For each case, i.e. item, several features are recorded. The features and their explanation can be found on the website that hostes the challenge [1].

The features can be split among the documents, items and events as can be seen in the entity-relationship diagram 1.

1.3 Challenge

The process owner has several questions that can be answered in the challenge. The following questions are of interest to the process owner [1]:

- Is there a collection of process models which together properly describe the process in this data. Based on the four categories above, at least 4 models are needed, but any collection of models that together explain the process well is appreciated. Preferably, the decision which model explains which purchase item best is based on properties of the item.
- What is the throughput of the invoicing process, i.e. the time between goods receipt, invoice receipt and payment (clear invoice)? To answer this, a technique is sought to match these events within a line item, i.e. if there are multiple goods receipt messages and multiple invoices within a line item, how are they related and which belong together?
- Finally, which Purchase Documents stand out from the log? Where are deviations from the processes discovered in (1) and how severe are these deviations? Deviations may be according to the prescribed high-level process flow, but also with respect to the values of the invoices. Which customers produce a lot of rework as invoices are wrong, etc.?

The decision of our team was to put focus on the third question. As for this question it is necessary to at least partially solve the first question as well, the following report will cover the identification of the deviations within the process preceded by creation of a suitable set of models. The following section closely

reflect the approaches taken in chronological order.
Section 2 describes the cleaning of the event log performed before the model creation. In section 3 various insights and relationships of different concepts in the data are described. In section 4 we provide an elaborate overview of the model creation and a thorough description of the models themselves. After the models were created, we aimed to determine the documents that do not conform to the defined standard processes by the models, which is described in section 5.

2 Data cleaning

The provided event log had a number of problems which needed to be addressed before performing further analysis. The following section provides a clear overview of the steps taken in in order to maximize the value of future findings and models constructed using the cleaned event log.

2.1 Missing values

Of all the features, only three had any missing values summing up to total of 3289 cases with missing values for 16294 events. This corresponds to about 1% of all cases. The features with missing values for these cases are:

- case Spend area text
- case Sub spend area text
- case Spend classification text

Interestingly, for each activity in a case these features are missing. To indicate so, the values is replaced with *unknown* as we do not consider it necessary to discard these cases, because the problem is not very severe. Moreover in the future these values might be traced back if their absence provides any insights.

2.2 Dates

Secondly, we examine the time frame and duration of the purchasing documents. Since the data description [1] indicates all data should occur in 2018 the events, cases and purchasing documents occurring outside of 2018 have to be looked into.The distribution of all the events, the starting events and ending events of purchasing documents is shown in Table 2. There are purchasing documents starting and ending before and after 2018, whereas the data should only contain purchasing documents created in 2018. For the purchasing documents starting after 2018, the assumption is made that these are future orders that are placed by clients that order ahead. For the cases that end after 2018 but start in 2018, a similar assumption is made, namely that activities are scheduled in the future necessary for the completion of the order and can only be done after certain time elapsed. For the cases that have a date earlier than 2018, it is assumed that this data is wrongly registered in the log, based on the challenge description. To investigate these purchasing documents further, the events that happen before

Table 2. Events, starting and ending starting purchasing documents per year

Year	Events	Purchasing Documents starting	Purchasing Documents ending
1948	10	1	0
1993	9	1	0
2001	22	15	0
2008	45	1	0
2015	3	2	0
2016	6	2	0
2017	223	75	0
2018	1550468	76241	65268
2019	45135	11	11079
2020	2	0	2

2018 are looked into. Only three out of forty-two event types occur before 2018, these event types can be seen in Table 3, with the value count and the time of the event occurring.

Table 3. Events, starting and ending starting purchasing documents per year

event concept:name	count	unique	time of events
Create Purchase Requisition Item	7	1	00:00:00
Vendor creates debit memo	27	1	23:59:00
Vendor creates invoice	284	1	23:59:00

As wrongly entered data and time might imply wrong order of the activities within a trace the constructed models might be based on a wrong order of the activities and also wrongly computed. Sub-optimal performance of a document corresponding to unusually long duration. Thus, it is decided to drop purchasing documents that start before 2018. All events connected to this purchasing document, including the ones happening in 2018 or later, will be dropped to guarantee the validity of the data.

3 Data analysis

Regardless of the purpose for which the data is used in the future, after the data cleaning we perform an initial data analysis to get general statistics and insights about the event log. Firstly the relationships between different concepts are investigated, after which we look at the distribution of different variables in the whole event log and their distributions within a trace. We analyze the frequency of different activities and dedicate a whole subsection to the time analysis.

3.1 Entity relationships

From Figure 1, which was constructed examining the given data, it can be derived that in one purchase document multiple items can be ordered. For all of these items the attributes specific to the purchase document are the same. It is always the same vendor delivering all the items ordered within one document. Interestingly, the Goods receipt flag is specific to purchase document, but GR-Based Inv. Verif. is not. From this it also follows that one purchasing document might contain items of different categories. One item, or case, contains one or multiple events, which differ in their timestamp, user that generated the event and other attributes. Surprising is that Document Category name is an attribute that varies within a single item within a single document.
The diagram also shows the updated counts for all of the three hierarchical entities present in the event log after the data cleaning.

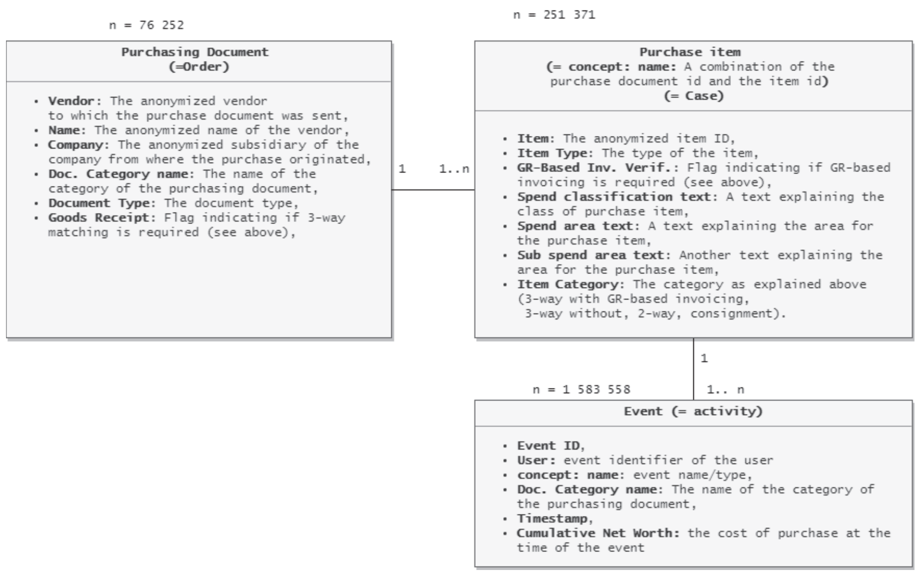


Fig. 1. Purchase documents, Items and Events shown in an entity relationship diagram showing their hierarchical relationship.

3.2 Activity frequencies

In the preprocessed log there are 42 different activities, of which following seven: *Record Goods Receipt*, *Create Purchase order Item*, *Clear Invoice*, *Record Invoice Receipt*, *Vendor creates invoice*, *Record Service Entry Sheet*, *Remove payment block* together create 90% of all events.

3.3 Attribute distributions

Firstly, the company distribution is very uneven. Out of the 4 companies present in the log the company with *ID_ 0000* is attributed to 99.66% of the events within the event log. Similarly, the Document Type is unevenly distributed, 97% of the events having *Standard_ PO* and *Framework order* and *EC Purchase order* only making small percentage of all. Of item categories *3-way match, invoice before GR* makes 77.9% of all events, followed by around 19.5% of event attributed to *3-way match, invoice after GR* and the rest split among *Consignment* and *2-way match*, see figure 2. For Item Type, the most occurring type is *Standard*, followed by *Service*, while the rest occur very little as shown in figure shown in Figure 2 as well. The most occurring values for Spend area text are *Packaging*, *Sales*, *Logistics* and *Trading & end products*.

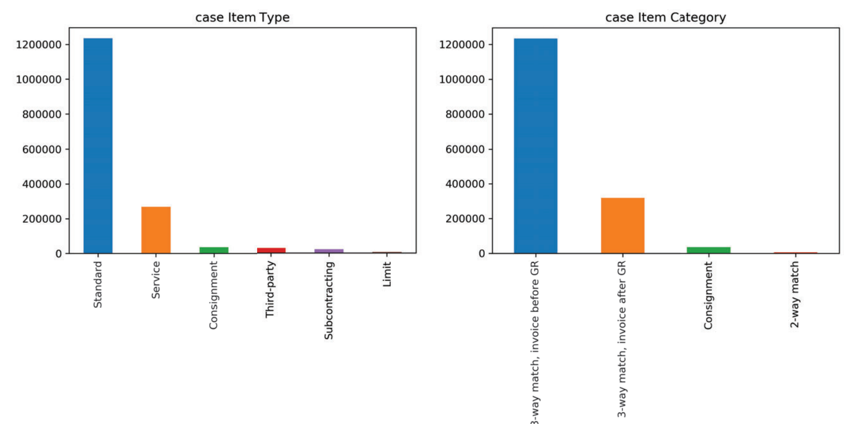


Fig. 2. Histogram showing the amount of events for each value for Item type and Spend area text

3.4 Time analysis

The median case duration after removal of the purchase documents with events from before 2018 is 64.3 days. The average lies at 70.6 days implying there is a small number of outliers that take much longer, which are probably the 5 cases that end after 2018 due to the assumed scheduled events in the future. The distribution of events generated on a certain day follows are regular interval of a work week, implying much higher number during a working day and a drop of the number over the weekends. Besides these drops, the number of events also drop heavily around Christmas period. Last regular pattern is a peak number of events generated on a single day, which happens monthly from March until November always between the 25^{th} and 30^{th} day of the month.

The number of cases open grows linearly from the beginning of the year until the 26^{th} of March at which point there are around 32000 case open. From April onward, the number of cases does not grow significantly anymore, but there is always a monthly peak of the number cases open between the 25^{th} and 30^{th} day of the month. The sudden drop after this day correlating with the peak in the number of events on that day suggest the events are closing events for these cases. On the 1^{st} of January 2019, there are still 20000 cases open which decreases to 5 until the 19^{th} of January.

Table 4. Events, starting and ending starting purchasing documents per year

Variable	Purchasing Document (max unique)	case concept:name (max unique)	Attribute of
case Goods Receipt	1	1	Purchasing Document
case Company	1	1	Purchasing Document
case Document Type	1	1	Purchasing Document
case Purch. Doc. Category name	1	1	Purchasing Document
case Vendor	1	1	Purchasing Document
case Name	1	1	Purchasing Document
case Source	1	1	Purchasing Document
case Spend classification text	2	1	case concept:name
case GR-Based Inv. Verif.	2	1	case concept:name
case Item Type	2	1	case concept:name
case Item Category	3	1	case concept:name
case Spend area text	4	1	case concept:name
case Sub spend area text	7	1	case concept:name
case Item	429	1	case concept:name

4 Models

It is desirable to produce a process model in the form of a classical Petri net, allowing for conformance checking with the original data and extensive analysis.

Multiple quality measures are taken into account when constructing the model. Further on, the generated process model will define the standard behaviour that will allow for deviation identification.

4.1 Model creation methodology

We split the data among the four line-matching categories as mentioned in the challenge description [1]. This is done because the four categories represent distinct process flows, specifically when it comes to 2-way and 3-way matching. The four process models created from these matching categories can then be properly compared against the event log, analyzing how a case in a category deviates from its respective process model. Additionally, as shown in Figure 2, the case item categories specifying the type of matching used, are not distributed evenly among the cases.

For model discovery, we used the ETMd Pareto Front Miner [2] in Live Mode. The ETMd algorithm is an evolutionary miner that creates random variations on process models at each generation. These random variations can then be used to propagate different types of models, ranging from relatively simple models (accuracy) to extremely complex models (fitness). At each generation, a population of new process models is created with the results of the previous generation as input, combined with a random factor. The algorithm trains on the previous generations and keeps the best mutations as an elite population to be passed along to the next generation.
The ETMd miner's Pareto front shows process models against two performance measures, where we chose fitness and precision to be determinant. An example of the Pareto front can be seen in seen in Figure 10. As we wish to analyze deviations, we typically do not want extremely high fitness, as it would cover deviations in the event log. However, achieving an equally high precision would result in an overly simplistic model, resulting in non-deviating but less frequent cases as deviating.
Using the ETMd miner, we created several process models for each matching category, ranging from higher-fitness/lower-precision to lower-fitness/higher-precision and compared these against the event log. For the ETMd parameters, we used a *population size* of 50 and an *elite count* of 5. We ran the ETMd miner for 500 generations for each of the four matching categories, enough for the Pareto front to converge. After a certain number of generations, the Pareto fronts were near-optimal and a marginally better model would have been found after an extremely long time or not at all in the case that there are no better models.
We used the Multi-perspective Process Explorer [3] plugin within ProM to do conformance checking of the process models generated from the ETMd miner, with the results denoted in table 5. More importantly, we used the Multi-perspective Process Explorer to export the alignments results between the created models and the event log, to be analysed in more detail in the following sections.

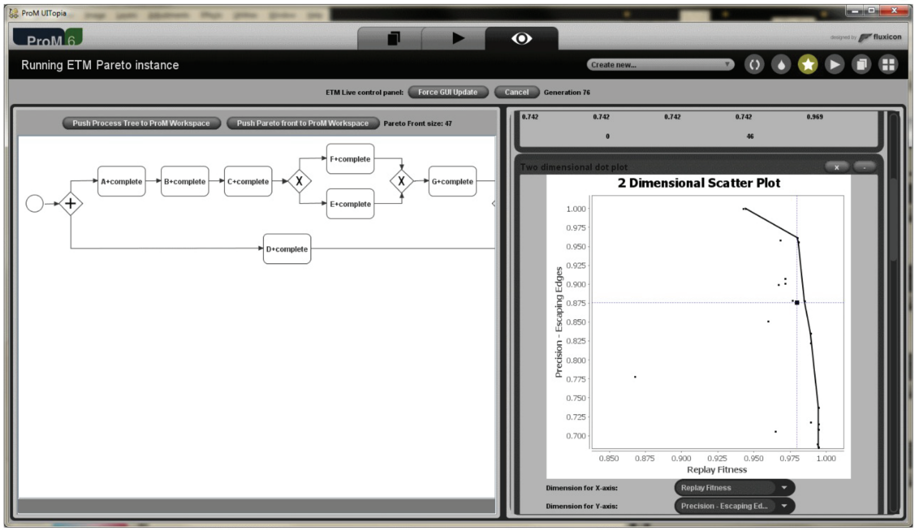


Fig. 3. Live Mode View view of the ETMd miner - Buijs, J. C. A. M. (2014). Flexible evolutionary algorithms for mining structured process models.

Category	Fitness	Precision
2-way match	87.2	92.6
Consignment	94.4	97
3-way-match Invoice after GR	87.3	84.6
3-way-match Invoice before GR	93.6	91.9

Table 5. Fitness and precision for Petri net models created against the event log of the corresponding category

4.2 Final models

Following section presents Petri nets corresponding to the process flow of one of the Item Categories. We provide a figure exported from the Multi-perspective Process Explorer with no modifications, despite the fact that automatically generated Petri net might contain redundant transitions or places that do not allow for extended behaviour. Instead we aim to interpret each of them and reflect on whether the model allows for behaviour that it should allow. The Petri nets contain black transitions with no name which shall be interpreted as silent (tau) transitions, which change state and correspond to no real activity of the system. Despite the fact, we use the Petri nets to check for conformance with the data, we also provide a process map from Disco that shows the absolute number of cases following certain paths.

2-way match The model for 2-way match4 contains a split initially, in which either *Vendor creates debit memo* happens, or *Vendor creates invoice* happens, or an arbitrary number of *Change Approval for Purchase Order* happens. These three options are structured in an or-split which mean only one of them shall happen in the model. Afterwards, *Create Purchase Order Item* should always happen, followed by again arbitrary amount of *Change Approval for Purchase Order*. Afterwards, both *Clear Invoice* or *Record Invoice Receipt*.

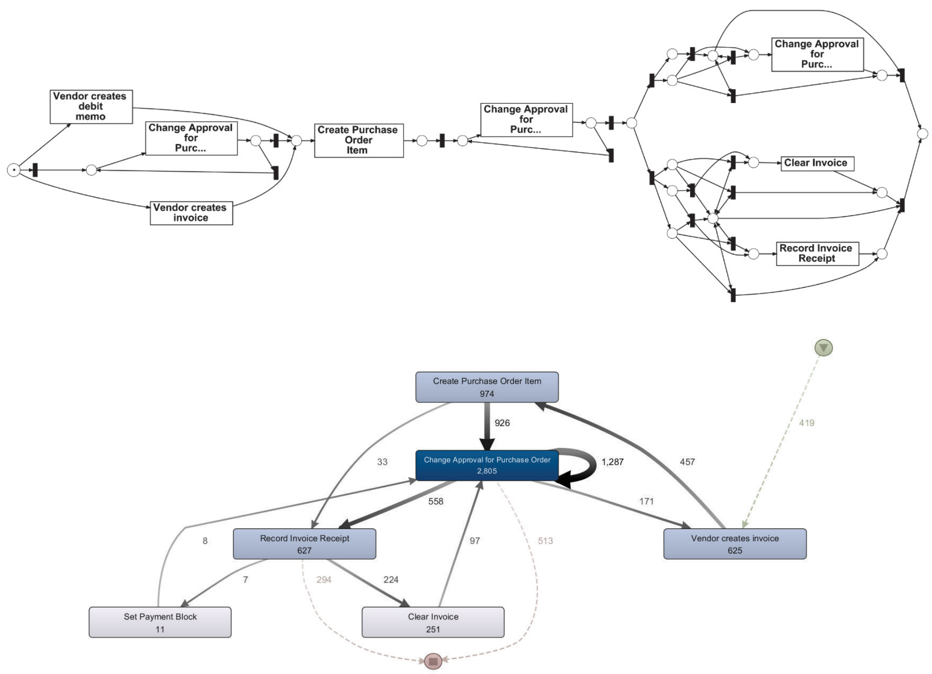


Fig. 4. Petri net used for the 2-way match flow and a process map reflecting the data.

Consignment The Petri net 5 is considerably straightforward in its initial part and the or-split can be interpreted as the process starting either with *Create Purchase Order Item* or *Change Price*, exclusively. Interestingly, no case within 2-way match data set after cleaning starts with the *Change Price*, neither there are cases in which both would be never followed by one another(or in another words exclude each other) as the model presents it. This is however not an issue

as the *Change Price* has relative frequency of 0.12%, so even if the model does not reflect the process perfectly, placing the *Change Price* activity in or-split with *Create Purchase Order Item* does not cause a problem.
Second part of the Petri net, following the silent transition has 5 activities: *Change quantity*, *Receive Order Confirmation*, *Delete Purchase Order Item*, *Record Goods receipt* and *Change Delivery Indicator*, which are in or-split, but the silent transition backwards allow for any order and any number of repetitions of these activities. This raises a question whether the model is not too loose and allowing for behaviour not present in the data, but the precision of the model is 97%, which ensures that the behaviour in the data is indeed closely represented by the model. The process map suggests that a few activities might happen after the order is created, in most of the cases, the last activity is *Record Goods Receipt*.

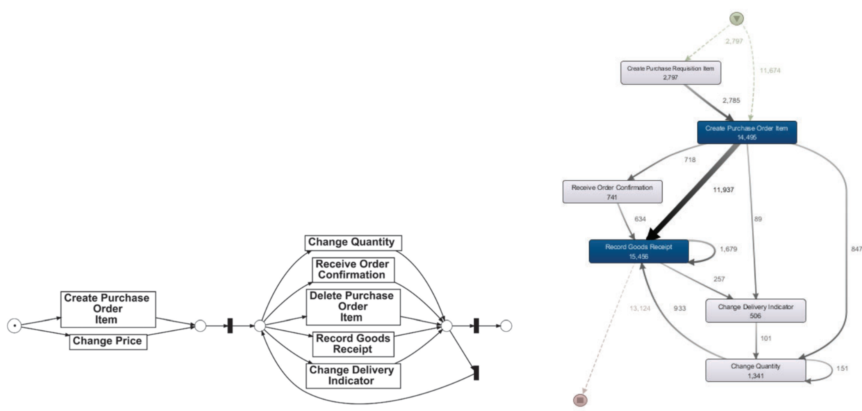


Fig. 5. Petri net model used for the consignment flow and a process map reflecting the data.

3-way match - Invoice after GR *Create Purchase Order Item* activity is optionally followed by *Receive Order Confirmation*. Afterwards there is an and-split, which implies that Vendor creates invoice should always happen at some point in the process before *Record Invoice Receipt*. In parallel, either different combinations of *Record Service Entry Sheet*, *Record Goods Receipt* and *Change Quantity* may happen, varying in order and the number of activities. An alternative to this is the combination of *Change Price* and *Vendor creates invoice* in any order. After *Record Invoice Receipt*, *Clear Invoice* and *Remove Payment Block* happen in arbitrary order.
Considering the process map 6, it is expected that the automatically generated SRM processes will be marked as non-conforming to our created model. The process map also shows many times after clearing an invoice, new invoice was created, which is also not accepted by the created model and will likely be marked

as non conforming in the deviations analysis. After examining the process map, it also becomes clear why the loop in the central part of the Petri net was added by the mining algorithm, considering the large number of repetitions of *Record Service Entry Sheer*, highlighted in dark blue.

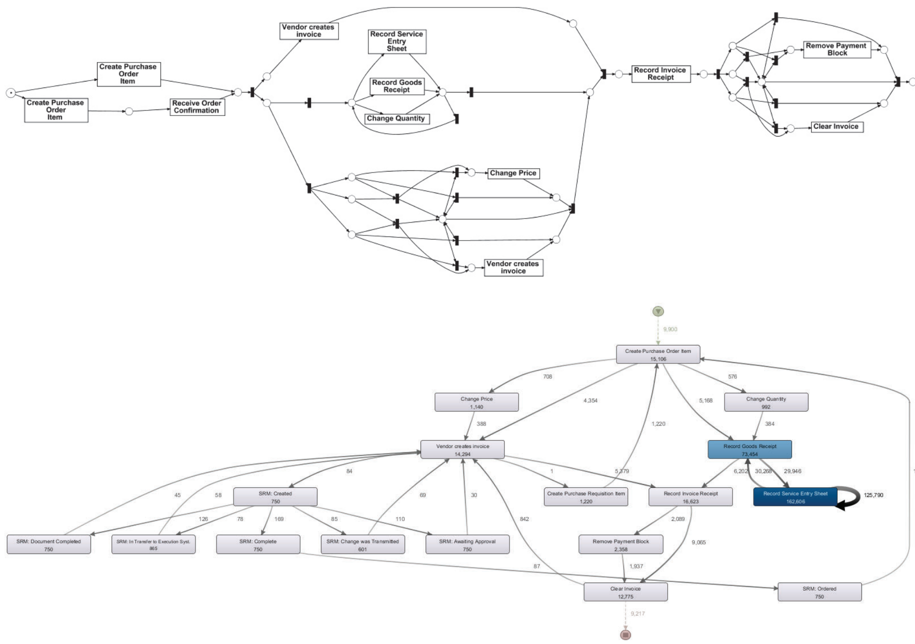


Fig. 6. Petri net used for the 3-way match - invoice after GR flow and a process map reflecting the data.

3-way match - Invoice before GR The process model 7 for this Item category is again rather straightforward, considering that this is the largest category of the four. The process model requires *Create Purchase Order Item* to happen first, after which it is necessary that *Record Goods receipt* happens sometime in the process before *Clear Invoice*. Interestingly, the model does not require *Vendor creates Invoice* to happen before the goods are receipt as the name of the item category suggests. Optionally, before *Clear Invoice*, *Remove Payment block* may happen.

5 Deviations identification

The data is analyzed to identify deviations, which is done to give an insight into which features create these deviations. The deviations are distinguished from

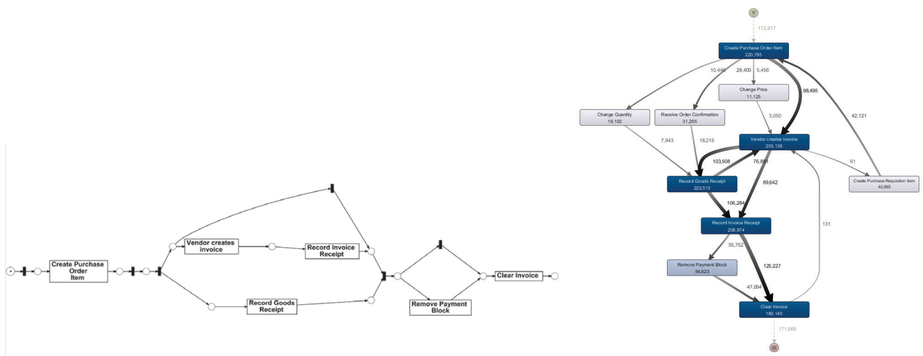


Fig. 7. Petri net used for the 3-way match - invoice before GR flow and a process map reflecting the data.

the standard process based on the models previously described for each Item Category. The process owner can use this information to look further into these features and see if improvements can be made to solve the deviations or find underlying reasons for the deviations.

5.1 Alignment fitness analysis

For the first of the analyses, we use the trace fitness calculated based on the models introduced in section 4.2 to see which factors have influence on the degree to which a case complies with the business process. In order to do so, it is assumed that the model represents the true process of the company. This assumption is reasonable as the fitness and precision of the models are quite high.

First, the mean case fitness is evaluated over the course of 2018. This is done for cases in general, but also for each of the four flow types. The result is depicted in Figure 8.

For all categories, the fitness remains more or less constant up until August; however, the case fitness rapidly decreases after that month. This suggests that the process flows have changed from those of the models depicted of section 4.2.

Besides time, other factors could be correlated with the fitness of a case. It was found that the case fitness exhibits significant correlations with the throughput time of a case and the cumulative net worth of the events of that case. The results are displayed in Table 6.

Clearly, cases that have a longer throughput time tend to comply less with the models, as one could have expected. interestingly, cases that tend to have events with a high cumulative net worth attribute, also tend to have a lower fitness.

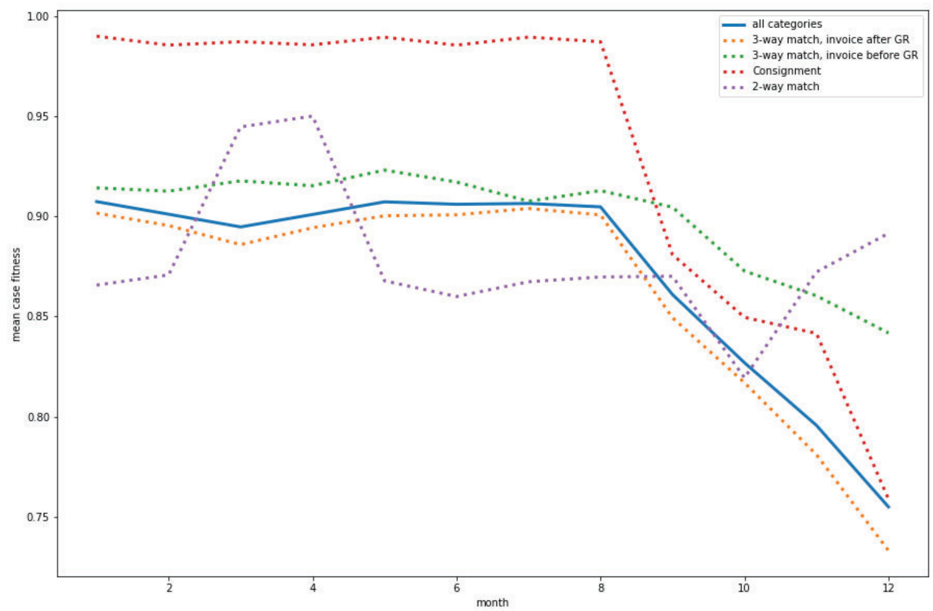


Fig. 8. Mean fitness of events plotted by the months of 2018

Also, the vendors of which the cases have the worst fitness were investigated. To do so, the mean fitness per vendor was calculated. The worst percent of vendors in terms of fitness are shown in Table 7. These vendors should be looked into to find why the fitness is so low. Special attention needs to be given to the vendor with Vendor ID 0582 since this vendor has over 1000 cases and is in the lowest percentile of fitness.

	Fitness	Throughput time	Cumulative net worth
Fitness	1	-0.624	-0.113
Throughput time	-0.624	1	0.339
Cumulative net worth	-0.113	0.339	1

Table 6. Correlations of Fitness, Throughput time and Cumulative net worth

Vendor ID	fitness	Vendor ID	fitness
0039	0.083	1592	0.0833
0046	0.083	1593	0.0833
0055	0.147	1594	0.1470
0065	0.147	1595	0.1470
0090	0.135	1596	0.1351
0582	0.067	1598	0.0674
1254	0.147	1599	0.1470
1572	0.147	1600	0.1470
1576	0.090	1604	0.0909
1583	0.147	1608	0.1470
1586	0.1470	1613	0.1470

Table 7. Users that stand out relative to the mean case fitness

5.2 Case duration analysis

In order to see what variables can delay a case the mean duration per variable are compared and outliers are identified. The duration is used instead of the throughput since duration it is the inverse of the throughput and easier to understand. Outliers are defined as cases that are more than two standard deviations higher than the mean case duration over the variable. The first interesting finding is that there are a lot of vendors with a high mean case duration. These vendors might be worth looking into to find an explanation for the long mean case duration. The vendors in Table 9 have a mean case duration of more than two standard deviations over the mean. The event users that stand out are shown in Table 8.

The previously mentioned users, vendors and sub spend area text are useful to investigate further. The process could be improved by reducing the mean case duration for each of the groups mentioned in Table 8 and Table 9.

In Figure 8 it is seen that the mean case duration for cases containing subcontracting as item type is higher than for the other item types. Further investigation into why this is higher might be useful to improve the process.

User ID	Mean case duration
u153	153.316667
u203	134.000000
u252	146.118881
u279	141.392857
u388	134.124031

Table 8. Users that stand out relative to the mean case duration

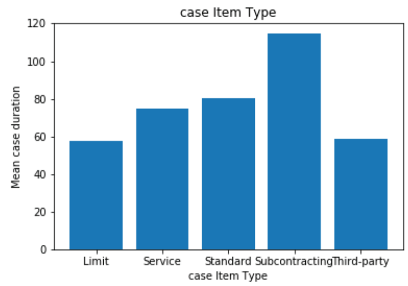


Fig. 9. Mean case duration per item type

Vendor ID	Mean case duration	Vendor ID	Mean case duration	Vendor ID	Mean case duration
0402	344.0	0780	214.684210526	1192	177.666666667
1914	328.0	0824	213.0	1430	176.0
1324	303.0	1024	211.0	1365	176.0
1318	297.0	1522	211.0	1542	175.0
1703	266.0	0532	201.0	0714	175.0
0335	263.0	1101	200.444444444	1386	174.0
1404	263.0	1750	197.0	1107	173.0
1413	260.0	1263	196.5	0025	171.0
0208	254.0	1534	196.0	1543	171.0
1175	247.0	1359	192.0	1472	170.333333333
1261	233.0	1162	192.0	1089	168.0
1194	227.0	1090	191.09375	1346	168.0
0002	224.0	1457	189.0	1523	167.0
0898	223.0	1382	187.310344828	1464	166.5
1428	218.5	0318	183.5	0555	166.458874459
0507	217.0	1489	183.0	1314	165.666666667
1085	216.045454545	1501	179.0	0871	164.5
0789	216.0	1023	178.222222222	1258	163.0

Table 9. Vendors that stand out relative to the mean case duration.

5.3 3-way Matching Analysis

Framework The 4 categories of matching as specified in section 1.2 require that each Purchase Document with line items follows a matching standard, with 3-way matching being dominant in the dataset. Thus, in this section, we provide further analysis on the verification and analysis of deviations in matching in the two 3-way matching categories. In both of these categories, the value goods receipts messages should be matched against the value of the invoice receipt messages and the value during creation of the item.

As both the number of Goods Receipt and Invoice receipt messages vary greatly among cases that themselves vary greatly in terms of duration, complexity and type, we use a framework to distinguish between 'matching' and 'non-matching' cases. To avoid taking into account duplicate entries for Goods Receipt or Invoice Receipt messages, we drop activities if they are duplicates regarding the following variables:

- Activity
- Case ID
- Complete Timestamp
- event Cumulative net worth (EUR)
- event User

Here, 'event Cumulative net worth (EUR)' represents the aforementioned 'value' that should be matched among (1) Goods Receipt messages, (2) Invoice Receipt messages and the creation of the line item, denoted as a (3) 'Create Purchase Order Item' message. These duplicate entries are do not occur frequently ($< 1\%$), but since the matching is dependent on an exact match of the cumulative values of the aforementioned three messages, a single entry that is duplicate within a case would result in a 'no-match'.

After removing duplicate entries, we calculated the values of the three messages (Goods Receipt, Invoice Receipt, Create Purchase Order) as follows, with V_e denoting the event Cumulative net worth (EUR) of an event with name e.

- Item Creation:
$$V_0 = V_{\text{Create Purchase Order Item}}$$
- Goods Receipt:
$$V_1 = \sum V_{\text{Record Goods Receipt}} - \sum V_{\text{Cancel Goods Receipt}}$$
- Invoice Receipt:
$$V_2 = \sum V_{\text{Record Invoice Receipt}} + \sum V_{\text{Record Subsequent Invoice Receipt}} + \sum V_{\text{Record Service Entry}}$$

We specify the following outcomes:

- Complete 3-way matching : *FULL_ MATCH* ($V_0 = V_1 = V_2$)
- Complete 3-way mis-match *NO_ MATCH*
- One or more messages missing, and as each of the three messages can be present or absent, there are $2^3 = 6$ possible missing outcomes: (e.g. *MISSING_ GOODS_ RECEIPT_ INVOICE_ RECEIPT*
- A partial match (e.g. *GOODS_ RECEIPT_ INVOICE_ RECEIPT*).

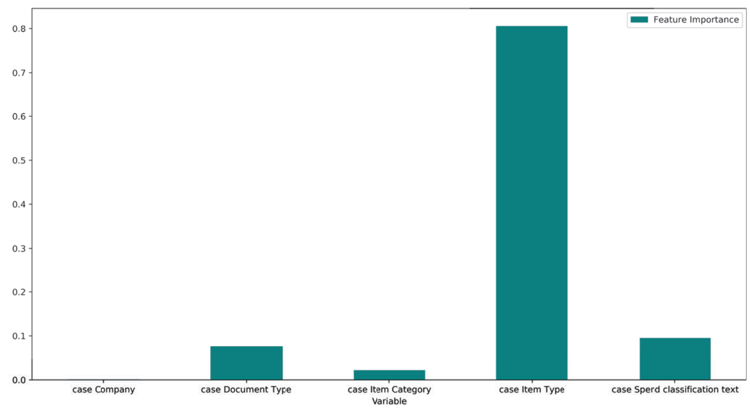


Fig. 10. Feature importance of variables [0,1] predicting matching outcomes using a Decision Tree Classifier

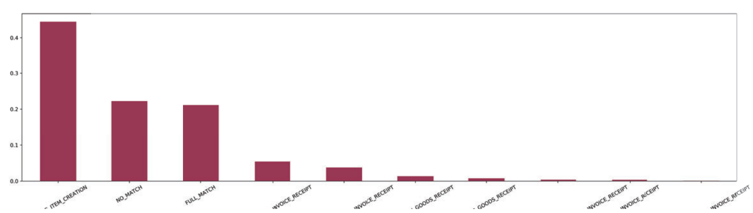


Fig. 11. Matching Results for item Type Service

Performance When calculating the matching on the entire dataset, we find that 87.26% of all cases are fully matching, requiring further analysis. To analyze what influences matching deviation, we used a Decision Tree Classifier using a train/test split of 80%/20% with the Match result as dependent variable, resulting in a test-accuracy of 88.07%. As seen in Figure 10, the *case Item Type* is the most important variable in predicting the matching result, and so we filtered on these cases to investigate further.

When strictly looking at a full matches, we find that the service type cases are considered outliers using the specified framework, this is seen in Figure 12. More than two-thirds of these mis-matching items are the result of these cases not having a *Goods Receipt* message, and many lack a *Create Invoice Receipt* message in their case or parent Purchasing Document, even though they occur completely in 2018.

Since these service-type items do not occur frequently (2.48%), we are interested in looking at the matching results for the rest of the dataset. Using the same train/test split of 80%/20%, the result is 95.16% test score, and now the

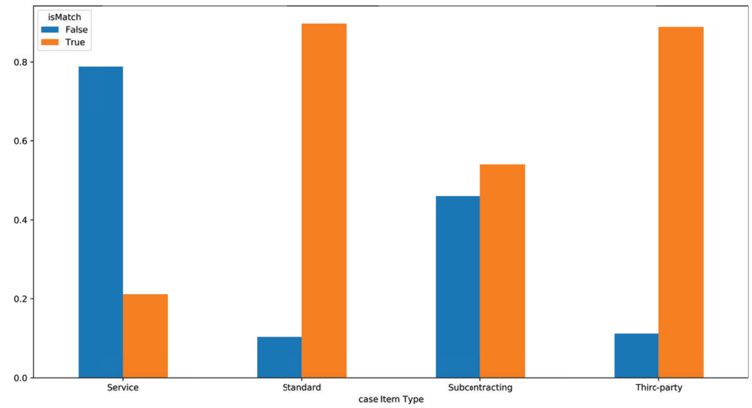


Fig. 12. Fraction of items completely matching (orange) for each case item type

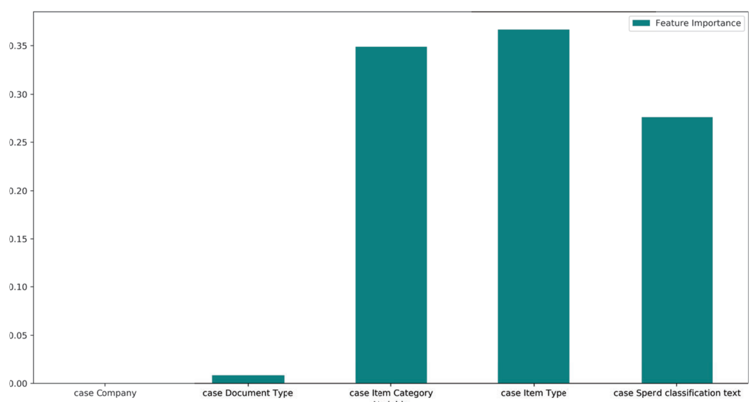


Fig. 13. Fraction of items completely matching (orange) for each case item type, without service items

case Item Category and Spend classification text together predict 99.17% of the matching outcome, as seen in 13. Without the service items, 90.94% of cases are fully-matching, and the following differences are observable for the non-matching cases:

- case Item Type: the type Subcontracting cases had less full-matches (53.69%) compared to Standard (89.36%) and Third-Party (85.04%) .

- case Item Category: The matching type Invoice-after-GR cases had less full-matches (80.93%) compared to Invoice-before-GR (88.77%).
- case Spend classification text: The cases with a case Spend classification text of 'Unknown' matches much less (62.20%) compared to the NPR, OTHER and PR cases. This 'Unknown' value is exactly what we used to fill the missing case Spend classification text values.

5.4 Conclusion

It is found that the case fitness changes from august onward. This might be due to a change in the process. It is also found that the fitness is negatively correlated with the throughput time. It is found that certain vendors tend to have a low case fitness. These vendors should be investigated further to see why this low fitness occurs. One vendor stands out due to the amount of cases this vendor handles. This vendor is the most important to look into and find an explanation for the low fitness.

It is found that there are a lot of vendors that have a high mean case duration compared to the mean. Looking into these vendors could be useful whilst trying to reduce the case duration. Users who have a high mean case duration are also identified, looking into these users is also useful in trying to reduce the mean case duration. Furthermore it is found that the item type subcontracting results in a higher mean case duration.

References

1. https://icpmconference.org/icpm-2019/contests-challenges/bpi-challenge-2019/
2. Buijs, J. C. A. M. (2014). Flexible evolutionary algorithms for mining structured process models Eindhoven: Technische Universiteit Eindhoven DOI: 10.6100/IR780920 https://research.tue.nl/en/publications/flexible-evolutionary-algorithms-for-mining-structured-process-mo
3. Mannhardt, Felix & de Leoni, Massimiliano & A. Reijers, Hajo. (2015). The Multi-perspective Process Explorer. CEUR Workshop Proceedings. 1418. https://www.researchgate.net/publication/282283719_The_Multi-perspective_Process_Explorer

Investigating Purchase-to-Pay process using Process Mining in a multinational corporation

Business Process Intelligence Challenge 2019

Aleksandra Rząd, Joanna Wojnecka, Maciej Rutkowski, Mateusz Guliński

PwC Data Analytics in Assurance, Polna 11, 00-633 Warsaw, Poland
[aleksandra.rzad, joanna.wojnecka, maciej.rutkowski, mateusz.gulinski]@pwc.com

Abstract. The Business Process Intelligence Challenge 2019 focuses on the compliance analysis of the Purchase-to-Pay process within a multinational company operating from The Netherlands. The participants are provided with a real-life event log data, which captures the information on the activities performed within the process of purchasing. Along with the data, the process owner suggests the specific scope for an analysis, which covers examination of at least 4 process models, throughput times of an invoicing process and spotting potential deviations. To address the challenges the analysis were divided into three main parts exploratory, process focused and predictive analysis, which were performed using various data analytics and process mining techniques including visualization tools, like SQL, PowerBI, Celonis and Python.

Keywords: process mining, purchase to pay, process discovery, data analysis, process compliance, BPI Challenge, predictive analysis.

1 Introduction

Due to dynamic development of the Information Systems field in the past decades, significant number of companies have made a decision to implement an IT system, gathering relevant amount of data on a daily basis. In an operational life of a multinational company, an integrated information system is a must. As a result, noticing the potential in data recorded, a business need for taking the advantage of the gathered data arises. In the times of Data Analysis, Data Mining and Business Intelligence emerged a discipline of Process Mining, combining those flows together.[1] As a fact-based insights provider, Process Mining do not require special means to collect the data. It relies on event logs derived from various operational datasets from different IT systems. Being a solution simple as such, dynamically growing popularity of the use of Process Mining techniques is observed. They allow to discover the actual as-is business processes and help to find potential inconsistencies, bottle-necks, nonconforming precedents, by providing a range of insights.

Within the Business Process Intelligence Challenge, a multinational company aiming at implementing means of control and surveillance to optimize its processes reaches out to Process Mining for a managerial support. Purchase-to-Pay process, as one of the core business processes, appears to have a lot of potential for improving the compliance within a company, ensuring it is in accordance with a prescribed set of norms, standards and policies. [2] This year's edition of BPI Challenge focuses on the compliance check of the Purchase-to-Pay process in an internationally operating company from The Netherlands, asking the participants to provide broad and unique insights.

This paper is divided into two main parts, which are Challenge overview and Analysis, conclusions and recommendations. The first part provides the necessary introduction into BPI Challenge 2019. Following, the analysis chapter incorporates the process of finding insights and growing conclusions. It is divided into 3 parts: exploratory, process focused and predictive analysis.

2 Challenge overview

This chapter includes an overview of the BPI Challenge 2019 and the approach taken towards its solution. Firstly, the case is briefly presented. Further on, the theoretical background for the Purchase-to-Pay process is brought up. In the last part of the chapter the approach towards the data is described, along with the tools overview and the basic data processing actions.

2.1 Case presentation

The organizers of the BPI Challenge 2019 provide participants with a real-life event log and ask them to analyze the data using techniques of a free choice. The process owner suggests some challenges to address, although the participants are allowed to provide a wide-range insights which reach outside of the given scope [3].

A company cooperating with the contest's organizers, called process owner, is a large multinational company operating from The Netherlands in the area of coatings and paints. The process owner provides the contestants with the basic information on the process, like the key flow of the data between Purchase Order creation, through Goods Receipt, until Invoice clearing. The flows are presented in more detail in the next subsection.

Challenges to address are focused on:

1. insights on the collection of 4 models corresponding to the 4 process flows mentioned above,
2. throughput times of the invoicing process,
3. deviations, rework activities, bottle-necks present in the process flows.

Despite being presented a set of suggestions, participants are free to analyze the case from different perspectives, in order to encourage the originality and usefulness of the outcome.

2.2 Theoretical background

Purchase-to-Pay process is recognized as one of the most important processes within a company's operational life. Purchasing provides core resources for leading a business on a daily basis and strongly influences overall costs and timing of production. Being based on the cooperation with suppliers, purchasing reaches out of a company's direct control. Due to the fact that the outer dependency is high, a great attention should be paid towards a purchasing strategy development. Such a strategy can be assessed using two factors. Whilst the first one is strategic importance of purchasing in terms of its impact on profitability, the second factor points to the complexity of the supply market and outer conditions. In order to reduce the risk of purchasing and supply to an acceptable minimum those two factors should be addressed by the top management and senior purchasing executives. [4] A discipline providing the support in this procedure is Process Mining. Having a possibility to look deep down into a process specificity potential issues can be discovered and addressed resulting in the process improvements. Not only bottle-necks, but also non-compliance procedures can be spotted. By pinpointing the development areas in the process it is possible to put the most of the improvement effort exactly where it is needed the most.

The main component of a supply procedure is Purchase-to-pay process. It is a coordinated and integrated set of actions taken to fulfill a requirement for goods or services in a timely manner at a reasonable price. [5] It involves a number of sequential steps, ranging from creating a purchase order, through goods delivery and paying an invoice.

In the Business Process Management exists a general ideal sequence of activities in the process, described as a desirable purchase process flow. However, there are several ways of confirming the compliance of the purchasing process. The organizers of the Challenge provide the participants with the description of 4 models, which determine the desired actions flow in the case. The models represent procedures of processing an invoice received from a supplier to ensure that a payment is complete and accurate. The goal of implementing the different types of procedures is to have a more precise control over the flows and highlight any discrepancies in the compliance between the three most important documents, which are purchase orders, goods receipts and invoices. [6]

The first of the models is called *3-way matching, invoice after goods receipt*. The value of the goods receipt should match the value present in the invoice corresponding to it, as well as purchase order item. Before the invoice is paid, accounts payable reviews what is ordered (purchase order), matches it with the received goods (goods receipt) and invoice to pay (invoice). If all documents are present and they match, an invoice is paid and cleared. In the second model *3-way matching, invoice before goods receipt*, the main difference to the previous one is that purchase items do not require a Goods-receipt invoicing. That means an invoice may be entered before goods are received, although it has a block set. Upon a receipt of the goods and a check if all of the three mentioned documents match, an invoice is unblocked and a payment is done. The third invoicing model is *2-way matching (no goods receipt needed)*. As the description suggests, in this type of a model only 2 documents must match, a purchase order item and an invoice. What is important is that a value of a purchase order might be consumed by multiple invoices, therefore a one-to-one match is not a must. The last category is *Consignment*. The check of a match in this model of invoicing is beyond the scope of the analyzed process, since it is handled in a fully separate one. There are no invoices registered in this event log and the analysis is conducted just for the first stages of purchasing. This model is mainly characterized by the item type named consignment.

2.3 Data overview

The data consists of over 1.5 million events recorded in 2018 within the Purchase-to-Pay process, without the workflow of the approval of POs and invoices. [7]

For the purpose of the contest, the dataset was anonymized. The company holds an anonymization key, therefore it is possible to translate the results and use the worth of the analysis in a real-life. Table 1 below presents the original dataset structure.

Table 1. The original dataset structure with the description of the fields.

No.	**Column name**	**Description**
1	Case Concept name	A combination of the anonymized purchase document id and the anonymized item id
2	Purchasing Document	The anonymized purchasing document ID
3	Item	The anonymized item ID
4	Item Type	The type of the item
5	GR-Based Inv. Verif.	Flag indicating if GR-based invoicing is required
6	Goods Receipt	Flag indicating if 3-way matching is required
7	Source	The anonymized source system of this item
8	Doc. Category name	The name of the category of the purchasing document
9	Company	The anonymized subsidiary of the company from where the purchase originated
10	Spend classification text	A text explaining the class of purchase item
11	Spend area text	A text explaining the area for the purchase item
12	Sub spend area text	Another text explaining the area for the purchase item
13	Vendor	The anonymized vendor to which the purchase document was sent
14	Name	The anonymized name of the vendor
15	Document Type	The document type
16	Item Category	The invoicing category
17	Event ID	The identification number of an event
18	User	The user ID recorded in the source system
19	Org. resource	The user resource involved in the process, always equals to User.
20	Event Concept name	The activity performed in the process
21	Cumulative net worth (EUR)	The anonymized value of an event
22	Timestamp	The date and time of an event

In order to carry out a process analysis it is necessary to create an event log which consists of case ID, activity ID and timestamp. Within the given dataset *Case Concept name* is suggested as a case ID, *Event Concept name* was chosen as the activity ID and column *Timestamp* was considered to be a timestamp.

To begin with the analysis appropriate data handling was required. To be able to perform transformations and adjustments the dataset was loaded into Microsoft SQL Server Management Studio. First, the data formats were reviewed. It was noticed that some of the values in *Cumulative net worth (EUR)* column are presented in the scientific notation (E-notation). A recalculation of these values into standard numeric format was applied and saved as *Cumulative net worth (EUR) new* column. Fig. 1 presents the values before and after the transformation.

	event Cumulative net worth (EUR)	event Cumulative net worth (EUR) new
1	2.899453E7	28994530
2	1.0652112E7	10652112
3	1.1475875E7	11475875
4	1.4202816E7	14202816
5	1.1475875E7	11475875
6	1.7213813E7	17213813

Fig. 1. Transformation of values from E-notion to a standard numeric format.

Secondly, in order to provide as much information as possible, a transformation supporting a creation of an event duration was implemented. The event log initially contained a start time of an event. To be able to calculate a duration of an event, the end time is needed. However, being given single timestamp per event, the exact calculation of a duration of an activity was not possible. Therefore it was agreed to calculate the time between the activities, so called lead time. Lead time provides the information of a duration of an event plus the time until another activity is performed. The start time of a next activity within one case ID was considered to be an end time of a current one and was moved to the current row into the new column capturing the end time of an event. The sorted activities, which took place within one case, the start times (marked respectively black and blue) were moved to the prior row and represent the end time of an event. Fig. 2 images the example of a change in order to support an understanding of a transformation.

	Timestamp	Case	Activity	Start time	End time	Row number
1	2018-01-02 13:53:00.000	2000000000_00001	SRM: Created	2018-01-02 13:53:00.000	2018-01-02 14:53:00.000	12
2	2018-01-02 14:53:00.000	2000000000_00001	Create Purchase Order Item	2018-01-02 14:53:00.000	2018-01-02 14:53:00.000	11
3	2018-01-02 14:53:00.000	2000000000_00001	SRM: Ordered	2018-01-02 14:53:00.000	2018-01-02 14:53:00.000	10
4	2018-01-02 14:53:00.000	2000000000_00001	SRM: In Transfer to Execution Syst.	2018-01-02 14:53:00.000	2018-01-02 14:53:00.000	9
5	2018-01-02 14:53:00.000	2000000000_00001	SRM: Complete	2018-01-02 14:53:00.000	2018-01-02 14:53:00.000	8
6	2018-01-02 14:53:00.000	2000000000_00001	SRM: Awaiting Approval	2018-01-02 14:53:00.000	2018-01-02 14:53:00.000	7
7	2018-01-02 14:53:00.000	2000000000_00001	SRM: Document Completed	2018-01-02 14:53:00.000	2018-01-02 14:53:00.000	6
8	2018-01-02 14:53:00.000	2000000000_00001	SRM: Change was Transmitted	2018-01-02 14:53:00.000	2018-01-02 23:59:00.000	5
9	2018-01-02 23:59:00.000	2000000000_00001	Vendor creates invoice	2018-01-02 23:59:00.000	2018-03-06 07:44:00.000	4
10	2018-03-06 07:44:00.000	2000000000_00001	Record Goods Receipt	2018-03-06 07:44:00.000	2018-03-06 08:53:00.000	3
11	2018-03-06 08:53:00.000	2000000000_00001	Record Invoice Receipt	2018-03-06 08:53:00.000	2018-03-29 15:06:00.000	2
12	2018-03-29 15:06:00.000	2000000000_00001	Clear Invoice	2018-03-29 15:06:00.000	2018-03-29 15:06:00.000	1

Fig. 2. The transformation performed to receive the end time of an event.

The lead time of an activity was then calculated by distraction of the start and end time and given in days. In cases where the activities are registered at the same time or the difference of time is below 24 hours, the calculated lead time equals 0. For the purpose of exploratory analysis such an assumption for calculation is well enough to get insights, although considering the predictive analysis the lead time (duration) include more detail and the lead time is given in a more precise format.

Another transformation considered the *User* column. The activities can be performed by automatic (batch) or manual users named *batch_xxx* or *user_xxx,* where *xxx* specifies particular user's number. In order to gain information on automation within the process in a facile manner, a new column was created. The values of the new column are 'B', 'U' and 'NONE' indicating automatic users, manual users and missing values, respectively.

Considering the presence of outliers in the data, it was noticed that some events recorded do not take place in 2018. Since the challenge overview clearly indicates that the analysis should correspond to purchase orders submitted in 2018, the ones not fulfilling this requirement were filtered out from the event log (~500 out of 1.5 mln events). The filtration was performed with respect to cases which started before January 2018 or after December 2018 and some of their activities took place in 2018. This means that if at least one event within a case happened in 2018, this case was incorporated. What is more, the events with the date later than 27.01.2019 were considered to be outliers, due to the fact that the data was published on 28.01.2019 and were filtered out (9 events). Moreover, looking at the timeline from the perspective of prior years, it was noticed that there are events taking place in e.g. year 1948. Having performed more detailed analysis it was found that the activities of the events happening in years 1948-2016 are Vendor Creates Invoice (284 events) and Vendor Creates Debit Memo (27 events) . Since the impact of those activities on the purchase process is not considered to be great from the company's operations' perspective, as they have mostly informative function from vendor's side, they were filtered out from the log. In a real-life process intelligence project the team reaches out to the process owner for the double-check of those cases and the nature of creating those two activities, although in the form of the contest the contact with the process owner is strongly limited, that is why those cases are spotted and approached in the form of deletion.

The changes and transformations applied to the data for predictive purposes are described in chapter *3.3 Predictive analysis*.

3 Analysis, conclusions and recommendations

In order to arrange the most applicable way of finding answers for the process owner's challenges, a specific approach towards the analysis division was taken. To begin with, the exploratory analysis was conducted providing the information on the overall process performance and supporting the recognition of potential deviations. Next, the process focused analysis was performed in order to gain insight into the activities sequence, sub processes performance, throughput times and potential deviations. The last part captures predictive analysis focused on anticipation of throughput time and occurrence of undesired activities, performed with usage of machine learning methods.

3.1 Exploratory analysis

The aim of the exploratory chapter is to perform an elaborate analysis, giving the most valuable insights and facilitating an effective discovery of Purchase-to-Pay process performance. In order to do that, a set of interactive dashboards was created using Microsoft Power BI visualization tool. Fig. 3, 4 and 5 present snapshots of the created dashboards. The application was published under **this link** and can be easily accessed by the reader. It helps to understand exploratory analysis performed within this chapter and provides a general overview on process KPIs as well as more detailed indicators and measures.

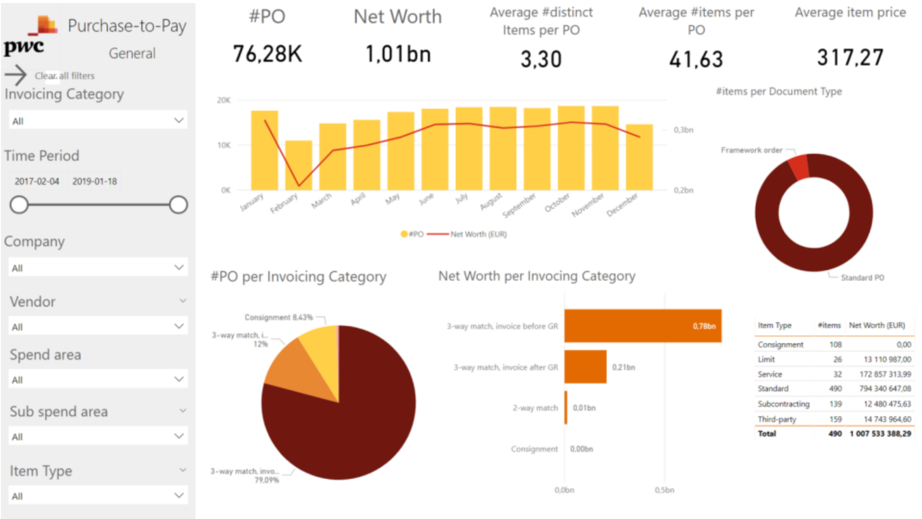


Fig. 3. General P2P dashboard of PowerBI application.

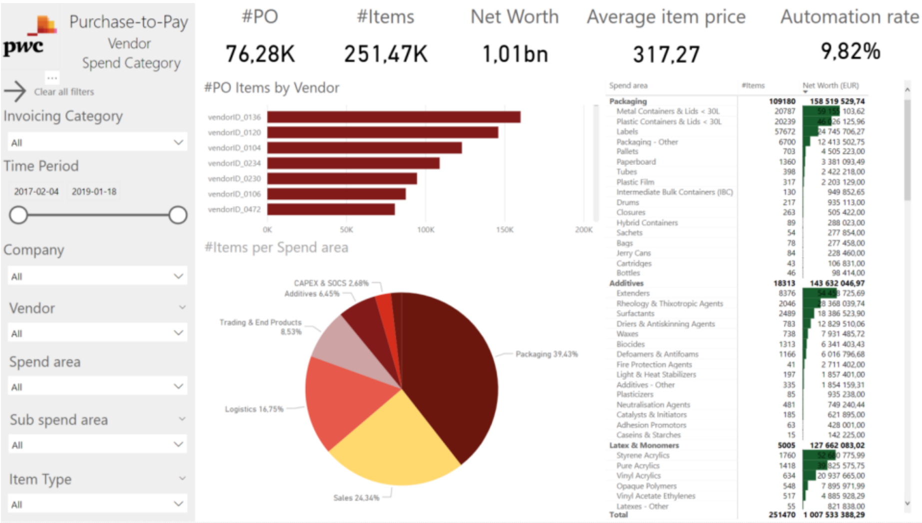


Fig. 4. Vendor and spend category dashboard of PowerBI application.

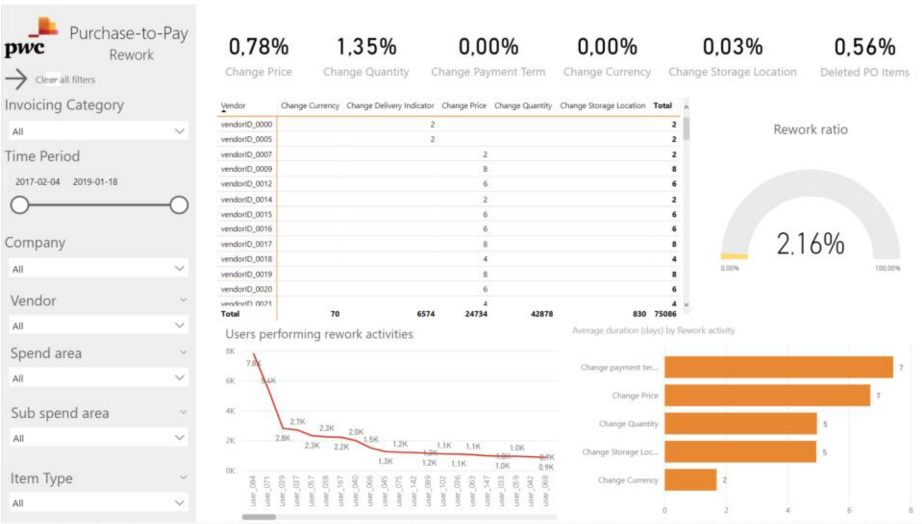


Fig. 5. Rework dashboard of PowerBI application.

PowerBI application. The Power BI tool served gathering findings within the scope of exploratory analysis and the application's construction is briefly described in the following paragraph. The application concentrates on giving the general background of the overall process characteristics and is divided into three thematic dashboards. The first one provides a general overview on the size and patterns within the process. Furthermore, it enables to investigate shares of respective invoicing types in the whole, regarding the net worth and number of purchase orders. The perspective on the timeline in 2018 is also provided to examine the changes of purchasing in time. Additionally, document types and item types are distinguished. Second dashboard concerns vendors and spend categories. The diagrams give the possibility to find insights on the importance of vendors, which the company collaborates with and spend areas of purchasing. The purpose of this dashboard is to gain knowledge on the subject of purchasing and the sources in terms of suppliers. The third dashboard is specifically process oriented and deals with rework activities. Not only the ratios of most popular rework activities are given, but also detailed visualizations of vendors and users performing them. On top of that, an average duration time of rework activities is presented in order to stress the impact they have on the time of a process.

The overall process from 4 invoicing categories' perspective was examined. Fig. 6 presents the proportion of shares of respective invoicing models in the whole process regarding the number of purchase orders.

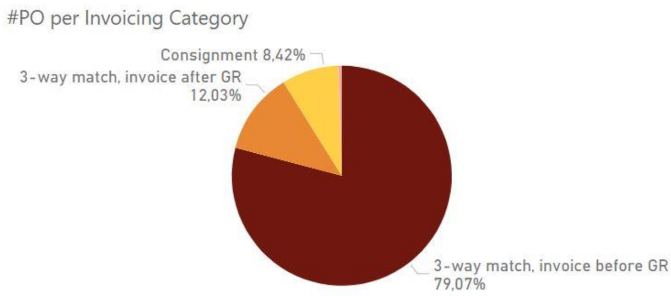


Fig. 6. Graph presenting the shares of invoicing categories in the process in terms of number of purchase orders.

3-way match, invoice before GR**.** The most popular invoicing model is *3-way match, invoice before GR.* It captures the largest net worth of purchasing (~€784M out of €1bn). At the same time it dominates significantly among other models as the one which serves the most purchase order numbers (~79% of all of the purchase orders). The timeline graph indicates that in February the purchasing is less intense than throughout the rest of the year, which correlates strongly with the total purchasing pattern. Document types are not diversified, since 98% of the purchase orders are of a *Standard PO* type. The definite majority of items is of a standard type also. Purchase orders contain on average 3.56 different items and around 40 on average of total items. The average net worth of an item equals around €318, which is relatively small (comparison to the other invoicing models €1.14 thousand for *2-way match*, €336 for *3-way match, invoice after GR).* This indicates that most of the purchases concern less valuable goods for this model, but they happen regularly in large quantities (62 thousand of purchase orders in 2018).

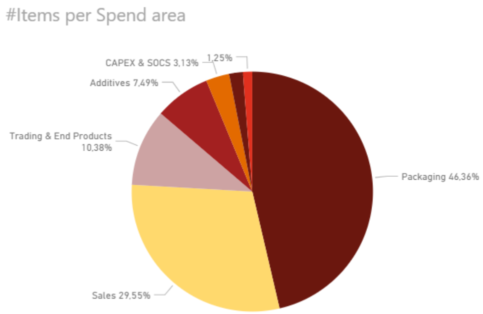


Fig. 7. Graph showing most important spend areas in *3-way match, invoice before GR* model.

Fig. 7 shows the main areas of spending regarding the number of purchase orders, which are *Packaging*, *Sales* and *Logistics*. Considering the net worth, an important spend area is also *Additives*, which captures a type of products needed for manufacturing. The amounts spent on *Packaging* and *Additives* are almost equal, followed by *Latex&Monometers* and *Titanium Dioxides*, which relate to the production processes. The most significant rework activity is *change quantity* (present in 1.55% of the cases). The time impact of the change of quantity activity is quite high. It prolongs the process by around 5 days. *Change price*, which ratio reaches 0.90% also seems to be an activity to avoid, since the time impact on the process equate up to 7 days. It can be noticed that in 2018 *user_84* performed the largest number of quantity changes whereas *user_071* was the main user changing prices.

3-way match, invoice after GR. Second most popular invoicing method regarding the net worth of the purchase orders and purchase orders number is *3-way match, invoice after GR*. This seems to be a category, which follows the most straight-forward and orderly rules, since the purchase order is first reconciled against goods receipt and then against invoice. Despite its intuitive matching nature it captures 12% of all of the purchasing transactions. The registered number of purchase orders is around 9,500. The average net worth of a purchase order item is €336 and average number of items per purchase order is around 66, whereas distinct items is 1.6. This means the purchasing done within this invoicing method concerns mostly the purchases of rather large quantities of undiversified items. The most of the costs are from *Road Packed* sub area from *Logistics* category. Also *Digital Marketing* costs underlie mostly this invoicing category. What is worth mentioning is a high automation rate, indicating the ratio of batch users to human users, amounting to more than 20%. Rework activity, which occurs the most often is *change price*, although the ratio of 0.38% is not too alarming. The user, who is involved the most into rework activities is user_038, which does mainly price changes and they correspond mostly to vendorID_0236 (744 out of 758 cases).

Consignment. Next, *Consignment* model is examined. It covers around 8.5% of all purchase orders and it is the third most common invoicing method in the process. However, what characterizes this type of a model is zero values in the net worth of orders. The reason for it is the nature of consignment items. A consignment is a business arrangement in which goods are left in the possession of an authorized third party to sell, while the ownership stays with the vendor. The goods bought within this type of an arrangement are possessed by the company and purchase orders are issued, they contain no value of the goods though. The invoicing process happens within a separate process, there exists no invoices for Consignment items in the dataset. Most of the purchases happen within *Packaging* spend area, followed by *Additives* and *Latex&Monomers*. Most of the rework activities concern *change quantity*, that is in 3.72% of the cases.

2-way match. *2-way match* model covers the least number of purchasing transactions (not even 1% of all purchase orders). Just one item type occurs within it, which is *Limit* and it is the item type that is only associated with *2-way match*. It is a specific type of an item, which does not have a price, but rather an upper financial limit set, which can be spend on it. What is interesting, just one subsidiary *companyID_0003* performs this type of a purchase. The Vendor/Spend Category dashboard unveils, that the spend areas are *Real Estate* (Sub areas: *Real estate services, Real estate brokers or agents*) and *Others* (Sub areas: *Government Payments, Taxation*), C*APEX & SOCS,* which explains the type of the invoicing method and the item type used. *2-way match* is namely used mostly in cases, where no physical goods are shipped and the purchase order is matched against an invoice directly. No fixed price of an item is therefore understandable in such cases. There is just one user user_602 responsible for that model, who handles all the activities within it. Additionally, the only document type used is Framework order. In the case of *2-way match* and *Consignment* models automation rates are kept on a low level, which is determined by the nature of those purchasing types (0.38% for *2-way match* and 1.90% for *Consignment*).

3.2 Process focused analysis

After having a glance on the characteristics of the process in the exploratory analysis part, a process mining was incorporated to perform process focused analysis. The software used for this purpose is Celonis. Not only the process maps available in Celonis were analyzed, but also the potential of its visualization capabilities were put to use. Additionally, a conformance functionality served as a reference point for the root cause analysis. The event log exists within the dataset, which exploratory analysis was based on. The process of preparing the event log is described in the 2.3 Data overview chapter.

Invoicing models' processes comparison. To distinguish between 4 invoicing models the process analysis was split and compared in terms of the time performance. Further on, however, the general common process map was analyzed regarding the deviations.

Due to the fact that the median is less vulnerable than average in terms of outliers' influence, a decision was met that all throughput times will be given in median instead of an average. To distinguish between 4 models Fig. 8 below presents median throughput times for all of the 4 invoicing models. The most common category *3-way match, invoice before GR* takes the longest to process (72 days). *2-way match* takes 62 days, it is the rarest category and at the same time the one that serves the most complicated cases from purchasing perspective. *3-way match*, *invoice after GR* is performed much faster, its median throughout time is 30 days, which is more than a half shorter than the similar "before" model. *Consignment* is characterized by relatively short overall process time, which equals 21 days.

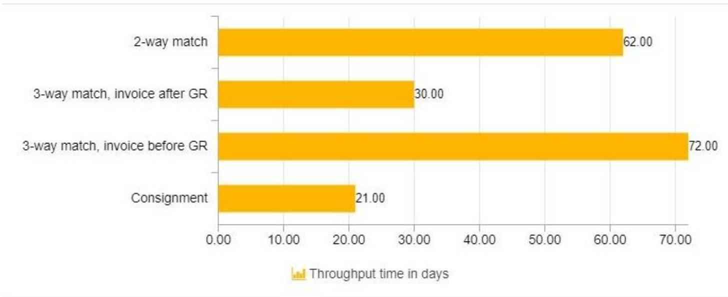


Fig. 8. Average process throughput times of invoicing categories.

The happy path (the most common path) of the overall process consists of Create Purchase Order Item, Vendor Creates Invoice, Record Goods Receipt, Record Invoice Receipt and Clear Invoice. Fig. 9 shows it together with median throughput times for activities. A median throughput time of the whole process is 72 days, whereas the cycles PO Item creation to Goods Receipt is 8 days, Goods Receipt to Invoice Receipt is 13 days and Invoice Receipt to Clear Invoice is 36 days.

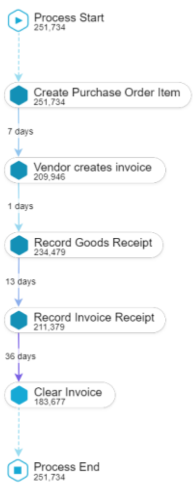


Fig. 9. Happy path of the overall process.

Role of SRM in the process. In order to perform efficient and upright analysis, the SRM activities are grouped into SRM group and kept aside the main flow of purchasing. SRM stands for Supplier Relationship Management, which is a software supporting requisition system. It handles creating requisitions, monitors approvals and documentation and passes the information on purchase requisitions further until a creation of a purchase order item. It is treated as an outer software solution, having a supportive role in the purchasing process. It serves 1,440 cases in the process and often happens in parallel to the creation of PO or Requisition Item (the same date of activities) and is mostly automated. It does influence the process time overall, since it is shorter in case of using SRM (median 65 days instead of 72). The process including SRM shortens the cycle time between invoice receipt and invoice clearing up to 17 days (versus 36 days with no SRM). However it doubles the goods receipt-invoice receipt cycle in comparison to no SRM process (26 versus 13 days). The vendors, which are associated with the developed SRM activities sequence are mainly vendorID_0003 and vendorID_0000. Those two vendors belong rather to the minor suppliers, since the annual number of PO is 133 together (out of ~76 thousand number of PO in 2018).

Vendor Creates Invoice activity. Vendor Creates Invoice is an activity, which is not usually taken into consideration in a theoretical presentation of the process, it is not a negative activity though. It appears in 83% of the cases and provides additional information on the process and it is assumed, that it happens in parallel to the shipment of the goods, since the time difference between Create Purchase Order Item and Vendor Creates Invoice is 7 days (median) and later until Record Goods Receipt it takes 1 day. In the cases where this step is not incorporated (41,788 cases) the process shows noncompliance on the stage between Record Goods Receipt and Record Invoice Receipt, due to the fact that its median throughput time takes 183 days. While performing root cause analysis for this issue it appeared that most of those purchases are done with cooperation with vendorID_282 (*CAPEX&SOCS*) and vendorID_0246 (*Trading & End Products*). The activity of recording an invoice is mostly performed by batch_01 automatic user. What is more, recording an invoice is the last step in the process, which indicates it is not cleared. These cases are unfinished by the sides, since there is no information from the vendor on the creation of an invoice, it takes enormously a lot of time to record it, recording is done by automatic user and invoice is not cleared. It was assured that the purchases of this type did take place throughout the whole year, therefore it is not the case that they are still on-going in the sense of being fresh (timestamps date back to January 2018).

An interesting phenomenon from the deviations' spotting perspective is the appearance of Vendor Creates Debit Memo and later directly Vendor Creates Invoice. It happens in more than 4 thousand of cases, therefore it is quite often appearing sequence. Since the Debit Memo, considered to be a correction to the invoice, is created before the actual invoice is prepared it seems to be an obvious inconsistency. The vendor, which happens to be associated with it the most is vendorID_0118 (*Trading & End Product*s) from *3-way match, invoice before GR* category.

Deletion of purchase orders. Another deviation occurring while no Vendor Creates Invoice activity appears in the process flow is Delete Purchase Order Item. In 8,839 cases, where is deletion of PO, 8,530 happen while there is no information on creation of an invoice from the vendor's side. Therefore it can be concluded that the correlation between those two is not accidental and information on creation of an invoice is thought to be a confirmation of receiving an invoice. Deletion of POs happens mostly within subsidiary company_0000 (just 31 cases come from another subsidiary). None of the vendors or users seem to cause that much of these reworks so that it would be outstanding. On the other hand, a part of deleted POs is reactivated. Fig. 10 below presents the sub process for reactivation of deleted POs. In reactivation the process flow goes in a generally compliant way. After a reactivation of PO, goods are received, invoice is created and recorded, then cleared. The only alarming point is the duration of the flow between recording invoice and clearing it (47 days). It is not much longer than the median throughput time in this cycle for the whole process though. However, this time is not considered to be optimal and a potential in speeding up this cycle is noticed.

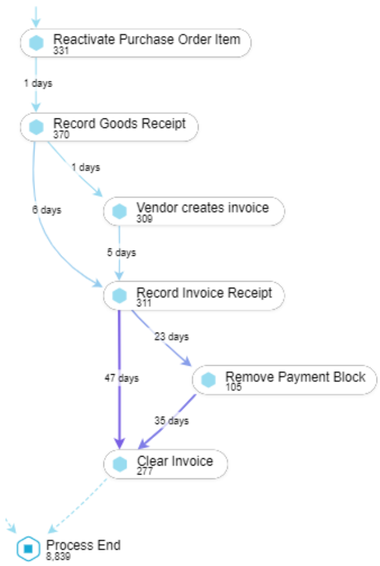


Fig. 10. Sub process map for reactivation of deleted POs.

Purchase order items creation not being a start activity. Create Purchase Order Item activity takes place in 100% of the cases, which is a positive check. However it does not always start the purchasing process. Fig. 12 includes a list of activities the process might start with, along with the case frequency. Although the PO item is created in some cases after a requisition item creation or some SRM activities (which is compliant), nearly 3,416 cases flow through Vendor Creates Invoice first. Worth mentioning is that the median process throughput time is 17 days shorter then. It is mostly clearing invoice cycle, that is faster (median 16 days). This kind of process deviation happens mostly while trading with vendorID_0550 from *Trading & End Products* spend area (1,213 cases).

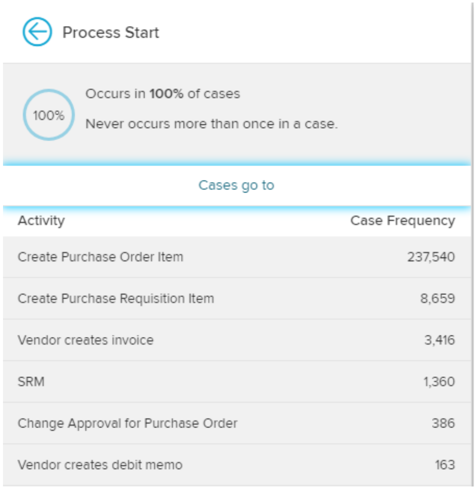


Fig. 11. The snapshot of the list of the process start activities.

Invoice clearing cycle. The flow between Record Invoice Receipt and Clear Invoice which is the final stage in the process is analyzed next. An average throughput time between those two equals 36 days, which is considered to be a long time in the context of the rest of connections. Remove Payment Block is an activity that appears in the process quite often (22% of all cases) and is quite understandable activity in the *3-way match, invoice before GR*, when the invoice is matched and the payment is set free to be transferred (which is at the same time the most often used matching model). The long throughput time in this cycle can be explained by the profits from the delay in the payment for the company. Postponing the payments to the maximum possible date allows to withhold the capital in the company. Fig. 12 presents invoicing cycle of the process.

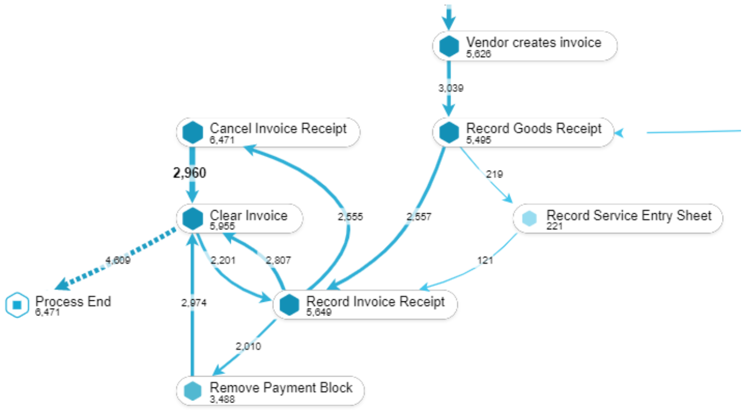


Fig. 12. The part of the process flow presenting the invoicing stage, covering 99% of activities and 56% of all connections.

Cancel Invoice Receipt is seen as a definite deviation. While exploring its connection it is concluded that after a cancellation of an invoice it proceeds with clearing (2,960 cases, that is 46% of all cancelled invoices) and later on new invoice is recorded. On the other hand the cases, which after cancellation of an invoice proceed with recording invoice receipt amount to 26% of all cancelled invoices. This sequence indicates that the invoices must be wrongly posted and new ones are issued by the vendor in order to correct the issue and finalize the purchase process. Performing further analysis on the issue it was discovered that in 99% of the cases, where invoice is cancelled and a new one is recorded, a Debit Memo was created upfront by the vendor. Debit Memo is thought to be a correcting, invoice based document, whose presence in the process justifies the cancellation of an invoice. Cancelling an invoice is done together with recording a new one and this cycle takes around 24 days. More governance of debit memos issued by the vendors in the process would bring a benefit of speeding up the process of multiple invoices handling and realization of payment in consequence.

Rework activities. Significant problems take place on the second stage of the process, namely between Create Purchase Order Item and Record Goods Receipt. Apart from positive activities, such as Vendor Creates Invoice mentioned before and Receive Order Confirmation, some unnecessary changes are applied. Those changes concern mostly quantity and price, which prolong the whole process by 7 days on average each and their ratios are 1.34% of all cases for quantity change, and 0.78% for change price. The overall rework ratio equals 2.15% and its scale is not alarming. Although the presence of rework is considered to be negative in principle, in a real-life they are hard to avoid. Changes in orders are considered as deviations, which could be avoided if orders were made with more care and closer cooperation with suppliers.

Change Quantity activity appears in 17,590 cases and the median throughput time for the whole process (incorporating Change Quantity step) is 18 days longer, than the initial one. Examining the root cause for presence of this rework in the process, it appeared they only touch companyID_0000. All of the change quantity actions are performed by manual users. More than half of purchases are done in Packaging spend area, especially in cooperation with vendorID_136. Moreover, Trading & End Products spend area seems to incorporate a lot of quantity change rework, where vendorID_0197 contributes to this issue the most. Fig. 13 presents the process map of vendorID_0197.

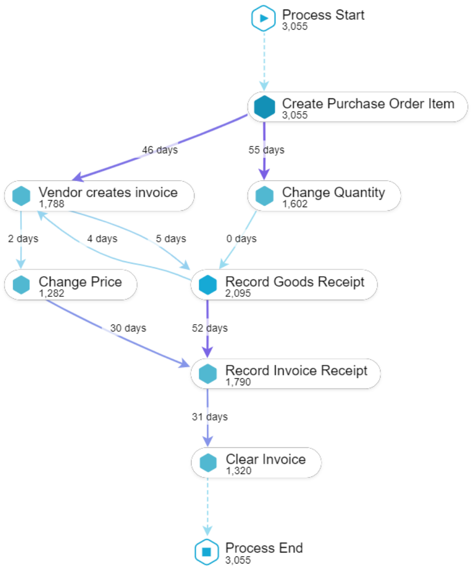


Fig. 13. Process map for vendorID_0197, as a supplier causing rework issues and long throughput times.

The median throughput time for supplier vendorID_0197 takes 133 days (versus 72 in general). One of the main problems in a trade with this vendor is a long throughput time between PO creation and the next step e.g. Vendor Creates Invoice (46 days). Also it takes long until goods are received (52 days). Along with quantity changes there are a lot of price changes applied (almost half of the POs). Users working with it are mostly user_071 (performing change price rework) and user_84 (change price rework).

Change Price rework appears in 4% of the cases. It happens in almost all of the cases within one subsidiary, namely company_0000 and they also are performed only by manual users. In the case of change price rework the most problems are caused by vendorID_0197, presented in the previous paragraph. No other vendor seems to contribute strongly to its presence. There exists a correlation between Change Price and Change Quantity. In 20% of cases, where there is Change Price also Change Quantity is present in the process.

Delete Purchase Order Item happens also on that stage of a process. It is natural that such cases as deletion of an order will sometimes take place at some point in purchasing. Due to the fact that it happens right after the creation of an item it does not contribute to growing problems, as it could if it happened on the latter stages. It can be treated as fairly acceptable then. Although aiming at the process excellence, if deletion of order items happen, an extra governance over their creation should be given.

Overall, more governance and control over purchase requesting would bring profits on the coherence of the process. Additionally better communication and closer cooperation with vendors would be profitable and let the company avoid unnecessary changes at different stages during the flow of the process.

Purchased services. Record Service Entry Sheet activity is present in 2% cases of the process. Service entry sheet is a corresponding document to goods receipt, which serves the same purpose just for services not goods. Hence, there is nothing surprising in terms of appearance of this activity in the process flow. It corresponds just to items from *Service* category and in nearly 94% *Logistics* spend area (with the focus on *Road Packed* sub area). Main vendors providing the services are vendorID_0234 and vendorID_0230. The median throughput time of processing services is relatively short and equals 27 days.

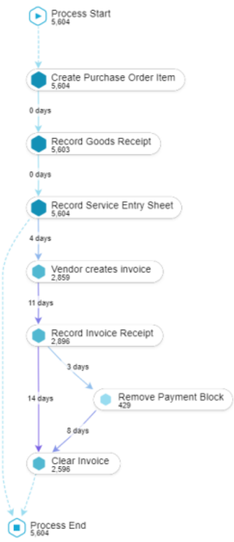


Fig. 14. The process flow map for services.

Fig. 14 captures the process map for services. What is attracting attention in its process flow is that it happens along with Record Goods Receipt activity and additionally at the same time, when the Purchase Order Item is created (0 days of throughput time between activities). This means the PO is created in the moment when the service is provided, not in advance.

3.3 Predictive analysis

In the previous parts of the report the process itself was examined and the most significant anomalies and deviations were identified. The purpose of predictive analysis, however, goes beyond the possibilities of exploratory and process focused analysis and enables to model and optimize corporate operations in order to maximize profits or minimize risks [6].

The aim of predictive analysis performed in the given project is to decide whether certain activities within the process path have significant impact on the overall process throughput time. It is reasonable to assume that different kinds of orders are being processed in different ways and potential delay causes may differ from one another [7]. All orders were clustered into three main groups and for each of these groups a separate delay analysis was performed. Undesired activities with the greatest impact on overall process time were identified and recommendations on which order elements should be most carefully checked were formulated. The process of identifying most troublesome activities is crucial in optimizing process throughput time and avoiding significant costs caused by delays [7]. An example of analysis' application was also described in the current chapter: from the examined dataset all of the unfinished processes were identified and classified into one of the clusters. It enabled to formulate recommendations for the most recent purchase order items.

In order to achieve the goal described above, it was necessary to prepare the purchase to pay process dataset for further analysis. Each purchase order item has its own characteristics, such as *case_spend_area*, *case_item_type* or *case_vendor*. These were extracted from the original dataset and used as clustering categories later on. In Fig. 15 there are presented exemplary cases with all of the relevant categorizing columns.

_case_key	c_spendarea	c_subspendarea	c_company	c_doctype	c_vendor	c_itemtype	c_itemcat
2000000003-00002	Enterprise Services	Office Supplies	companyID_0000	EC Purchase order	vendorID_0003	Standard	3-way match, invoice before GR
2000000009-00001	CAPEX & SOCS	Laboratory Supplies & Services	companyID_0000	EC Purchase order	vendorID_0008	Service	3-way match, invoice after GR
2000000015-00001	Marketing	Digital Marketing	companyID_0000	EC Purchase order	vendorID_0014	Service	3-way match, invoice after GR

Fig. 15. Exemplary rows of the dataset used in purchase order items clustering.

What is more, there was created an additional pivot table consisting of activities' names in columns and dummy values *1* and *0* for activities that were conducted or not, respectively, within specific purchase processes. Additionally, another data transformation was performed in order to calculate for each purchase order item, purchasing duration time expressed in days, purchase order item cumulative net worth, number of distinct activities, automation rate and number of distinct users involved in particular purchase process. Exemplary cases, after transformations described above, are shown in Fig. 16.

_case_key	BlockPurchaseOrderItem	CancelGoodsReceipt	CancelInvoiceReceipt	CancelSubsequentInvoice
2000000003-00002	0	0	0	0
2000000009-00001	0	0	0	0
2000000015-00001	0	0	1	0

_case_key	case_days_no	net_worth	case_activities_distinct	case_days_avg	case_automation	case_users
2000000003-00002	27	37.00	12	3	0.58	5
2000000009-00001	39	1039.00	13	3	0.54	5
2000000015-00001	256	457836.00	18	3	0.12	11

Fig. 16. Exemplary cases of the dataset used in modelling purchase process time.

The aim is to identify activities causing the most significant process delays. It entails incorporating division between activities that are accepted as a part of the purchase process (e.g. *Clear Invoice*) and activities that might and should be avoided (e.g. *Change Quantity*); there are also activities which prolong the process but exist only as a consequence of other activities (e.g. *Reactivate Purchase Order Item*). Therefore all of the activities were divided into categories of Accepted, Undesired or Immaterial. Table 2 presents activities recognized as Undesired.

Table 2. Activities recognized as Undesired in the purchase to pay process.

Undesired Activities in P2P Process		
Block Purchase Order Item	Change Currency	Change Storage Location
Cancel Goods Receipt	Change payment term	Delete Purchase Order Item
Cancel Invoice Receipt	Change Price	Record Subsequent Invoice
Change Approval for Purchase Order	Change Quantity	Update Order Confirmation

The analysis is conducted for purchases taking place in 2018, therefore the dataset consists of completed purchases as well as the ones which by the end of the year (to be specific, until 27th Jan 2019, please see detailed description in Chapter 2.3) were still being processed. The aim of this chapter is to indicate the length of unfinished purchases and to betoken most significant alterations. In order to achieve that the whole dataset (251,270 cases) was divided into a train set on which the models were built and a predictive set for which model predictions were concluded. The train set consists of completed cases (191,971 rows) and the predictive set consists of unfinished ones (59,299 rows).

Machine learning methods were applied for obtaining predictive analysis results. First, all of the completed purchases were divided into groups using clustering, namely, K-Modes clustering. Three clusters were distinguished and in the second part of the analysis three linear regression models were build, one for each cluster. Purchase process throughput time was estimated using Ordinary Least Square method. The technologies used were Scikit-learn, Statsmodels, PyPI and SciPy open source libraries written in Python.

Purchase Order items clustering. Clustering is an unsupervised machine learning approach used in variety of cases, ranging from customer segmentation to anomaly detection [8]. It aims at finding similar objects in one cluster and dissimilar objects far from one another. Clustering can be done in multiple ways based on the type of data and business environment. In this analysis dataset contains categorical information about purchase order items (spending area, spending subarea, company name, document type, vendor name, item type, item category) therefore clustering K-modes method was applied. Rather than calculating the distance between any two observations, it counts occurrences of the same values and clusters the most similar cases.

The summary of each case category is presented in Table 14. It may be noticed that there are no missing values which ensures that clustering is conducted properly.

Table 3. Purchase order items' categories summary.

Category	Count	Unique	Top	Frequency
Spend area	191,971	21	Packaging	80,763
Spend subarea	191,971	134	Products for Resale	53,258
Company	191,971	4	companyID_0000	191,676
Document type	191,971	3	Standard PO	190,125
Vendor	191,971	1,680	vendorID_0136	10,945
Item type	191,971	6	Standard	181,926
Item category	191,971	4	3-way match, invoice before GR	181,311

Based on elbow method results, clusters' number was set to three. Number of times the K-Modes algorithm is being run was set to five and Huang initialization method was chosen [9][10].

The result of clustering are three order items' groups (clusters). The most frequently occurring values in each cluster's categories, called *centroids*, are presented in Table 15. One may notice that differences between centroids appear between vendors, spending areas and spending subareas. In each cluster, however, most numerous company, document type, item type and item category are the same.

Table 4. Most frequently occurring values in each order items' cluster (centroids).

Category	Cluster A (54,538 order items)	Cluster B (80,818 order items)	Cluster C (56,615 order items)
Spend area	Trading & End Products	Packaging	Sales
Spend subarea	Trading products (old structure)	Labels	Products for Resale
Company	companyID_0000	companyID_0000	companyID_0000
Document type	Standard PO	Standard PO	Standard PO
Vendor	vendorID_0118	vendorID_0136	vendorID_0108
Item type	Standard	Standard	Standard
Item category	3-way match, invoice before GR	3-way match, invoice before GR	3-way match, invoice before GR

Linear regressions for purchase order items' clusters. In order to establish which activities cause the longest delays, linear regression on each cluster was applied. All three models were built with respect to linear regression requirements and initial tests, that is, no significant variables (factors) in explaining variability of purchase process duration are omitted, all of the insignificant variables are eliminated and that the variables are not correlated [11]. The initial version of each model contained a dependent variable *case_days_number* and 37 explanatory variables, to be specific, 5 numerical variables and 32 dummy variables indicating whether an activity was present in the particular purchase process. Insignificant explanatory variables were then gradually removed from the models. It enabled to establish final version of each model and to specify which significant explanatory variables are undesired activities listed in Table 13. The activities with the greatest impact on overall purchase process time, differing between individual clusters, are listed below in Table 16.

Table 5. Undesired activities with the greatest impact on overall process time. Linear regression.

No.	Cluster A	Cluster B	Cluster C
1.	Record Subsequent Invoice [19.3 days]	Change Approval for Purchase Order [26.1 days]	Change Payment Term [24.4 days]
2.	Change Quantity [17.8 days]	Record Subsequent Invoice [20.3 days]	Record Subsequent Invoice [10.9 days]
3.	Cancel Goods Receipt [17.4 days]	Block Purchase Order Item [15.5 days]	Cancel Invoice Receipt [6.8 days]

Random Forests regression for purchase order items' clusters. First approach to identify activities with significant impact on purchase to pay process throughput time was to build linear regression models. Second approach, however, was based on Random Forest Regressor method. It was run for each cluster separately and enabled to compare and verify results obtained earlier.

Each of the three datasets containing cases from three different clusters were randomly divided into train sets (70% of all cases from each cluster) and test sets (the remaining 30%). Random forests were built based on train sets' data and the results were cross validated on test sets' data, which enabled to control accuracy of the models. What is more, it was quantified how much including a particular variable improves the entire random forest prediction and therefore undesired activities with the greatest impact on the overall process time were identified in a new way. The results are presented in Table 6.

Table 6. Undesired activities with the greatest impact on overall process time. Random forests.

Cluster A	Cluster B	Cluster C
Change Price Change Quantity	Change Approval for Purchase Order	Change Quantity

Clustering of predictive set's data. The analysis of completed orders aimed to identify potential threats to an optimum purchase to pay process flow and was performed for completed orders. Nonetheless, by the end of 2018 some of purchase orders were still being processed. Based on K-Modes method it was possible not only to cluster completed processes, as described above, but also to classify the unfinished ones. The classification was conducted in accordance with individual order items' characteristics into clusters defined above and enabled to predict the most possible process interruptions. Out of 59,299 order items the majority (54,750 cases) was identified to belong to Cluster A, whereas 4,384 cases were assigned to Cluster C and 165 to Cluster B.

Prediction analysis results and recommendations. Clustering on predictive set helped to understand that the majority of uncompleted, by the end of 2018, processes revealed close similarity to completed processes from Cluster A. To be specific, their main spending area and subarea is *Trading & End Products*, *Trading products (old structure)*, the main subsidiary involved is *companyID_0000* and the main vendor is *vendorID_0118*. Two independent models built on the train set indicate that the most significant delays in process' total throughput time will be caused by quantity change. Additional problematic activities are price changes, subsequent invoice records and goods receipt cancellations. It is recommended to thoroughly check the initial orders for the correct quantity and price as well as the justifiability for placing these orders. If there is more governance introduced, it saves hours of manual users' work later on.

In predictive analysis *vendor_0118* was identified as the main vendor of most numerous Cluster A items, whereas in process analysis it was pointed out that *vendor_0917* was the one with the biggest number of quantity changes in year 2018. Combining the conclusions from both approaches, it is advisable for the process owner to engage in delay causes detection with these two main vendors. Such cooperation may result in significant improvements in the purchase to pay process flow.

It is worth mentioning that the characteristics of order items with quantity changes are partly consistent with process analysis performed above. However, it widens the scope of significant vendors engaged in the activities and enhances the importance of some threats for the nearest future. If the most current orders were placed on different items, different undesired activities would be identified in the presented predictive analysis as the most urgent. Nowadays, an ability to dynamically foresee process threats using artificial intelligence is an area where companies gain more and more competitive advantage.

4 SUMMARY

In the digitalization era, the possibilities of implementing IT systems supporting Process Mining techniques are constantly growing. Among different business processes Purchase-to-Pay process has a lot of potential for an improvement. In the presented report a thorough analysis was performed. With the usage of Power BI data visualization tool exploratory analysis was conducted. Further on, process focused analysis with the usage of Celonis Process Intelligence tool was described. The last part of the report, carried out with the use of Python Machine Learning libraries, was predictive analysis and recommendations.

In exploratory analysis four different invoice matching models were visualized and described. For every invoicing model (*3-way match, invoice before GR; 3-way match, invoice after GR; Consignment; 2-way match*) their characteristics were briefly described, including spend areas and subareas, main vendors, seasonality and the most common rework activities. The analysis enabled to learn main purchase to pay process features, recognize differences and became useful in process focused analysis conducted in the next chapter.

The main goal of process focused analysis was to describe process flow and to identify process deviations. First, so-called happy path of the process was discovered and further on its violations were identified. Possible explanations, root causes and alarming points, derived from detailed analysis supported with Celonis application, were also provided.

Predictive analysis aimed at identifying undesired activities with the greatest impact on overall process throughput time and predicting the most possible threats for the upcoming orders. All of the completed order items were divided into three clusters, with order items' categories different from one another as much as possible. For these three clusters separate models were built and most influential (that is, causing the longest delays) activities were identified. It enabled to recommend the most important precautions against purchase to pay process delays for the orders which were already initialized but not yet completed.

In summary, the analysis presented significant insights into Purchase-to-Pay process characteristics, identified process deviations and provided useful recommendations. If data provided is de-anonymized, the process owner may apply given results and improve Purchase-to-Pay process flow which will result in significant cost reduction. Also, in the future, more accurate conclusions may be obtained with usage of wider range of data, more specific analytical questions and possibility of communication between analysts and process owners.

5 REFERENCES

1. Van der Aalst, W.: Process Mining. Discovery, conformance and Enhancement of Business Processes, Springer-Verlag Berlin Heidelberg (2011).
2. Sadiq, S., Governatori, G., Namiri, K.: Modeling Control Objectives for Business Process Compliance, BPM 2007. Lecture Notes in Computer Science, vol 4714. Springer, Berlin, Heidelberg (2007).
3. BPI Challenge 2019, 9th International Business Process Intelligence Challenge https://icpmconference.org/icpm-2019/contests-challenges/bpi-challenge-2019/, last accessed 2019/04/17.
4. Kraljic, P.: Purchasing must become supply management. Harvard Business Review (1983).
5. Bhuvaneswaran, Sh.: The Ultimate Guide to a Truly Effective Procure-to-Pay Process, https://kissflow.com/procurement-process/guide-to-procure-to-pay-process/, last accessed 2019/04/17.
6. Murphy, K.: 3-way matching process. Why it's a critical Step in the Accounts Payable Process, PurchaseControl (2017), https://www.purchasecontrol.com/uk/blog/what-is-3-way-matching/.
7. Van Dongen, B.F., Dataset BPI Challenge 2019. 4TU.Centre for Research Data. https://doi.org/10.4121/uuid:d06aff4b-79f0-45e6-8ec8-e19730c248f1, last accessed 2019/04/17.
8. Van der Aalst W., Reijers H., Weijters A., Van Dongen B., Alves de Medeiros A., Song M., Verbeek H.: Business Process Mining: An Industrial Application, Information Systems, 32(5), pp. 713-732.
9. Van der Aalst W., Schonenberg M., Song M.: Time Prediction Based on Process Mining, BPM Center Report BPM-09-04 (2009), BPMcenter.org.
10. Singh P.: Machine Learning with PySpark: With Natural Language Processing and Recommender Systems, Apress (2018).
11. Huang, Z.: Clustering large data sets with mixed numeric and categorical values, Proceedings of the First Pacific Asia Knowledge Discovery and Data Mining Conference, Singapore (1997), pp. 21-34.
12. Huang, Z.: Extensions to the k-modes algorithm for clustering large data sets with categorical values, Data Mining and Knowledge Discovery (1998), 2(3), pp. 283-304.
13. Mycielski J.: Econometrics, University of Warsaw (2009).

6 APPENDIX


BPI Challenge 2019
Purchase-to-Pay
process
Management summary
PwC RAS Data Analytics
Warsaw, Poland
06.05.2019
pwc


1
Case overview

Case overview

BPI Challenge 2019 focuses on analysis of a Purchase-to-Pay process within an international company operating from The Netherlands.
4 main invoicing models were examined using three perspectives: exploratory, process focused and predictive analysis.

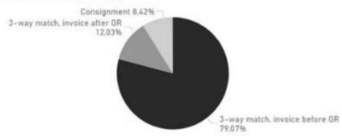
Dataset

- Over 1.5 milion events
- Timeframe capturing purchasing in year 2018
- ~251 thousand of purchase order items
- 4 subsidiaries of the company performing purchases
- Almost 2 thousand vendors
- Purchasing of the worth of more than 1 bilion €
- 4 invoicing models

Approach

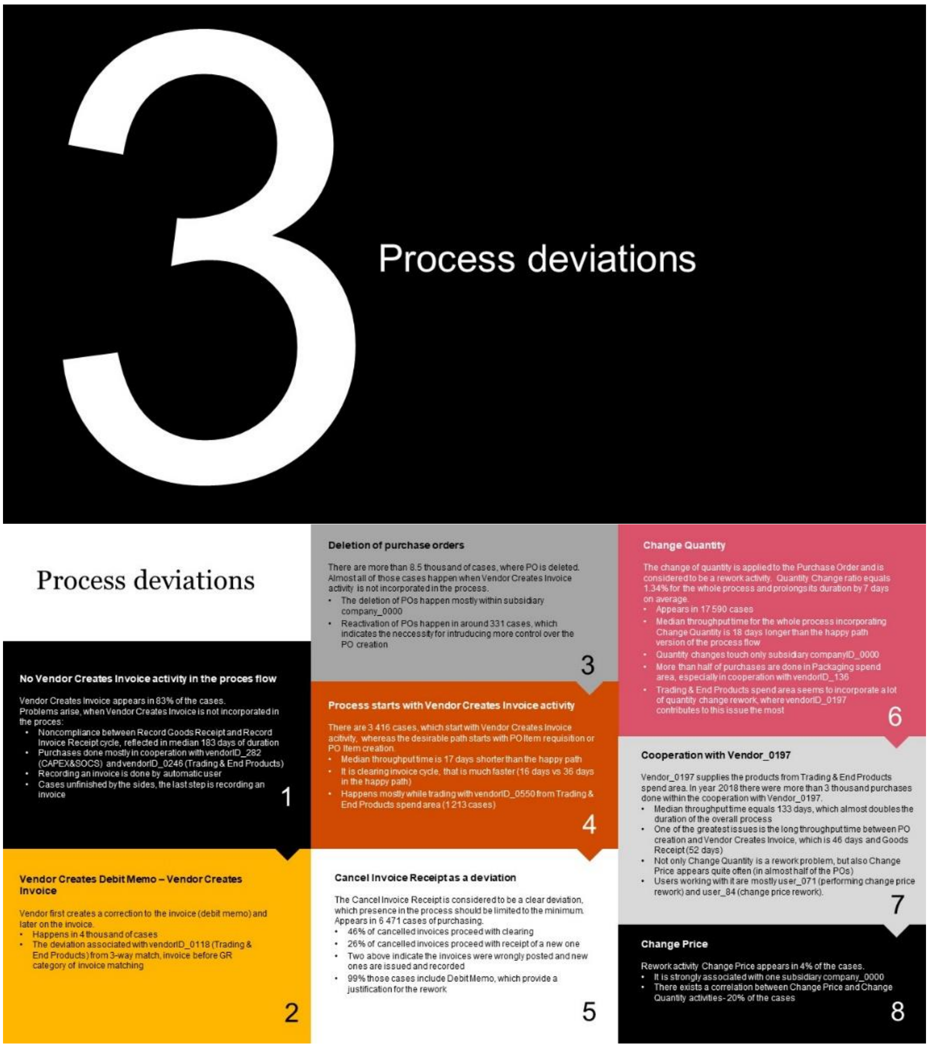
- **Process exploration using Microsoft PowerBI**
- **Deviations investigation with Celonis**
- **Predictive techniques application in Python**

#PO per Invoicing Category

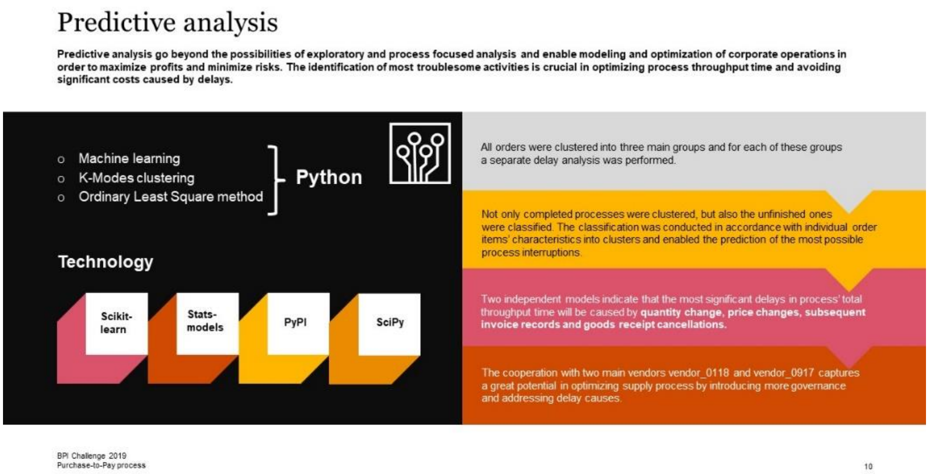


2 Invoicing models overview


3-way match, invoice before GR
The value of the goods receipt is matched against the value of an invoice corresponding to it, as well as purchase order. Before the invoice is paid, accounts payable reviews purchase order, matches it with the received goods and invoice to pay. If all documents match, an invoice is payed and cleared.
1. Main characteristics
Most common purchasing model ~79% of all the purchase orders
€784.90M in 2018
62k of PO in 2018
98% of Standard PO type
Less purchases in the 1st quater of the year
Average item price €317
Areas: Packaging, Sales, Logistics, Additives
Rework ratio equals 0.69%, half of which builds up change price (0.37%) having impact of 5-6 days of delay in the process
Change quantity in 0.31% of the cases
3-way match, invoice after GR
The main difference to the previous invoicing model is that purchase items do not require a Goods-Receipt-Invoicing. An invoice may be entered before goods are received, although the payment is done after receiving a goods receipt and matching the documents.
1. Main characteristics
~12% of all the purchase orders
€234.12M in 2018
9.5k of PO in 2018
Average item price around €742
POs undiversified- 1.6 distnct items per PO
Areas: Road Packed (Logistics), Digital Marketing
High automation rate, more than 20%
Developed SRM activities net
Overall rework ratio equals 0.69%, half of which builds up change price (0.37%)
Change quantity in 0.92% of the cases, prolonging the process up to 7 days on average (user_071)
2-way match
In this type of a model only 2 documents must match, a purchase order and an invoice. What is important, is that a value of a purchase order might be consumed by multiple invoices, therefore a one-to-one match is not a must.
1. Main characteristics
Least popular invoicing model (around 1% of all cases)
€12.05M in 2018
373 of PO in 2018
Average item price €2.53k
One susbsidiary company_003
Purchases happening mainly in March/April, November/ December 2018
Item type Limit
One user user_602
Areas: Real Estate, Government Payments, Taxation
Consignment
A consignment is a business arrangement in which goods are left in the possession of an authorized third party to sell, whereas the ownership stays with a vendor. Due to the specific nature of this matching model it is handled in a fully separate process. There are no invoices registered in this event log and the analysis is conducted for the first stages of purchasing.
1. Main characteristics
Cover around 8.5% of all the cases
Zero values for PO
More than 6.6k of PO in 2018
One susbsidiary company_0000
Item type Consignment
One user: user_602
Areas: Packaging, Additives, Latex&Monomers
Change quantity in 3.71% of the cases, prolonging the process by 5 days on average


3
Process deviations
Process deviations
No Vendor Creates Invoice activity in the proces flow
Vendor Creates Invoice appears in 83% of the cases.
Problems arise, when Vendor Creates Invoice is not incorporated in the proces:
• Noncompliance between Record Goods Receipt and Record Invoice Receipt cycle, reflected in median 183 days of duration
• Purchases done mostly in cooperation with vendorID_282 (CAPEX&SOCS) and vendorID_0246 (Trading & End Products)
• Recording an invoice is done by automatic user
• Cases unfinished by the sides, the last step is recording an invoice
1
Vendor Creates Debit Memo – Vendor Creates Invoice
Vendor first creates a correction to the invoice (debit memo) and later on the invoice.
• Happens in 4 thousand of cases
• The deviation associated with vendorID_0118 (Trading & End Products) from 3-way match, invoice before GR category of invoice matching
2
Deletion of purchase orders
There are more than 8.5 thousand of cases, where PO is deleted. Almost all of those cases happen when Vendor Creates Invoice activity is not incorporated in the process.
• The deletion of POs happen mostly within subsidiary company_0000
• Reactivation of POs happen in around 331 cases, which indicates the neccessity for intruducing more control over the PO creation
3
Process starts with Vendor Creates Invoice activity
There are 3 416 cases, which start with Vendor Creates Invoice acitvity, whereas the desirable path starts with PO Item requisition or PO Item creation.
• Median throughput time is 17 days shorter than the happy path
• It is clearing invoice cycle, that is much faster (16 days vs 36 days in the happy path)
• Happens mostly while trading with vendorID_0550 from Trading & End Products spend area (1 213 cases)
4
Cancel Invoice Receipt as a deviation
The Cancel Invoice Receipt is considered to be a clear deviation, which presence in the process should be limited to the minimum. Appears in 6 471 cases of purchasing.
• 46% of cancelled invoices proceed with clearing
• 26% of cancelled invoices proceed with receipt of a new one
• Two above indicate the invoices were wrongly posted and new ones are issued and recorded
• 99% those cases include Debit Memo, which provide a justification for the rework
5
Change Quantity
The change of quantity is applied to the Purchase Order and is considered to be a rework activity. Quantity Change ratio equals 1.34% for the whole process and prolongs its duration by 7 days on average.
• Appears in 17 590 cases
• Median throughput time for the whole process incorporating Change Quantity is 18 days longer than the happy path version of the process flow
• Quantity changes touch only subsidiary companyID_0000
• More than half of purchases are done in Packaging spend area, especially in cooperation with vendorID_136
• Trading & End Products spend area seems to incorporate a lot of quantity change rework, where vendorID_0197 contributes to this issue the most
6
Cooperation with Vendor_0197
Vendor_0197 supplies the products from Trading & End Products spend area. In year 2018 there were more than 3 thousand purchases done within the cooperation with Vendor_0197.
• Median throughput time equals 133 days, which almost doubles the duration of the overall process
• One of the greatest issues is the long throughput time between PO creation and Vendor Creates Invoice, which is 46 days and Goods Receipt (52 days)
• Not only Change Quantity is a rework problem, but also Change Price appears quite often (in almost half of the POs)
• Users working with it are mostly user_071 (performing change price rework) and user_84 (change price rework).
7
Change Price
Rework activity Change Price appears in 4% of the cases.
• It is strongly associated with one subsidiary company_0000
• There exists a correlation between Change Price and Change Quantity activities - 20% of the cases
8


4
Predictive analysis


Predictive analysis
Predictive analysis go beyond the possibilities of exploratory and process focused analysis and enable modeling and optimization of corporate operations in order to maximize profits and minimize risks. The identification of most troublesome activities is crucial in optimizing process throughput time and avoiding significant costs caused by delays.
Machine learning
K-Modes clustering
Ordinary Least Square method
Python
Technology
Scikit-learn
Stats-models
PyPI
SciPy
All orders were clustered into three main groups and for each of these groups a separate delay analysis was performed.
Not only completed processes were clustered, but also the unfinished ones were classified. The classification was conducted in accordance with individual order items' characteristics into clusters and enabled the prediction of the most possible process interruptions.
Two independent models indicate that the most significant delays in process' total throughput time will be caused by quantity change, price changes, subsequent invoice records and goods receipt cancellations.
The cooperation with two main vendors vendor_0118 and vendor_0917 captures a great potential in optimizing supply process by introducing more governance and addressing delay causes.
BPI Challenge 2019
Purchase-to-Pay process
10


Thank you
© 2019 PwC. All rights reserved. Not for further distribution without the permission of PwC. "PwC" refers to the network of member firms of PricewaterhouseCoopers International Limited (PwCIL), or, as the context requires, individual member firms of the PwC network. Each member firm is a separate legal entity and does not act as agent of PwCIL or any other member firm. PwCIL does not provide any services to clients. PwCIL is not responsible or liable for the acts or omissions of any of its member firms nor can it control the exercise of their professional judgment or bind them in any way. No member firm is responsible or liable for the acts or omissions of any other member firm nor can it control the exercise of another member firm's professional judgment or bind another member firm or PwCIL in any way.

Analysis and prediction of purchasing compliance using process mining

Vincent F. A. Meyer zu Wickern, Mounisha Juluru, Anshu Roy, Thi-Thu-Hang Nguyen, and Viet-Hung Vu

Faculty of Communication and Environment, Rhine-Waal University of Applied Sciences, Friedrich-Heinrich-Allee 25, 47475 Kamp-Lintfort, Germany

Abstract. The annual Business Process Intelligence (BPI) challenge is an opportunity to find and apply new techniques for process mining analyses on provided datasets, which in this year 2019 is an event log of purchase-to-pay events of a multinational paints and coatings company. The anonymized process owner is interested in gaining information on the compliance of their process through the discovery of a collection of process models, the analysis of throughput times in their invoicing process and the separation of deviating purchase order items. This paper is focused on the detection of incompliant purchase order items to support the process owner's compliance overview. Compliance of purchasing may be violated through value differences between the company's order items, their goods receipts, and their invoices. Further, the defined process flow may be violated. Multiple process models are discovered belonging to different purchase order item attributes to understand the process landscape of the anonymized company. Additionally, a method to determine incompliant cases is developed. Predictive process mining is used to identify influences of purchase order item attributes, derived additional features and process flow to the probability of breaching compliance. With the consideration of these influences, more incompliant process executions may be detected or prevented.

Keywords: Process Mining · Three-Way Matching · Purchasing Compliance · Predictive Process Monitoring.

1 Introduction

The Business Process Intelligence (BPI) Challenge 2019 is a challenge in the field of process mining related to the International Conference on Process Mining (ICPM) conference which is to be held in Aachen, Germany [5]. The BPI Challenge 2019 requests all participants to analyze an event log of a purchase-to-pay process provided by an anonymized process owner. The process owner is especially interested in insights into their compliance. The BPI Challenge is split into student and non-student submissions, and this paper strives to analyze the given dataset on compliance as a student submission. For this, the major influencing factors on purchasing compliance and an opportunity to improve it were sought after. Predictive process mining was used to identify attributes of the purchase order (PO) items, the additional derived features, and the process flow, which increase the probability of breaching compliance.

The following tasks should be answered to support the process owner in the procurement compliance improvement as defined in the challenge description [5]:

- Find one or more process models that could accurately describe the processes in the event log.
- Find the throughput time of the invoice process and suggestions to mitigate the bottlenecks.
- Find unusual activities and documents in the event log.

These questions are expected to be solved with the help of process mining. Process mining is a research discipline that aims to discover, monitor, and improve real processes by extracting knowledge from event logs that are available in organizations' information systems [3]. For this paper, multiple machine learning techniques were used to support the analysis of the company's purchasing compliance and to potentially further leverage machine learning techniques in process analyses. The main goal of the paper is to find correlating attributes in the given event log that may provide insights into the occurrence of incompliant cases. Predictive process mining was therefore used to classify PO items based on their compliance and find predictors for incompliant PO items both with the complete knowledge of the event log and with the restriction of knowing only information up to a given event time in a PO item process.

As the quality of the prediction may be influenced by a reasonable preprocessing of the dataset, the dataset was first explored for exceptional and potentially unusable data points in section 2. An understanding of the company's processes may additionally be beneficial for a classifier's success, so the processes as seen in the data were visualized. Trace clustering techniques were used in section 3 to identify different process areas and thereby enhance the simplicity and precision of the models. A calculation technique for the second request of the company regarding throughput time is provided in section 4, followed by the creation of additional dimensions in section 5, in which deviating PO items were identified. The key focus in the detection and analysis of deviating PO items was laid upon incompliant PO items and PO items with rework. The classifiers to predict incompliant behavior are further described in section 5.

2 Dataset exploration and preprocessing

The provided anonymized data consists of a list of PO item events submitted in 2018, which were collected from a coating-and-paints company in the Netherlands [5]. From the dataset, each PO item (or case) can be uniquely identified using the *_case_concept_name_* column representing the concatenated PO number and the line item number [5]. There can be multiple events (or activities) from one PO item, each of which has its own event name (*_event_concept_name_* column) and timestamp (*_event_time_timestamp_* column). In general, the data is split into attributes that are valid for an entire case, i.e., a PO item, and attributes that are valid only for single events, i.e., the activities that were performed. The PO items could be pre-classified into four main categories as defined by the BPI challenge:

- 3-way match, invoice before GR
- 3-way match, invoice after GR
- 2-way match
- Consignment

Not all of the presented events could be used for the analyses in this paper, because they may not be comparable. The lack of comparability may be due to the following reasons, which are taken into account in this chapter:

1. Events did not occur in the extraction timeframe
2. Cases are not complete

A special focus of the analyses in this pager is set on the compliance of cases, which is not provided in the BPI challenge dataset. Therefore the specifications of compliance used in this paper are explained in this chapter.

2.1 Timeframe exclusion

The BPI challenge description defines that there are "over 1.5 million events for purchase orders submitted in 2018" in the dataset [5]. Therefore the validity of PO items created before or after this timeframe could not be ensured, and the dataset needed to be filtered on the timeframe for further processing of the data. As the data was submitted on 28.01.2019, this was the final date that was accepted in the preprocessing of the event log. The earliest point that was accepted in the data was the beginning of 01.01.2018.

2.2 Compliance of cases

As the process owner has compliance questions, compliance is critical in the analyses created in this paper. To verify whether a PO item is compliant, the necessary activities have have been reviewed, and the monetary values set in the events had to be checked. For this, first, a compliance check was made verifying all possible compliance criteria. In the second step, the non-compliant PO items, which have not been completed yet, were separated from full cases that are non-compliant due to a mismatch in 3-way matching or 2-way matching. The word “incompliant” does not necessarily indicate any mistakes made in the process, because also ongoing and therefore incomplete cases have to be regarded as incompliant, e.g., a PO item without payment cannot be seen as compliant. Therefore, the split into complete and incomplete cases is necessary.

Verification of compliance
To review the compliance of the PO items, the monetary values, and counts of PO items, goods receipts (GRs), and invoice receipts were checked. Both the

counts and monetary values should be equal for the PO items to be compliant. However, the case item category has to be considered because not for all case item categories, invoices and GRs can be expected. Since the invoicing process for consignment order items is not conducted in the observed system, no invoices or clearing of these invoices can be expected. For items of the category *2-way match*, no GRs are required, therefore also PO items without GR have to be regarded as compliant. The count of GRs and invoice receipts were reduced if cancellations were recorded. Because the purchase-to-pay process requires a payment to meet the customer's obligation, the *Clear Invoice* activity was seen as necessary for compliance. Consequently, the rules presented in the following were applied to categorize incompliant PO items in relation to their item category:

All, except *2-way match*: Number of GRs > 0

All, except *Consignment*:

- Number of *Clear Invoice* events > 0
- Number of invoice receipts > 0

Consignment: Cumulated value of the GRs divided by the number of GRs = PO item value

3-way match, invoice after GR, *3-way match, invoice before GR*:

- PO item value = Cumulated value of the invoice receipts divided by the number of invoice receipts
- Cumulated value of the GRs = Cumulated value of the invoice receipts
- Number of GRs = Number of invoice receipts

Incomplete PO Items

A subset of the incompliant PO items are the incomplete cases. These are not incompliant due to any wrong entries or behavior, but only because they have not yet been finished. Inclusion of these items in defining criteria that lead to incompliant cases may, therefore, lead to incorrect findings, and the items have to be excluded for such an investigation. Incomplete PO items can be recognized by their end activities, which is assumed to be *Clear invoice* for all cases except *Consignment* PO items, which do not involve an invoicing process. The recording of a GR is assumed to be the end activity for *Consignment* PO items. Furthermore, more than one GR and invoice receipt are possible in the purchasing process, which may also lead to multiple clearings of invoices. In *3-way match: invoice after goods receipt* order items, invoices have to follow goods receipts, so incomplete orders are those, in which fewer invoice receipts were received than goods receipt, reduced by cancellations. For *3-way match: invoice before goods receipt* order items, the order of GRs and invoice receipts is insignificant, so both a lower count of goods receipts or invoice receipts compared to the respective other implies an incomplete PO item.

Figure 1 presents an overview of the four main clusters with the number of compliant and complete PO items within them. It is evident that all incomplete

PO items are also incompliant, but that complete PO items are split into compliant and incompliant ones. Especially for *2-way match* PO items, there are many incomplete cases in the dataset and the number of these even surpasses the count of complete cases. These cases lack payments and can therefore not be seen as complete. Considering that *Limit* PO items, which these *2-way match* items are, are typically open for a long time, this observation reflects their estimated distribution for a one-year extraction limit, in which many PO items may be assumed to still be open. Similarly, *3-way match: Invoice after GR* PO items facilitate the receipt of multiple deliveries with GR-based invoicing and therefore were assumed to have a comparatively longer throughput time than *3-way-match: Invoice before GR*. This is also visible in the ratio of cases that were not completed within the extraction time frame.

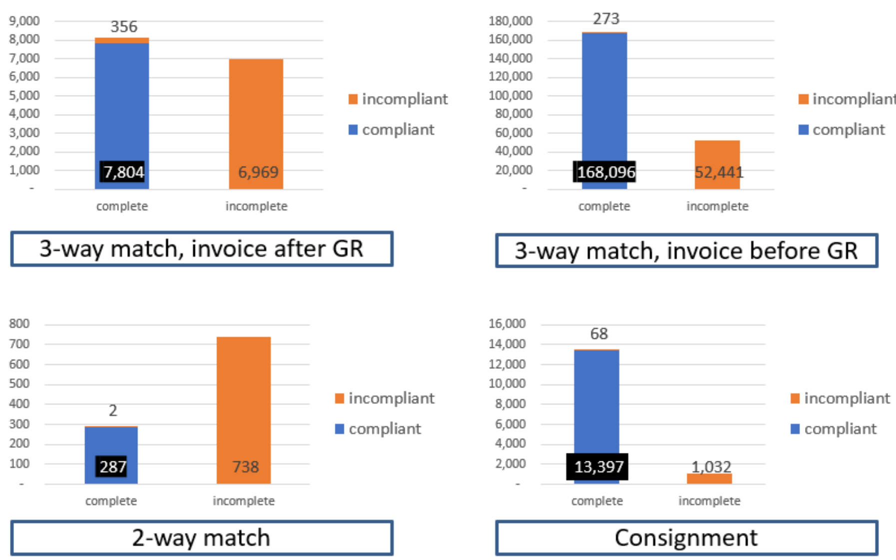


Fig. 1. The four main clusters by compliance and completeness

3 Process models generation

The first main request of the company in focus is the description of its process models that can reflect the current situation of purchasing process execution in the company. Therefore, the analysis in this chapter is directed on the discovery of the as-is process models. Activities, which are not regularly considered part of a standard purchase-to-pay process, such as *Change quantity* or *Change price*, may be part of these models if they are conducted in the majority of cases. This may enable the company to compare these models to the normative models that they have set up. However, to avoid the consideration of a high level of noise and deviations, only compliant cases as defined in chapter 2 were used for determining the process models.

Each generated model should satisfy the four competing success indicators of *fitness*, *simplicity*, *generalization*, and *precision* as described in [4]. Generalization and precision are indicated by the new behavior that is possible with the process model and cannot be observed in the event log. While generalization is positively influenced by enabled additional behavior, the precision indicator is designed to restrict the process model to the observed behavior. As such, in this paper, a high but not maximized precision is aimed for, especially restricting precision when designing a model for a smaller log. Fitness refers to the ability to replay the behavior of the event log in the process model. Fitness of the models in this paper was ensured by allowing for as many activities and connections as necessary to replay the event log. Fitness was measured with alignment-based replay implemented in the *Multi-Perspective Process Explorer* ProM plugin [2] in this paper. This plugin was also used for measuring model precision. Including more connections and activities in the process models may improve model fitness but may damage the simplicity, which, however, can be improved with a reasonable splitting of the dataset to logical sub-groups and with filtering infrequent activities and connections. Therefore, in the following, the grouping of the dataset into reasonable sub-groups is described.

3.1 Split into case groups with similar process flows

The goal for the creation of the models was to have a small number of process models that together can describe the traces encompassed by them. In the first split, the dataset was grouped into four main groups that were already defined in the BPI challenge. In figure 1 in section 2, the counts of these groups can be seen together with their classification on compliance. The process models are based on the filtered data as described in chapter 2 and only compliant cases are considered. The notion of incompliance does not necessarily indicate a wrong behavior in a case, but may also be due to a case that was not conducted until the end and in which for example a payment may be missing.

In the next step, the event log was analyzed for a pattern in the event count and event description. In the event count, it is noticeable that seven activities with the prefix *SRM:* were all executed in 1,119 cases. This led to the assumption that an SRM (Supplier Relationship Management) system was used for a subset of the PO items. In a further analysis of the PO items with SRM activities, it was found that all occur for the document type *EC Purchase Order* and that they were only conducted for *3-way match, invoice before GR-* and *3-way match, invoice after GR*-cases. Since the inclusion of an SRM system indicates other process flows, it was decided to group the event log further into cases including SRM activities and cases without SRM activities.

To analyze, whether other case attributes would enhance the clarity of case process flows for these clusters, the numbers of distinct values for the different case attributes were reviewed. In table 1, these numbers are listed. The first two attributes *_case_Purch__Doc__Category_name_* and *_case_Source_* both have a value of 1, which means that another split on these attributes is not possible. The columns *_case_GR_Based_Inv__Verif__*, *_case_Goods_Receipt_* and *_case_Item_Category_* are all used to build the first level grouping as seen above and would not split the data further so that these attributes could further be neglected. *_case_concept_name_*, *_case_Purchasing_Document_* and *_case_Item_* are identifiers of the POs and their items, so these should also be neglected as split criteria for a broad grouping of purchases. Other attributes, which allow for many distinct values are *_case_Spend_area_text_*, *_case_Sub_spend_area_text_*, *_case_Name_* and *_case_Vendor_* ranging between 21 and 1,552 values. To ensure simplicity in understanding the collection of models, these columns were further neglected due to the high number of process models that would have been generated when splitting by these columns. The columns left to be analyzed were *_case_Company_*, *_case_Document_Type_*, *_case_Spend_classification_text_* and *_case_Item_Type_*.

column name	count of distinct values
_case_Purch__Doc__Category_name_	1
_case_Source_	1
_case_GR_Based_Inv__Verif__	2
_case_Goods_Receipt_	2
_case_Company_	3
_case_Document_Type_	3
_case_Item_Category_	4
_case_Spend_classification_text_	4
_case_Item_Type_	6
_case_Spend_area_text_	21
_case_Sub_spend_area_text_	133
_case_Item_	318
_case_Name_	1487
_case_Vendor_	1552
_case_Purchasing_Document_	55994
_case_concept_name_	189584

Table 1. Count of distinct values per case column

The dissimilarity within each generated group was analyzed to verify whether a split by these columns would add more precise process models than with a broader dataset. If the sum of the within-sum-of-squares (WSS) was significantly lower in the split groups than in the overall dataset, a split by these groups would be reasonable. A simple version of trace profiling was used to measure the dissimilarity of cases within a cluster. For each case, the number of occurrences for each activity was counted, and to measure the dissimilarity between two cases, the Euclidian distance between two activity count vectors was calculated. To measure the WSS of one cluster, the dissimilarity of each case vector to the center of this cluster was added to one sum. For the comparison to the dissimilarity of only four groups by item category, all WSS values for one split criterion, e.g., document type or company, were added. In figure 2, the WSS values for all split criteria are displayed.

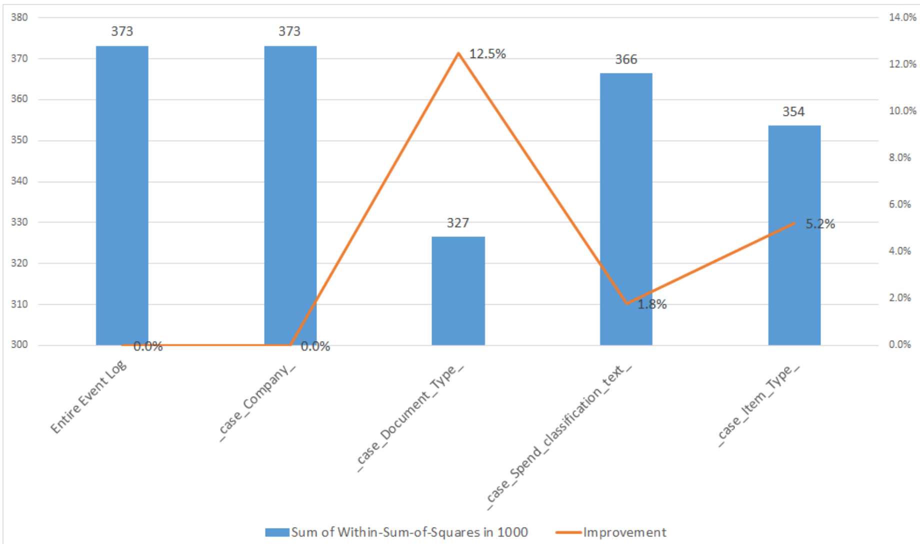


Fig. 2. Comparison of the inner cluster dissimilarity by case criteria

With figure 2, it becomes evident that the case attributes *_case_Document_Type_* with 12.5% improvement and *_case_Item_Type* with 5.2% improvement achieve the highest differences in WSS values. The difference induced by the split by *_case_Document_Type_* matches the above-explained difference of process flows through the inclusion of an SRM system in EC Purchase Orders. Based on these split criteria, the process models were created. When the process models were generated, it was found that the datasets for framework orders and standard POs for *Service* items in the category *3-way match, invoice after GR* yielded the same process models, hence these were grouped. Since a difference between framework orders and standard POs then did not lead to splitting, the split of document types was changed to a split into PO items with SRM activities and PO items without SRM activities. Similarly, the groups of Subcontracting and Third-Party item type PO items for standard POs in the item category *3-way match, invoice after GR* were grouped, because they also yielded the same process models. The new groups are displayed in table 2. The cluster IDs for these groups are used for reference in the following.

cluster ID	name	count of PO items
01_01_01	3-w. match, inv. after GR (with SRM; Item Type: Service)	318
01_01_02	3-w. match, inv. after GR (with SRM, Item Type: Standard)	96
01_02_01	3-w. match, inv. after GR (without SRM; Item Type: Service)	676
01_02_02	3-w. match, inv. after GR (without SRM, Item Type: Standard)	6446
01_02_03	3-w. match, inv. after GR (without SRM, Item Type: Subcontracting and Third-Party)	268
02_01	3-w. match, inv. before GR (with SRM)	705
02_02_01	3-w. match, inv. before GR (without SRM, Item Type: Standard)	161613
02_02_02	3-w. match, inv. before GR (without SRM, Item Type: Subcontracting)	1611
02_02_03	3-w. match, inv. before GR (without SRM, Item Type: Third-Party)	4167
03	2-way match	287
04	Consignment	13397

Table 2. Case groups by item category, SRM inclusion and item type

3.2 Description of the process models

With the split event log, all process models could be generated. To be able to compare the process models, the same filter and mining parameters were set for all of them. With a heuristics filtering of a minimum event observation as well as start event and end event observation of 10% for each group, the simplicity of the generated models was enhanced. After this, the inductive miner was applied with a noise threshold of 20%, reducing the rarely observed edges in the graphs and enhancing simplicity. As the models are intended to be understood by the process owner, BPMN as a standard for business process modeling was used for the visualization of the models [1]. The fitness and precision of the initially created Petri nets and the simplicity in the form of the number of nodes and edges of the converted BPMN models for the case clusters are presented in table 3. The predominantly high values in precision and thereby low values in generalization are accepted due to the simplicity of the models, which implies a lower degree of overfitting to the event data. Weighted averages of 93% for fitness and 99% for precision could be achieved. All process models are attached in the appendix of this paper.

In the following, the main activities and building blocks of the purchasing process are described without specifying the activity sequence, since the sequence and repetition of activities may be different for every process model.

In the process models, a PO item is created either with an SRM system or another not further specified system. The creation of the PO item may be combined with other activities for the request and approval of this PO item, e.g., *Create Purchase Requisition Item* or *SRM: Awaiting Approval*, but this approval is said to not be mandatory in the challenge description [5]. In the progress of a PO item then a form of goods or service receipt is recorded with the activities *Record Goods Receipt* and for PO items of the item type *Service* additionally the activities *Record Service Entry Sheet*. The vendor creates an invoice, expressed in the activity *Vendor creates invoice* and this invoice receipt is recorded with

cluster ID	number of cases	fitness	precision	number of nodes	number of edges
01_01_01	318	87%	78%	18	22
01_01_02	96	96%	100%	15	15
01_02_01	676	91%	62%	13	17
01_02_02	6,446	94%	100%	10	11
01_02_03	268	94%	100%	8	8
02_01	705	95%	100%	15	15
02_02_01	161,613	93%	100%	10	11
02_02_02	1,611	90%	88%	15	19
02_02_03	4,167	92%	91%	12	13
03	287	97%	70%	10	11
04	13,397	94%	100%	6	6
Sum / Weighted Avg.	**189584**	**93%**	**99%**	**10**	**11**

Table 3. Process model quality indicators

the event *Record Invoice Receipt*. Invoices are cleared with the activity *Clear Invoice*. If a payment block was prior set for the invoice or PO item, e.g., due to a missing GR for the invoice, this payment block has to be removed before the clearing of the invoice with the activity *Remove payment block*. Additional to the described activities, there may be more activities relevant to only some cases.

The main difference of the process models can be recognized in the following areas, which are described in the following.

- Inclusion of the SRM system
- Loops of GRs, invoice creations, invoice receipts and invoice clearings
- The sequence of GRs and invoice receipts
- Integration of payment block removals
- Other special activities

Inclusion of the SRM system

An SRM is included in the clusters 01_01_01, 01_01_02, and 02_01 and supports the creation of the PO item at the start of the purchasing process. The SRM block always starts with the *SRM: Created* activity and additionally a PO item is created, which is mandatory in clusters 01_01_01 and 02_01 and only optional in group 01_02_01. After the document creation, approval is awaited in all clusters. Afterward, there is a change transmittance, which is followed by the *SRM: Complete* activity and a *SRM: Document Completed* event in all case groups. The two activities *SRM: Ordered* and *SRM: In transfer to Execution System* are conducted subsequently; however, in the different process models, different direct sequences were discovered. While cluster 01_01_01 identifies *SRM: Ordered* and *SRM: In Transfer to Execution System* as parallelly running activities, the groups 01_01_02 and 02_01 discover a direct succession from *SRM: In Transfer to Execution System* to *SRM: Ordered*.

Loops of GRs, invoice receipts and invoice clearings

In the analysis of the activities of GRs, invoice receipts, and invoice clearings the possibilities for different frequencies of these activities per PO item could be observed in the various case groups. The following possible process flows can be distinguished:

1. One GR, one invoice receipt, one invoice clearing
2. Multiple GR, multiple invoice receipts, multiple invoice clearings
3. No GR, one invoice receipt
4. Multiple GR, no invoice receipt

The first group with one GR, one invoice receipt, and one invoice clearing could be found for all item types except the *Service* item type. The possibility for multiple GRs, multiple invoice receipts, and multiple payments is observable in process models of the item type *Service* represented by case groups 01_01_01 and 01_02_01. The occurrence of multiple GRs and invoices, therefore, appears predominantly for service items, while standard, subcontracting and third-party items at least in the vast majority of cases do not show this behavior.

An invoicing without GR is conducted for PO items of the item category *two-way matching* according to the process model generated for them in case group 03. This matches the definition of two-way matching in the BPI challenge 2019 description [5].

For consignment PO items, one GR record and no invoice receipt can be created according to its process model for case group 04. This also matches the specifications of the BPI challenge 2019, because the invoicing process for consignment orders are conducted in another system that is not included in the event log [5].

Sequence of GRs and invoice receipts

A difference in the sequence of the GRs and invoice receipts for PO items can be seen in the process models discovered from the event log. One possibility observed in the process models is that a GR is succeeded by one or multiple invoice receipts, which is true for the majority of the process models and PO items. If there are loops of multiple combinations of GRs and invoice receipts, in which for the loop iterations the invoice receipt(s) succeed the receipt of the goods, the process models are also counted to be in the GR-invoice receipt succession group. The following case groups follow this pattern: 01_01_02, 01_02_02, 01_02_02, 01_02_03, 02_01, 02_02_01, 02_02_03, 03 and 04.

Another behavior observed for the case groups 01_01_01 and 02_02_02 is a parallel execution of GR and invoice receipt, in which the order of these is not critical. Although a GR and invoice receipt parallelization describes the specification of the item category *3-way match, invoice before GR*, the majority of cases follows the GR-invoice receipt succession.

Inclusion of payment block removals

For three case groups, the activity *Remove payment block* can be seen in the process models. This activity is primarily described as used for releasing the payment after a mismatch between invoice receipt and GR is solved by an incoming GR [5]. This behavior would be assumed in the item category *3-way match, invoice before GR* due to the possible parallel execution of *Record Invoice Receipt* and *Record Goods Receipt*. Two of the case groups are from this item category, i.e., the case groups 02_02_01 and 02_02_03 for the item types *Standard* and *Third-Party*. Additionally, the activity can be found in the case group 01_02_02, which is valid for the PO items of category *3-way match, invoice after GR*, and type *Standard*.

Other special activities

There are activities which are only present in process models for some case groups. The reasons for including or excluding these are not given and provide a lead for further investigation.

A purchase requisition item is only included beyond the noise threshold in case cluster 04. In the BPI challenge 2019 description, it is said that approval workflows are not in focus. However, it may be valuable for the company to check whether the purchase requisition items are created according to the rules that have been set up for them.

The activity *Change approval for Purchase Order* can be found for the case group 03. In cluster 03 for 2-way-matching, the approvals for the PO may even be changed multiple times and at least one time. It is not evident, what the meaning of these approval changes are and why they occur; however, this may also be a valuable area to research for the company in focus.

Prices changes and quantity changes are part of the process model for case group 02_02_02. Here it is a choice in the process model to change the price after the PO item creation. Quantity changes are even regarded as a mandatory activity. With this finding, an opportunity for investigation is released asking why so many price changes and quantity changes occur that they are included despite the applied noise threshold.

3.3 Validation of case group and process model quality

While the generated process models lead to satisfactory model criteria in fitness, precision, and simplicity, the respective process models do not always represent the expected behavior and possibilities. As the case group 02_02_01 covers the most cases with 161,613 process models, this group is reviewed more closely.

In the process model presented in figure 3, a recording of a GR always precedes a recording of an invoice receipt. This does not reflect the behavior for PO items of item category *3-way matching, invoice before goods receipt*, for which "invoices can be entered before the goods are receipt" [5].

To review, whether an infrequent, but correct behavior was filtered by noise filtering, the trace clustering ProM plugin ActiTraC was used, which allows

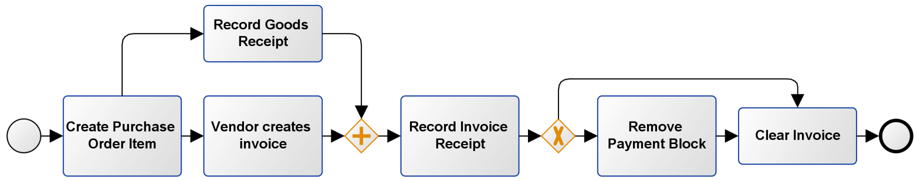


Fig. 3. Process model for case group 02_02_01

grouping traces into groups of similar behavior. With this technique, two large groups of traces could be identified, and the third group with additional behavior seen in the log was created. All process models with their number of PO items are shown in figure 4.

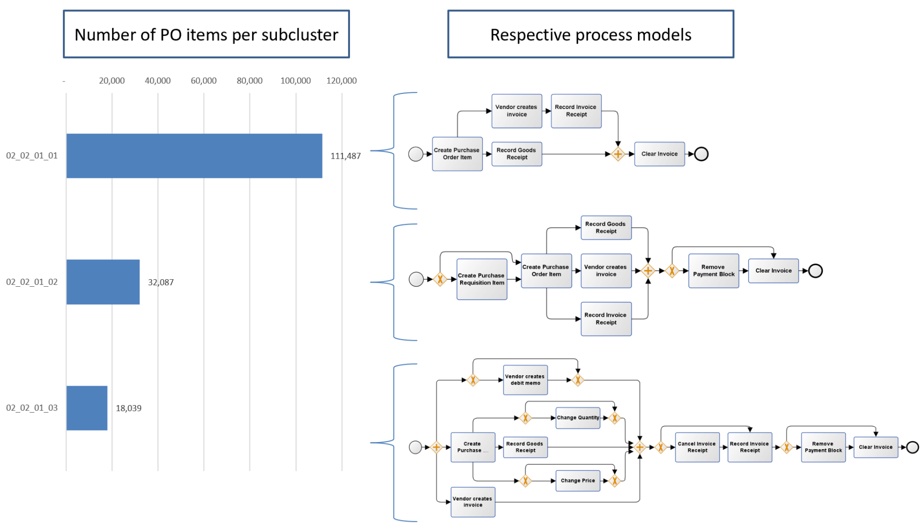


Fig. 4. Subcluster models for case group 02_02_01

In the first two process models, no order is defined for the sequence of the activities *Record Goods Receipt* and *Record Invoice Receipt*. In the third process model, which is applicable to ca. 11% of the PO items, an indirect succession from a GR to an invoice receipt is defined. It can be derived that for the majority of PO items, a GR is followed by an invoice receipt and the cases, in which an invoice receipt is followed by a GR, are filtered in the noise threshold for the combination of all cases of the group 02_02_01 or for the subcluster 02_02_01_3.

Due to the possibility of finding subclusters by ActiTraC, it was inspected whether there is a possibility to split the event log further while still taking into account the case attributes with more distinct values, namely the columns _case_Spend_area_text_, _case_Sub_spend_area_text_, _case_Name_, and _case_Vendor_. The asynchronous correlation of these columns to the found clustering was calculated with the uncertainty coefficient, and the values are presented in table 4.

column	uncertainty coefficient
_case_Spend_area_text_	0.0208
_case_Sub_spend_area_text_	0.0505
_case_Name_	0.1736
_case_Vendor_	0.1780

Table 4. Uncertainty coefficients of case attributes with many distinct values

Although the values depicted in table 4 indicate a small level of correlation, this meets the expectations since there are many values for both the dependent and the independent variable. A maximum uncertainty coefficient of 0.1780 did not reason a further split of any these columns to decide upon the process models. Therefore the presented groups are assumed to be natural presentations of user and PO item process behavior of the same category or linked to an attribute, which may not be provided in the dataset.

4 Throughput Time Analysis

The performance of an organization can be measured using time-related indicators like the throughput time of a process. The average time that a PO item spends in the process is termed throughput time. For this challenge, in this section the focus is to answer the challenge question "What is the throughput of the invoicing process, i.e., the time between GR, invoice receipt, and payment (clear invoice)? " [5].

As per the as-is processes and the models explained in section 3, for one PO item, there can be multiple GR messages and multiple invoice receipt messages. To identify which of them belong together within one PO item and further calculate the time taken for the execution of the events, the sequence of the below steps is executed.

1. A consecutive number of rows is assigned using ROWNUMBER() in SQL by sorting on the _eventID__ in ascending order OVER partitioning by _case_concept_name_ and _event_concept_name_. This newly identified column is termed as *iteration_number*.
2. The time between applicable activities *Record Goods Receipt*/*Record Invoice Receipt* to *Record Invoice Receipt*/*Clear Invoice* that belong to the same iteration_number and same PO item is calculated and the average value of them over every _case_concept_name_ is taken as the time between GR, invoice receipt, and payment.

The above-mentioned technique bears the risk that not all the above-defined activities are present in every case. Further, the existence of the activities *Cancel Invoice Receipt* and *Cancel Goods Receipt* can result in negative values which lead to inaccuracy in the throughput time of the invoicing process. The entire data is filtered by the rules below, to avoid this inaccuracy.

1. All complete cases, as explained in section 2 are considered.
2. Cases that contain at least one of the activities *Cancel Invoice Receipt* and *Cancel Goods Receipt* are not considered.
3. Cases in which the invoice is cleared before the invoice receipt are not considered.
4. Cases belonging to _case_item_category_ *3-way match, invoice after GR*, for which invoice was created before the receipt of the goods are not considered. Finally, the time difference between the GR and the invoice receipt as well as the time difference between the invoice receipt and the invoice clearing are calculated.
5. For cases which belong to _case_item_category_ *3-way match, invoice before GR*, the invoices can be recorded before the receipt of goods. Hence, the entire data of this category is divided into the following subcategories:
   - PO items, for which invoices were recorded before good receipts: The time difference between the invoice receipt and the GR as well as the time difference between the GR and the invoice clearing are calculated.
   - PO items for which invoices were recorded after GRs: The time difference between the GR and the invoice receipt as well as the time difference between the invoice receipt and the invoice clearing are calculated.
6. Cases belonging to _case_item_category_ *2-way match*: Since there is no separate GR message, the time difference between the invoice receipt and the invoice clearing is calculated.
7. Consignment cases were not considered since no invoice receipts were recorded in the system.

The implemented results of the above-mentioned steps, along with case attributes and event attributes were imported into *celonis*. The throughput time of the invoicing process for different values of _case_item_category_ was analyzed by grouping on _case_item_type_ and _case_Document_Type_ using the column chart component in *celonis* as shown in figure 5 and figure 6.

The interpretation and further analysis of the throughput times are not followed due to the following reasons:

1. There could be a delay in the company receiving the invoice receipt and the user recording the same invoice receipt in the system, but these cannot be known from the data.
2. The payment terms are not specified. However, the time between the invoice receipt and the payment depends on these. If the payment terms were known, the adherence to these could be calculated with the throughput times as described above.

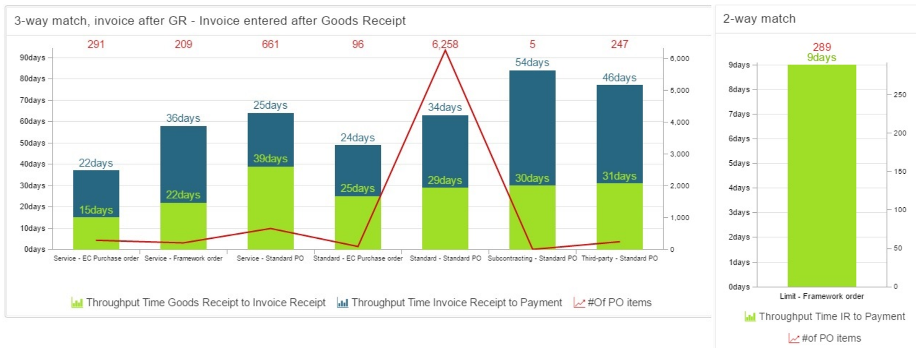


Fig. 5. Throughput times for *3-way match - invoice after GR* and 2-way match PO items

Fig. 6. Throughput times for *3-way match - invoice before GR*

5 Analysis of process deviating purchase order items

In this chapter, the given event log is analyzed on the compliance and the factors for incompliant and thereby deviating PO items. For this, the first additional dimensions are proposed which subsequently can be used in classifiers to detect incompliant cases based on their attributes and events. One of the additional dimensions aims to identify rework activities, which are presented in subsection 5.1. Fostered by the additional dimensions, the compliance classifiers are then described, and findings in their interpretation are presented.

5.1 Additional dimensions

Additional dimensions were created to analyze other parameters influencing the cases to turn out compliant or incompliant. The focus was to introduce other features, deduced from the event log to evaluate their contribution towards the case compliance.

Since a typical case involved several events associated with it, additional dimensions were further calculated on both case level and event level and could thereby be used for either post-mortem based predictions or pre-mortem based predictions, as explained in subsection 5.3. The approach was to gain more in-depth insight into the case compliance using all the information available at the case level. On the event level, the same features were fed to the classifiers that were available until the timestamp of a particular event occurrence.

Number of handovers: On the case level, the number of users excluding "batch_user" and "NONE" who were involved in a particular case defined the sum of handovers of work per case. At the event level, the dimension was calculated to understand the sequence in which each user was handed over the work against each event in a case.

Count of rework activities: Any event with _event_concept_name_ that indicated rework was identified and counted. These activities generally include "change," "cancel," or "delete" keywords in their activity names. A Boolean indicator "is_rework" was calculated against each event and later summed up by _case_concept_name_ to measure the final count of rework as the case level dimension and until a particular event for the event level dimension. Rework activities are explained in detail in subsection 5.2.

Segregation of duty (SOD): The dimensions *sod_create_poi_and_gr* and *sod_create_poi_and_ir* would be true against each case if the same user was responsible for creating the purchase order item with the user who recorded goods receipt or invoice receipt. On the event level, this dimension would inform whether the segregation of duty was achieved or not at each event occurrence.

Retrospective PO items: To understand the deviations from the process flow, particular records, in which the receipt of the goods was logged before the creation of purchase order item or in which the invoice was recorded before logging the receipt of the goods, were marked as retrospective purchase order items by a boolean. The background of this measure was that retrospective PO items may indicate that the purchasing process was not followed and the order may have been placed without creating a PO item in the system.

Resource Workload: This dimension provided the information about the work performed by each user in the past two days and seven days to understand the workload of the resource at the time of occurrence of the event. To use these values, they were aggregated on the case level with their average and their maximum values. Since it is possible that the quality of some activities is affected differently by higher workload than for others, this dimension was drawn in relation to the activity names.

Create Order net value: The total worth of a PO was calculated and appended in the case level attributes.

Throughput time: The throughput time is defined as the time from creating a PO item to the last event in the particular case. This time was further calculated based on the timestamp of occurrence of each event in days.

Material count: It was noticed that some POs contain items belonging to different groups of _sub_spend_area_text_. Therefore, this dimension was added to count the different materials in one purchase order.

Missing material: This dimension indicates in a Boolean value, whether a_sub_spend_area_text_ was given in the data or not. The assumed relation to compliance was drawn, because missing sub spend areas may indicate missing master data and thereby a higher chance of wrong entries.

Process cluster: As explained in section 3, all cases with their events were grouped into clusters, and this metric was also used as an additional dimension or feature for the decision tree learning.

Number of orders placed on the same day to the same vendor: This feature measures the count of orders on one day to the same vendor and applies this to the creation days of the PO items. The reason that this indicator was drawn is that confusion and potentially wrong allocations may arise between different orders that were placed on the same day for the same vendor.

5.2 Rework analysis

In this section, the rework steps were analyzed using *celonis* to identify the number of rework activities and the number of PO items with rework activities. Initially, a variable *reworkActivities* was defined with the set of defined activities as listed below:

Change Price, *Change Quantity*, *Cancel Invoice Receipt*, *Cancel Goods Receipt*, *Delete Purchase Order Item*, *Change Approval for Purchase Order*, *Change Currency*, *Change payment term*, *Change Final Invoice Indicator*, *Change Delivery Indicator*, *Change Rejection Indicator*, *Change Storage Location*, *Cancel Subsequent Invoice*, *Change Rejection Indicator*

The total distribution of the case and event data explained in section 2 was considered and a component filter on the defined set of *reworkActivities* was implemented on the column chart to visualize the number of PO items and the number of events in which rework was performed. The results are visualized in figure 7. The rework activities *Change Quantity*, *Change Price* and *Delete Purchase Order Item* were observed to occur most often in the total distribution.

5.3 Compliance classification on case level

The identification of the origins of incompliant cases was aimed to be solved with predictive techniques of machine learning. Using the relevant data to learn which features provide information about the compliance of a PO item, a classifier was built that could categorize a set of features to belong to either an incompliant or a compliant PO item. Using a *post-mortem* dataset with the complete knowledge about complete PO items, it was verified whether such a classification is possible. In a productive environment, however, during the execution of a purchasing process, not all information about a case is known, and only the information up to a specific event can be taken into account. Therefore, a *pre-mortem* dataset,

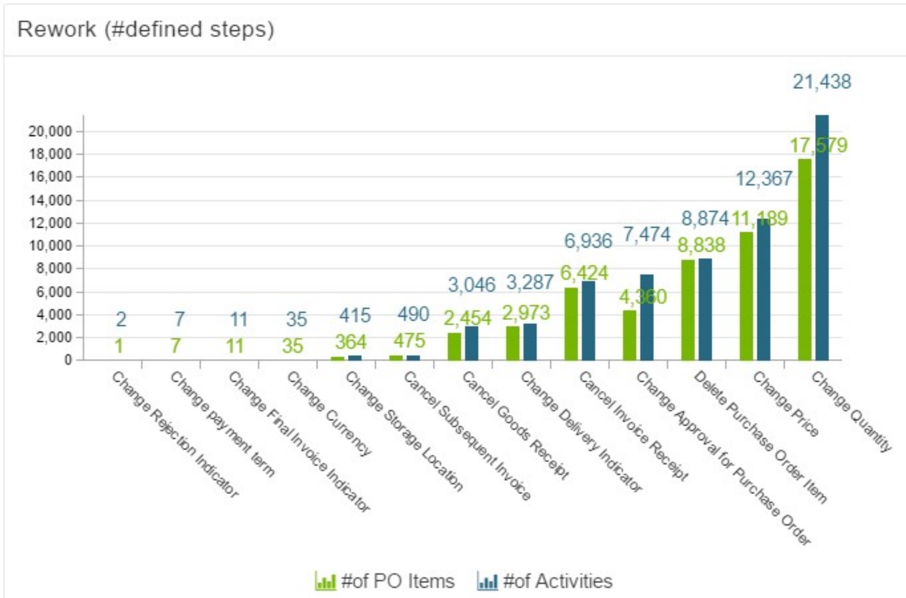


Fig. 7. Number of rework events and cases per rework activity name

was additionally used to indicate whether a particular combination of features during process execution has a higher chance to lead to incompliant behavior in a case.

The primary goals of the implemented classifier encompassed the understandability to learn about the influencing factors on compliance and the ability to find a large share of the incompliant cases with a low amount of false predictions. A decision tree algorithm was the classification technique of choice due to the inherent possibility of visualization and the manageability of simplicity. However, with an increasing number of depth and leaves, decision trees grow more complex, so the achieved prediction correctness was balanced out with the complexity of the decision tree. Analyzing an imbalanced data set with 99.6% compliant PO items, accuracy as the measure of correct predictions was not an insightful model quality measurement. 99.6% of accuracy could be reached by only guessing all PO items to be compliant. Therefore, recall as the ratio of correctly predicted incompliant cases of all incompliant cases and precision as the ratio of the false positives from all positive predictions, i.e., the proportion of actually incompliant cases from all predicted incompliant cases, were used as the key measures of the model. As the harmonic mean between recall and precision, the F1-score was additionally used as a classifier quality indicator.

In the preprocessing of the decision tree classifier, all relevant data was imported into Python and the categorical variables were binarized. The deduced values of the GRs and invoice receipts, as well as their receipt quantities, were included in neither case level nor event level prediction. The exclusion was due to the reason that factors leading to these values were sought after, and the values could nevertheless be calculated directly. In the following, all used data columns are presented.

Categorical attributes:

- _case_Document_Type_
- _case_Item_Category_
- _case_Spend_classification_text_
- _case_Item_Type_
- _case_Sub_spend_area_text_
- Process cluster

Numerical attributes:

- Presence of events per activity name
- Average resource workload per activity type
- Maximum resource workload per activity type
- Number of handovers
- Number of rework events
- Number of sub-spend areas in the superordinate PO
- Missing material flag
- SOD: PO created and GR recorded by the same user
- SOD: PO created and invoice recorded by the same user
- Create Order net value
- Flag to indicate whether it is a retrospective PO item
- Total throughput time in days

The data was fitted to a balanced decision tree classifier, which was created by splitting the leaves by the best possible *gini* information gain optimization. With four layers of depth in the decision tree, satisfactory results in recall could be achieved at an acceptable level of complexity. Applying 5-fold cross-validation to the dataset and classifier, a mean recall of 88.40% with a standard deviation of 3.34 percent points was measured. For the same dataset and classifier, the precision amounted to 5.06% with a standard deviation of 1.61 percent points. A mean weighted F1-score of 96.11% with a standard deviation of 1.24 percent points was measured with the same parameters. These numbers indicate that a high majority of incompliant PO items are detected at the expense of low precision. Therefore, applying this decision tree's learnings could mean that almost all incompliant cases could be found without knowing the monetary values of goods and invoice receipts. However, only ca. one in every twenty predicted incompliant PO items would actually be incompliant. Since a general suspicion of all PO item with a more in-depth review of all PO items would result in the precision of 0.37%, the precision could still be improved with the findings. In this paper, the concentration was set towards achieving a high recall, but another approach for another direction may be to maximize precision and thereby find incompliant PO items with as little effort as possible.

The four-layer decision tree that was generated is presented in figure 12 and 13 in the appendix. It describes the combination of features that indicate whether a PO item has a higher chance to be compliant or incompliant. Despite the relatively high simplicity of the decision tree, it is still difficult to understand, why particular splits were taken and which features influence the decision tree in which way. Therefore, the inspected features were further analyzed on their importance on the classification. In figure 8, the importance values for the encoded features of the decision tree with the importance of more than 0.09% are listed. The term *encoded* here refers to the binarization for the categorical variables.

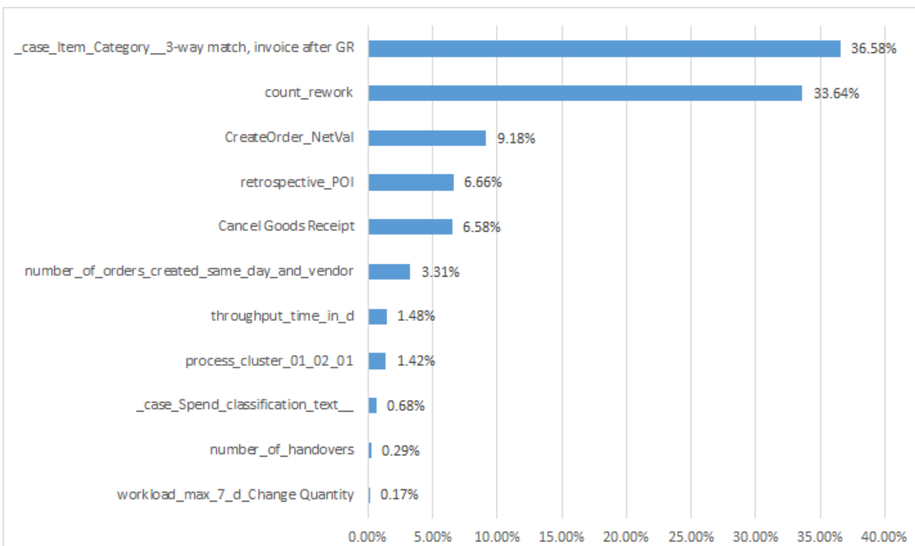


Fig. 8. Feature importance values for case level decision tree

In this step, the contributing features to compliance were known, and the decision tree indicated how a particular feature contributes to the compliance prediction. However, the question was left open, whether certain feature values increase or decrease the chance of compliance. Hence, for the attributes that the decision tree algorithm indicated as relevant, the distribution of compliant and incompliant cases was calculated. The change of proportion of incompliant PO items of all PO items filtered on a specific value range of the feature was calculated to see if the value range increases or decreases the proportion of incompliant cases. The ratios were compared to the ratio of incompliant cases within the total distribution amounting to ca. 0.37%. For categorical variables, all possible categories were taken as inputs, and for the numerical values, the ranges proposed by the decision tree were used. As an example, 20 highest values of difference to the incompliance ratio in the total population are presented in table 5.

column	value	imcompl. ratio	change to total incompl.
_case_Sub_spend_area_text_	Road Packed	40.69%	+10,978%
Cancel Goods Receipt	>0.5	11.82%	+3,119%
_case_Sub_spend_area_text_	Marketing Support Services	10.00%	+2,622%
_case_Sub_spend_area_text_	Sea	9.09%	+2,375%
_case_Sub_spend_area_text_	Government payments	9.09%	+2,375%
_case_Sub_spend_area_text_	Third Party Labor	6.25%	+1,601%
_case_Sub_spend_area_text_	Polyurethane Resins	6.06%	+1,550%
_case_Item_Category_	3-way match, invoice after GR	4.36%	+1,088%
_case_Sub_spend_area_text_	Packaging	4.18%	+1,039%
_case_Sub_spend_area_text_	Other Logistics Services	4.11%	+1,019%
_case_Sub_spend_area_text_	Digital Marketing	3.85%	+947%
_case_Sub_spend_area_text_	HR Services	3.39%	+823%
_case_Sub_spend_area_text_	Commercial Printing	2.79%	+660%
_case_Spend_classification_text_	OTHER	2.02%	+450%
_case_Sub_spend_area_text_	Waxes	1.73%	+370%
_case_Sub_spend_area_text_	Customers	1.68%	+358%
retrospective_POI	>0.5	1.33%	+262%
count_rework	>0.5	1.27%	+245%
CreateOrder_NetVal	>12729.5	1.23%	+234%
_case_Sub_spend_area_text_	Design	1.20%	+228%

Table 5. Feature values with the 20 highest incompliance ratios

In these 20 value distributions, some strong influences of feature values on compliance are already observable. Cancellations of goods receipt strongly increase the chance of impliance as well as when a PO item is entered retrospectively and with a high net value. If rework activities are registered, the chance for incompliance also more than doubles. Multiple sub-spend areas show a higher than average ratio of incompliant cases, so a focused review of these sub-spend areas may support the company in reducing incompliant cases.

Analyzing only these 20 highest feature values intervals or values with their incompliance may already provide the company with a basis for a guided improvement of their incompliance. The mentioned sub-spend areas pose a higher risk for incompliance and their processes can be observed more closely or a stricter approval process could be introduced for the related goods. Cancellations of goods receipts not only create additional effort, but also have a stronger tendency to turn out incompliant, so a filter of these PO items from a purchase manager could reveal the incompliant cases, before they are completed or clear open questions. The fact that higher net order PO items are affected with a higher impliance ratio even increases the issue of incompliance for these

PO items, so a more frequent review of these PO items may be used to reduce incompliance.

5.4 Compliance classification on event level

Although the decision tree on case level may enhance the understanding of the factors leading to incompliant PO items, it would have only limited applicability in a running system. This is because it uses all information that is only known after a completed purchasing process. Therefore, a further decision tree was implemented on the event level using only a pre-mortem dataset with only the information that is known up to the timestamp of a given event. A classification into compliant and incompliant cases should then be given for this particular case at the time of the given event. As the indicators recall and precision are expected to depend on the number of events in the case that is classified, not every event is expected to provide enough information for accurate predictions. The goal of the development of the classifier is the verification of the underlying concept, and the classifier may benefit from further information fed to it. The primary use cases aimed for with this concept is that warnings could be given to process managers when the probability of incompliance of a given case surpasses a defined threshold. The potential independent features comprise of the case level attributes known after the PO item creation, the presence for each activity as well as all additionally created dimensions. All of this information is restricted to the data that was recorded before or at the given timestamp and event-ID within the case. In the designed concept, not only the current information of a specific event is taken into account, but also the history up to this event. The categorical features were the same as for the case level detection, but the numerical features changed and are listed below:

- Missing material flag
- Retrospective PO item flag
- Count of rework activities
- CreateOrder_NetVal (PO item worth)
- Number of sub-spend areas in the superordinate PO
- SOD: PO created and GR recorded by the same user
- SOD: PO created and invoice recorded by the same user
- Total throughput time in days
- Task load past two days / seven days

Building, similarly to the case level decision tree, a four-layer model, similar features were used by the model to predict compliance. The most influential attributes were the *_case_item_category_*, the process cluster and the net value of the PO item. While these attributes are known from the case level and already known during PO item creation, also features on event level, like the throughput time in days, the task load and the presence of activities such as *Record Goods Receipt*, *Cancel Goods Receipt* and *Change Delivery Indicator* are used.

Applying 5-fold cross validation on the classifier, a recall of 79.47% with 4.16 percent points standard deviation could be achieved.The precision amounted to

8.50% with a standard deviation of 1.86 percent points, and the mean weighted F1-score was 96.05% with a standard deviation of 1.20%. These values indicate that the risk of breaching compliance of a PO item can be predicted with such a decision tree. These findings support the gained understanding from the case level decision tree that some attribute values are related with a higher risk of incompliance. The created decision tree is shown in figures 14 and 15 in the appendix.

The pre-mortem predictive incompliance detection could be used in the process owner's indication to warn the respective employees of the high risk of incompliance in a given PO item. Such a warning system could reduce the incompliance by raising awareness and concentration in the right moments. As a more severe consequence, the purchase managers could be also directed to the high risk PO items for a general review.

6 Discussion and future work

In this paper, the purchase-to-pay process could be described by multiple process models designed for different process areas given by a group of PO item attributes. Additionally, multiple calculated indicators could be created that may increase the understanding of the process and may support in the advisory of its quality. Among these measures were the compliance of the process and the throughput times between GRs, invoices, and payments. Consequently, the compliance influencing factors could be determined with the type of their influence and the strength of their impact. The predictive process mining techniques applied for this also were used on a pre-mortem dataset to verify the possibility of a compliance warning method in a running system.

The analysis of the influences was not exhaustive, and another research into the complementary forces between attributes could be a field of study in the future. However, the simplification of the decision tree findings may be applied similarly to other process mining analyses in the future and thereby constitute one good practice of interpretation. While also the process model creation may be abstracted to a general methodology, the creation of additional dimensions and calculations between the given variables and events is a unique result to this project. In this, it is specific to a purchasing process and even the given data columns. Therefore, the findings of this paper divide into the followed analysis steps and methodology as a potential good practice for the community of process mining scientists and practitioners and into the dimensions and findings that affect the particular company and purchasing analysts.

The designed classifiers to predict compliance can be used in the process owner's organization to reduce incompliance in a guided approach by determining the influential factors beforehand. With the pre-mortem decision tree, a continuous monitoring of the compliance and a shift of factors may be observed. A warning system or an introduced focused review system of PO items with higher incompliance risk may be introduced on basis of the implemented classifiers. However, if additional factors potentially influencing the compliance can be assumed in the future, these would first need to be implemented as additional dimensions.

7 Conclusion

Structured along with the BPI challenge requests, the contributions of this paper can be grouped into three parts.

First, a group of process models was created, representing the as-is conduct of purchasing within the process owner's organization. An analysis was conducted into the case attributes that may be used to split the data and which features show the highest within sum-of-squares similarity.

In the set-up of a throughput time calculation method, the rules for obtaining throughput times between GRs, invoice receipts, and payments even with multiple iterations of these documents were developed. The throughput times between invoice receipt and payment mainly depends on the payments terms within the organization. Furthermore, the time between GR and invoice receipt is both dependent on the supplier's invoice delivery as well as the organization's invoice recording speed. As data for both payment terms as well as invoice arrival time were not available, no further analysis was performed into the throughput times.

In a third step, the purchasing compliance of the organization was analyzed by creating additional dimensions combined with the given dataset to feed a decision tree classifier. With this classifier and subsequent analyses, the influencing factors to purchasing compliance could be identified. The concept of a compliance warning decision tree in a running system was explored using pre-mortem perspectives of the data. By obtaining results, with which more incompliant cases could be detected early in the process, this concept was found to be valid to reason further research and company implementations.

While therefore the findings of this paper may support the process owner in their process understanding and compliance transparency, the applied methodology and the proved concept into predictive compliance detection with process mining may lay the ground for new scientific advancements in this direction.

References

1. Business process model & notation^{TM} (bpmn^{TM}), https://www.omg.org/bpmn/
2. Hompes, B.F., Buijs, J., van der Aalst, W.M., Dixit, P.M., Buurman, J.: Discovering deviating cases and process variants using trace clustering. In: Proceedings of the 27th Benelux Conference on Artificial Intelligence (BNAIC), November. pp. 5–6 (2015)
3. van der Aalst, W.: Process mining: discovery, conformance and enhancement of business processes, vol. 2. Springer (2011)
4. van der Aalst, W.: Data science in action. In: Process Mining. Springer (2016)
5. van Dongen, B.F.: Bpi challenge 2019. https://doi.org/10.4121/UUID:D06AFF4B-79F0-45E6-8EC8-E19730C248F1

8 Appendix

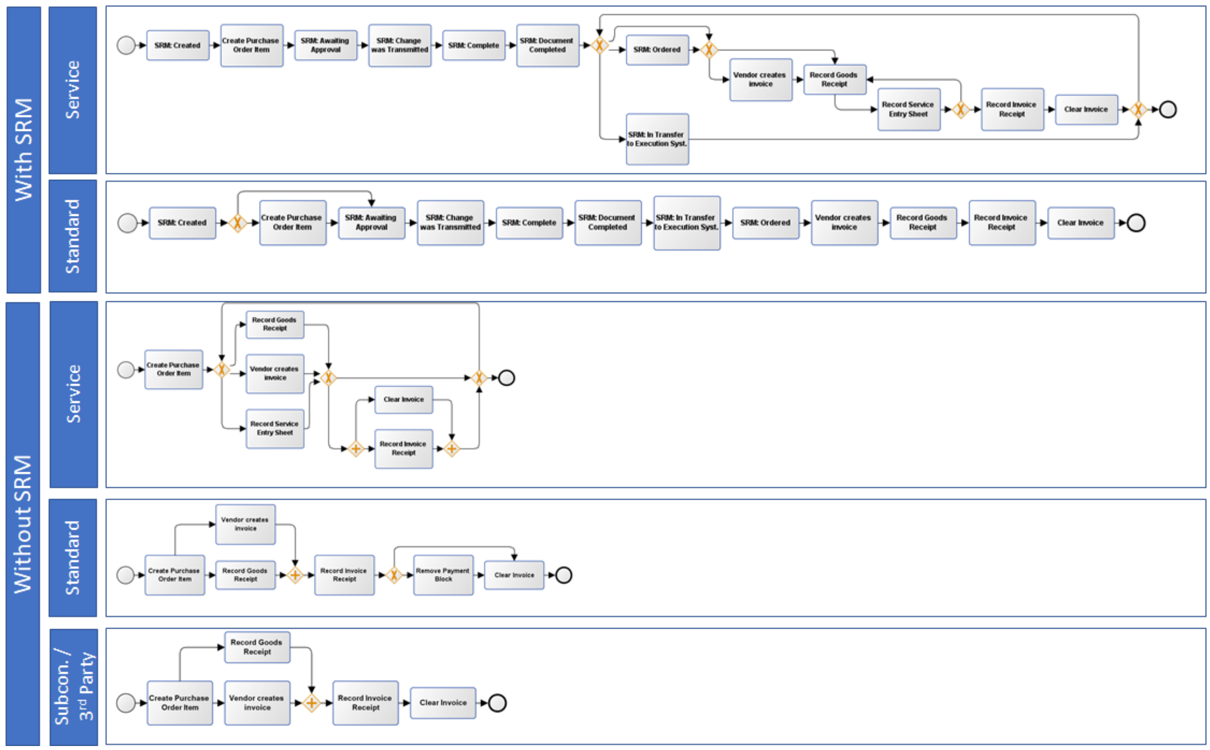


Fig. 9. BPMN models for *3-way match: Invoice after GR* order items

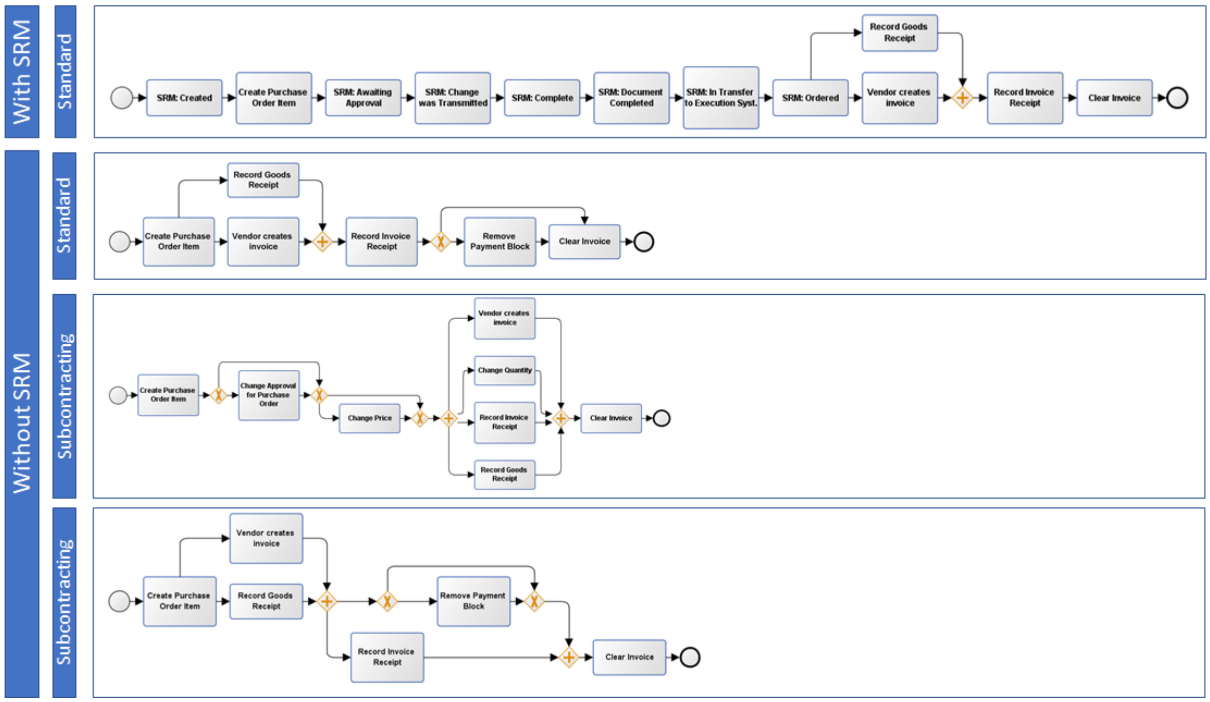


Fig. 10. BPMN models for *3-way match: Invoice before GR* order items

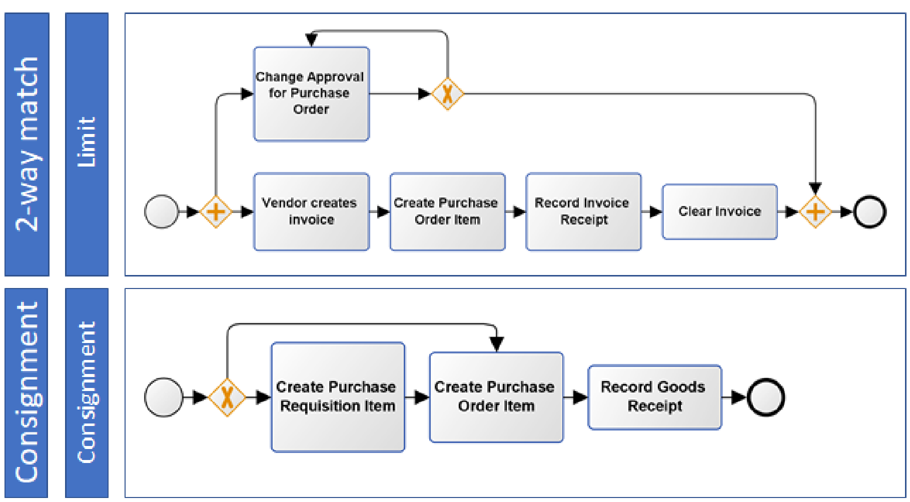


Fig. 11. BPMN models for *2-way match* and *Consignment* order items

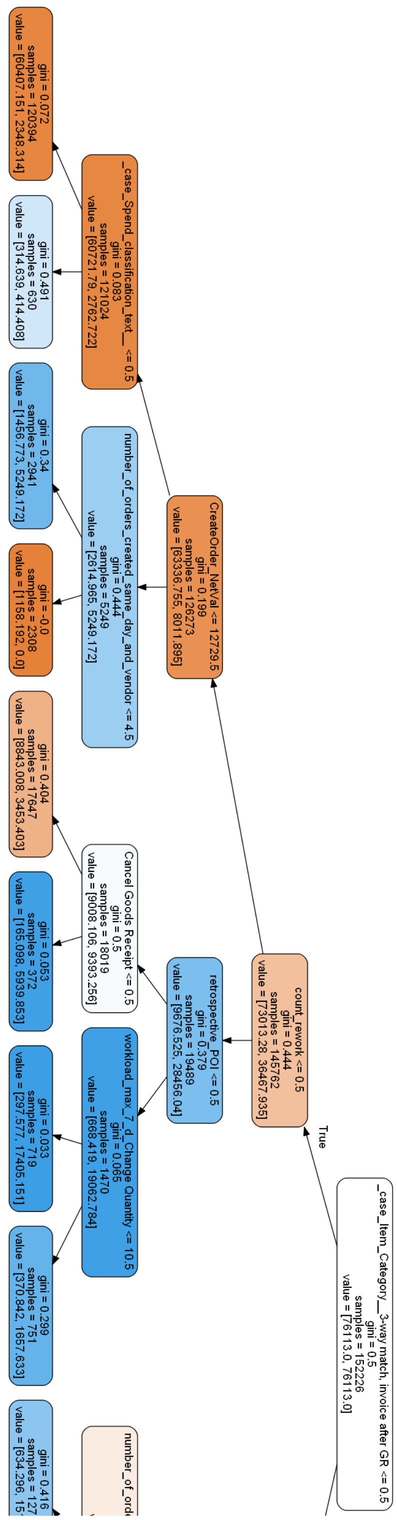


Fig. 12. Four layer case level decision tree [Left part]

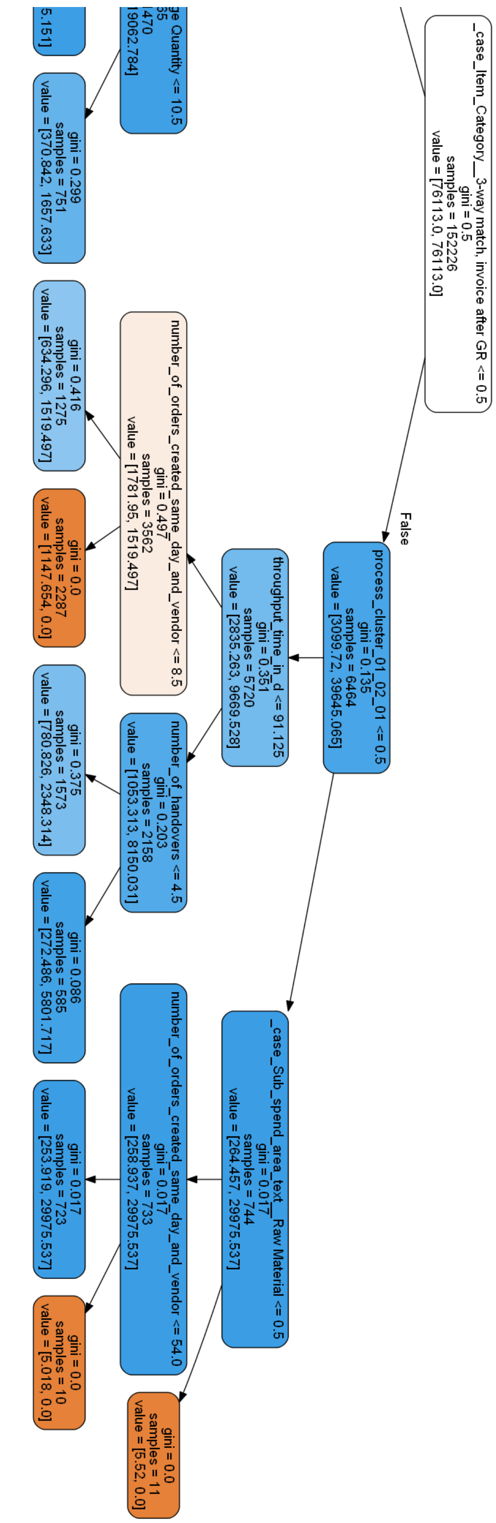


Fig. 13. Four layer case level decision tree [Right part]

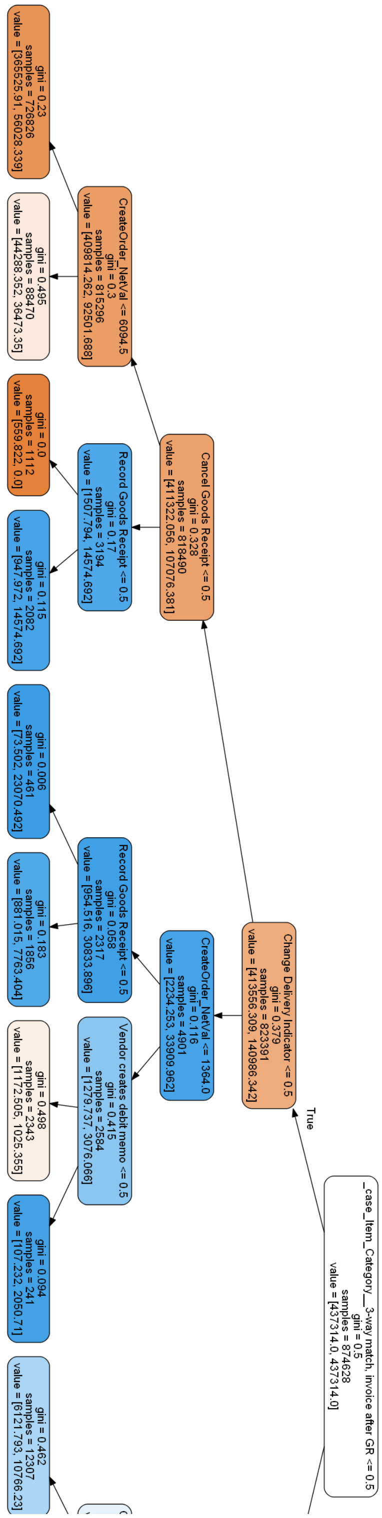


Fig. 14. Four layer event level decision tree [Left part]

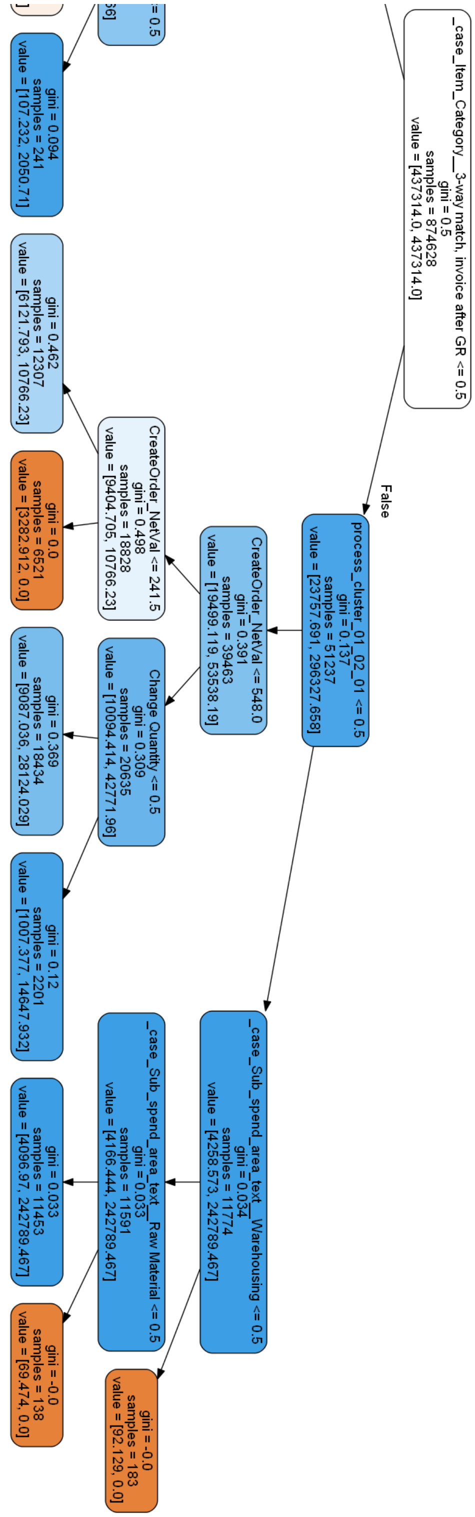


Fig. 15. Four layer event level decision tree [Right part]

Process Mining for optimization of a P2P process of a company in the coatings and paints industry

Florian Guschl[1, 2], Christopher Tzakov[1, 3], Peter Lingner[1, 4], Albert Fichtenau[1, 5], Kira-Marie Liebig[1, 6]

[1] Leopold-Franzens-Universität Innsbruck, Innsbruck 6020, Austria
[2] Florian.Guschl@student.uibk.ac.at
[3] Christopher.Tzakov@student.uibk.ac.at
[4] Peter.Lingner@student.uibk.ac.at
[5] Albert.Fichtenau@student.uibk.ac.at
[6] Kira-Marie.Liebig@student.uibk.ac.at

Abstract. Business operations and the processes behind them usually do not run smoothly. Why this is the case is often difficult to determine in large companies. Process Mining is a method to solve tasks of this range. The Business Process Intelligence Challenge offers the possibility to practice process mining techniques with real data. Hence, this paper analyzes a real-life event log of a Dutch organization, operating in the field of coatings and paints. First, all processes are modelled according to the BPMN 2.0 standards. For the subsequent challenges the purchase-to-pay process, particularly focusing on 3-way-matching, invoice before goods receipt is analyzed. The data is additionally restricted to the top ten vendors in the area of packaging for the year 2018. This paper aims to identify optimization possibilities in the area of compliance and throughput time. The tool Celonis is used throughout the analysis. The most relevant bottleneck for all of the top ten vendors is the direct process flow from the activity Record Invoice Receipt to Clear Invoice, which makes up a major part of the throughput time. The majority of the top ten vendors has compliance issues, hence recommendations for areas of improvement have been formulated.

Keywords: Process Mining · BPIC 2019 · Purchase to Pay · Compliance · Celonis.

1 Introduction

Process mining techniques allow for extracting information from business data and can be used to discover models describing processes, organizations, and products [1]. Relevant information of companies is often stored in so called event logs and enable the possibility of process mining [2]. Event logs contain information like Case ID, the timestamps of the start and end times, and other attributes of the event recorded by an IT system. Thus, an event log maps one or more cases of a business process but can also be a documentation of several, related business processes.

Process mining has experienced some major development over the last decade because companies need to learn how their processes perform in the real world [3], and the amount of data collected via information systems has increased substantially [4]. This discipline is the focus of this year's Business Process Intelligence Challenge (BPIC) organized by the Process Mining Conference 2019. For this challenge participants are provided with a real-life event log from a large multinational company operating from the Netherlands in the area of coatings and paints. The purchase order handling process for some of its 60 subsidiaries should be investigated, with special attention towards compliance questions.

Our approach to the challenge is structured as follows. This chapter explains the structure and research questions of this paper. The following chapter describes our used tools, gives an overview about the provided data set and clarifies the research objective and the used methods. The third part is about the three main challenges stated by the company. To solve the first challenge, the following question is answered:

1. How can the underlying processes be modeled according to BPMN 2.0 standards?

This helps to get an overview and first understanding of the process. Since we participate in the BPI Challenge as a team of the student category, we focused on specific cases for the second and third question. Those are:

2. How is the invoicing process performance in terms of throughput time affected by different process characteristics?

3. Can particular purchase documents of the invoicing process be successfully identified as potentially improvable irregularities? Which vendors are mainly responsible for deviations, regarding compliance?

The process to restrict the data on certain factors is described in detail in the next chapter. In order to generate meaningful insights, we further focus on issues regarding the invoicing process, with the major activity's vendor creates invoice, record invoice receipt and clear invoice. To structure possible findings in a meaningful way, we formulated the following research question that will be answered in the chapter of the respective challenge: Are there any optimization possibilities in the area of compliance and throughput time regarding the top 10 vendors?

2 Description of the data and methods

The provided and in this paper analyzed dataset is a fully IEEE-XES compliant event log and contains 76,349 purchase documents containing in total 251,734 items. Hence, there are 251,734 cases. In these cases, there are 1,595,923 events relating to 42 activities performed by 627 users (607 human users and 20 batch users). The vast majority of events took place in 2018. The dataset contains around 1.5 million events for purchase orders (PO) of the dutch company submitted in 2018. The data shows the purchase-to-pay process (P2P) without the approval workflow of the PO's and the invoices. The data refers to many different categories of goods and services and include many different types of vendors.

The assignment in the student category is to select a specific aspect of interest. We found an area playing an important role for the company. This is the spend area of ordered goods related to packaging. As seen in figure 1, we selected the top 10 vendors in this area by the number of cases. They are involved in around 67,000 cases that correspond to 26% of the total cases in 2018.

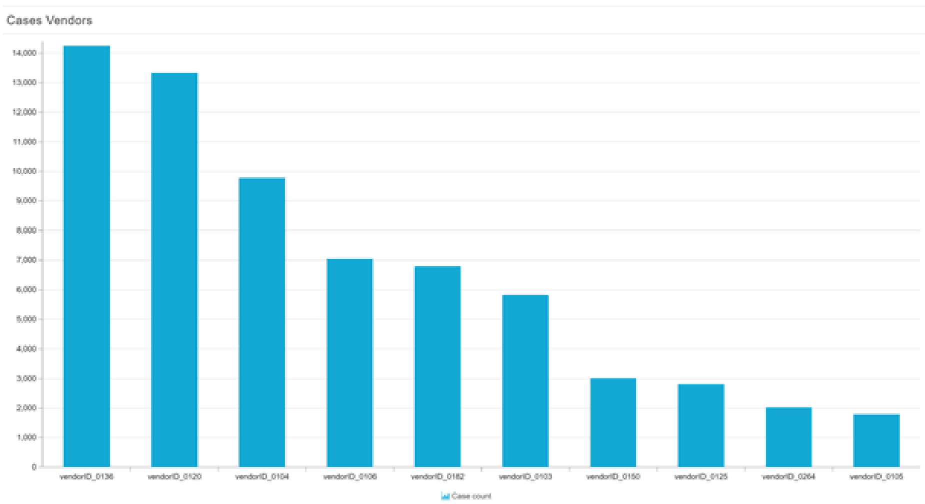


Fig. 1. Top 10 vendors in the spend area packaging (Own Depiction).

The main tool used for the analysis is Celonis. It is a suitable tool to investigate cases and events in detail but at the same time not losing the overview of the processes. Additionally, Microsoft Access and Microsoft Excel are used to conduct special investigations for a small number of cases but where more specific calculations are required.

3 Analysis of processes regarding the stated three challenges

3.1 Modeling of the main process flows

In this section an overview of the main processes is provided establishing a foundation for the subsequent analysis. The process flows are modelled according to the BPMN 2.0 standards and were simplified, so the most relevant path is depicted. The vendor is modeled as an empty lane as the majority of the vendor process model would be based on presumptions. Thus, the models, depicting the happy path, contain almost no message flows between the company and the vendors. Only "Vendor creates invoice" (VCI) was associated with the vendor as an activity for the purpose of illustration.

3-way matching, invoice after goods receipt

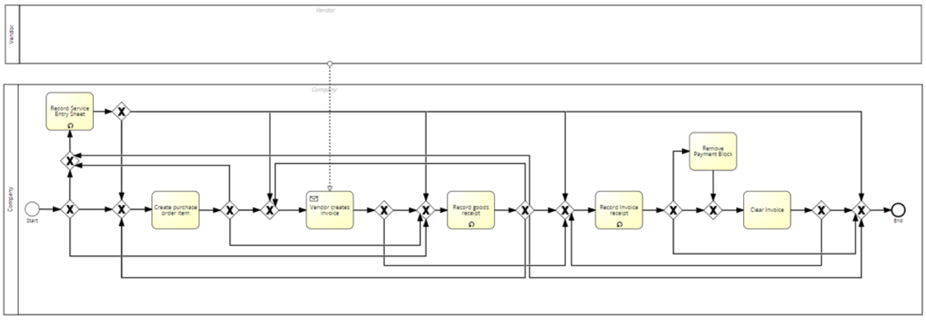


Fig. 2. BPMN 2.0 model of "3-way matching, invoice after goods receipt" (Own Depiction).

The "3-way matching, invoice after goods receipt" process flow consists of 15,182 cases (make up for 6% of total cases). The main focus in this process is that the "Record Invoice Receipt" (RIR) happens after "Record Goods Receipt" (RGR). For these items, the value of the goods receipt message should be matched against the value of an invoice receipt message and the value put during creation of the item. In figure 2 the process is modelled with 96.2% of all activities and 91.6% of all connections. The process usually starts with the creation of a purchase order item, followed by the creation of the invoice by the vendor. Then, the goods' receipt is recorded, followed by the record of the invoice receipt. In the end, the invoice is cleared. This most common path is also called the happy path and happens in 15.3% of the cases.

3-way matching, invoice before goods receipt

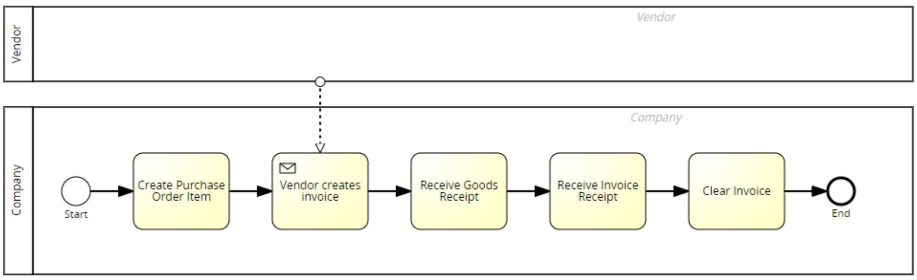


Fig. 3. BPMN 2.0 model of "Happy Path: Three-way matching: invoice before goods received" (Own Depiction).

The "3-way matching, invoice before goods receipt" process flow consists of 221,010 cases (make up for 87.8% of the cases). The items require a goods receipt message, while they do not require Goods Receipt-based invoicing. Invoices can be entered before the goods are receipt, but they are blocked until goods are received. This unblocking can be done by a user, or by a batch process at regular intervals. Invoices should only be cleared if goods are received and the value matches with the invoice and the value at creation of the item. The happy path (figure 3) for this process flow starts out with the purchase item being created and presumably send to the vendor. Upon receiving the order, the vendor creates the invoice, which is likely documented through a confirmation being send to the company. Next, the company receives the order, which is documented through a goods receipt, followed by the invoice receipt from the vendor. The last step in the happy path is the clearing of invoice. This happy path accounts for 47,957 of the 221,010 cases falling under the selected category. This means that roughly 21.7% of the cases are represented. The reason for such a low percentage is, that the process contains a vast collection of activities and connections which occur only infrequently, but which cause the majority of the cases to deviate from the happy path in terms of extra activities.

Interestingly, the goods receipt actually arrives before the invoice receipt in the happy path, which is counterintuitive given the name of this category. Filtering for cases in which the invoice does in fact arrive before the goods received leads to an entirely new happy path. The happy path also starts with the purchase order item being created, moving on to the vendor creating an invoice, which again we presumably know due to a confirmation message, which is followed by the invoice receipt, which in turn leads to the goods receipt which causes the payment block to be removed leading to the payment of the invoice. This new path is only present in 14,830 cases and the following happy path represents 59.6% of these (8,836), as such not being very representable for the 221,010 cases categorized as invoice before goods receipt.

Comparing the probabilities or rather frequency in which the two paths occur leads to the conclusion that while the invoice may be created early on this is probably only derived from an order confirmation message or later from the invoice date and it does not actually suffice for proceeding to the payment stage because in most cases the company only clears the invoice after another process step called: record invoice receipt – and even when both events occurs before goods receipt in only 685 of the 221,010 cases the invoice is actually cleared before the goods receipt is booked, which might indicate that these cases represent exceptions.

2-way matching (no goods receipt needed)

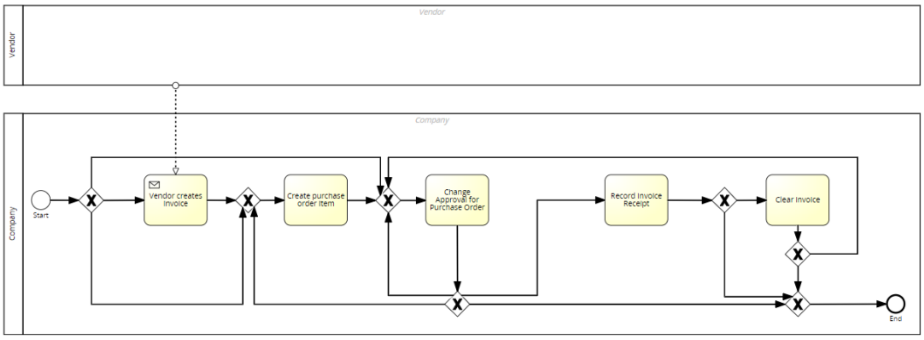


Fig. 4. BPMN 2.0 model of “2-way matching (no goods receipt needed)” (Own Depiction).

The “2-way matching (no goods receipt needed)” process is the process with the least amount of cases. It consists of 1,044 cases (makes up for 0.4% of total cases). For these items, the value of the invoice has to match the value at creation, but there is no separate goods receipt message required. In figure 4 the process is modelled with 99.1% of all activities and 94.7% of all connections. Due to the relatively small amount of cases it is not purposeful to model the process in even more detail. In contrast to the two 3-way matching processes, 2-way matching does not require a goods receipt and it inherits the activity “Change Approval for Purchase Order” (CAPO). Usually the process starts by either a vendor creating an invoice, creating a purchase order item or change approval for purchase order. The last two options mentioned, happen in every single process. If a vendor creates an invoice, then an invoice receipt is recorded in the end.

Consignment

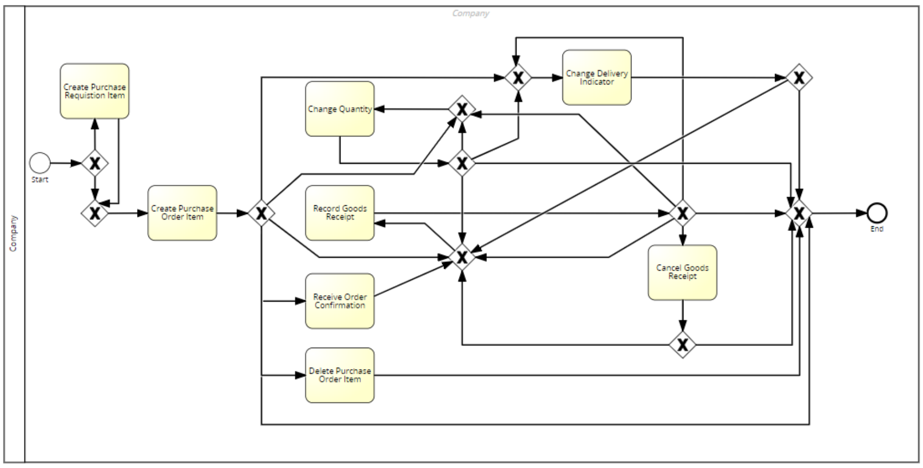


Fig. 5. BPMN 2.0 model of “Consignment” (Own Depiction).

The consignment process consists of a similar amount of cases as the 3-way matching, invoice after goods receipt process (14,498). The major difference to all other processes is that no invoicing is required. Instead the goods are already owned by investigated company and only shipped from the vendor. Vendor_0188 makes the most significant contribution with almost a third of all cases. Due to the lack of invoicing required the happy path is simply the creation of purchase order items followed by recording goods receipt. 60.5% of all cases follow the happy path. In figure 5 the consignment process is modelled with 99.7% of all activities and 99% of all connections. The next more noteworthy activity is the creation of a purchase requisition item before the actual creation of the purchase order item. After the creation of purchase order items, a change in quantity is required before the goods receipt activity which sometimes re-occurs before the process ends. All other parts of the process refer to less than 1,000 cases.

3.2 Investigation of the invoicing process

Throughput Time

To answer the second question as formulated in the introduction, the same selection process is chosen as already explained in chapter one. To investigate the average throughput time of cases inheriting specific activities, the data set is modified to only depict cases including the activities "Record Goods Receipt" (RGR) / "Record Invoice Receipt" (RIR) and "Clear Invoice" (CI). This selection matches 20% of all the cases analyzed.

The average throughput time for the cases involving the activities RGR / RIR and CI is 100 days, as presented in figure 6. The Happy path for the selected cases is a process consisting of five activities: "Create Product Order Item" (CPO) → RGR → VCI → RIR → CI.

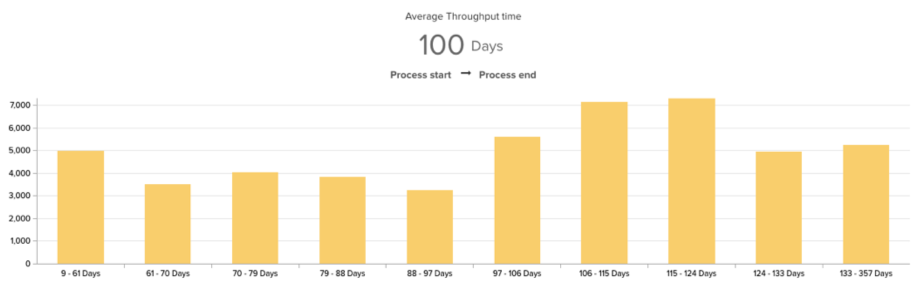


Fig. 6. Illustration of Average Throughput Time for cases involving the activities record goods receipt/ record invoice receipt and clear invoices (Own Depiction).

The average throughput time for the last two activities (RIR → CI) is 75 days affecting 67% of the cases. These two directly following activities constitute the slowest bottleneck. How this can be interpreted depends on the company's internal defaults. However, the investigated company could reflect upon why this process takes such a relatively long time. When analyzing whether there is a specific user who does not pay or does not pay fast, we can see that more than 50% of the cases are inherited by an unknown user (user_NONE), where the activity CI does not flow through the process. User_NONE indicates no user was recorded in the source system making the finding difficult to investigate further. Therefore, we consider the two directly following largest user, user_235 and user_029 as shown in figure 7.

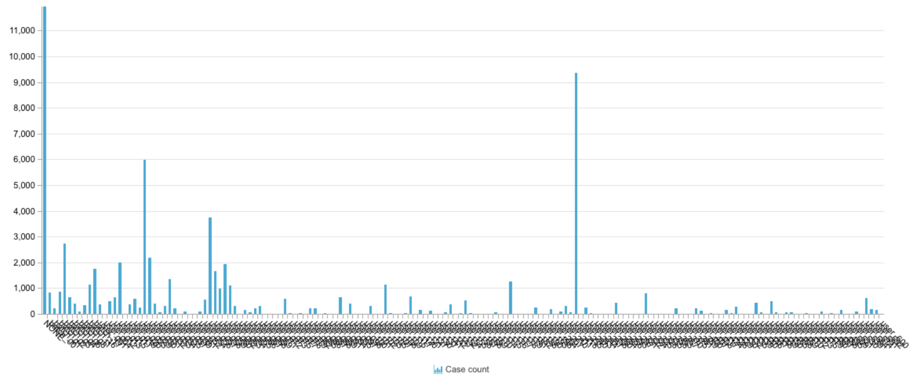


Fig. 7. Number of Cases depending on Users for cases that have not been paid (Own Depiction).

For user_235 we found that most cases which are not paid, start with the activity "Create Product Receipt" (CPR) while the cases which do contain CI, start with CPO.

For the vendors we found, that vendor_0136 and vendor_0120 are mostly involved in cases which do not flow through the activity clear invoice. But as depicted in figure 8 there is no great distance between most of the vendors.

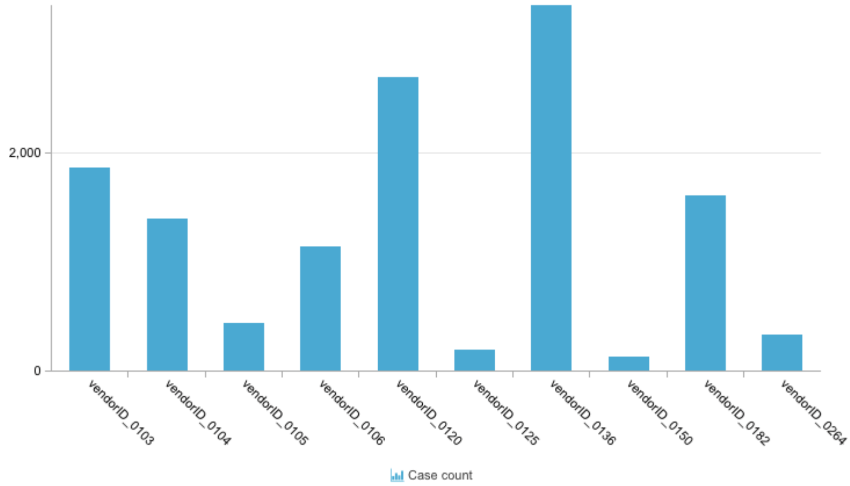


Fig. 8. Number of Cases depending on Vendors for cases that not have been paid (Own Depiction).

To answer the question, whether there are users which do pay relatively fast, we analyzed the throughput time depending on the users. We found no big difference between most of the users, while batch_01 sticks out as their throughput time is twice as long, shown in figure 9. This could be a result of the batch being many users, making coordination relatively hard, or this could also be a default instruction to hold back payments by the company investigated.

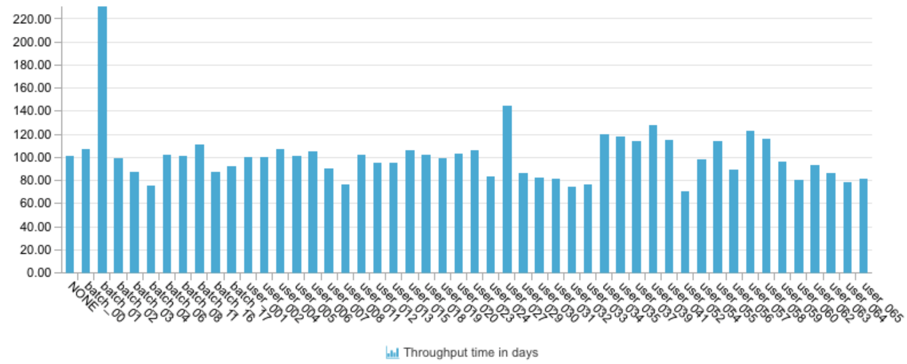


Fig. 9. Throughput time in days depending on the users for cases that flow through the activities Record Goods Receipt/ Record Invoice Receipt and Clear Invoices (Own depiction).

Furthermore, we investigated whether there is a specific user or vendor receiving or creating the activity, “Vendor Creates Debit Memo” (VCDM). We found that user_002 is involved in the biggest share of VCDM activity, as shown in figure 10. While this finding could also result from the fact of user_002 being in charge of all payments or a related company structure reason it might be worthwhile to focus on this user in future research.

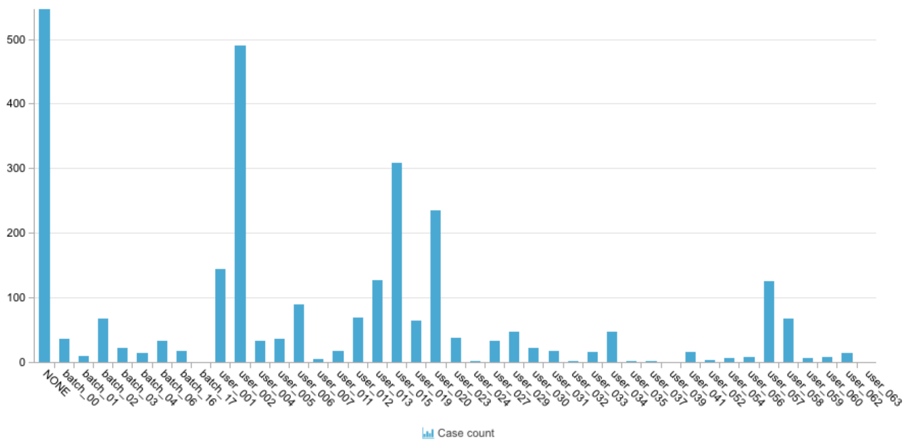


Fig. 10. Number of cases depending on users for cases involving the activity vendor creates debit meme (Own Depiction).

More interesting is the finding of the vendor_0136 creating with great distinct the most debit memos as shown in figure 11. Here one could interpret the vendor_0136 may have different regulation, for after which time limit the debit memo is sent out. To optimize this process, the investigated company could get in contact with the vendor and find a solution on how to minimize the debit memos. This could be of advantage, as the process gets more complex and employees have to file the memo within the internal database.

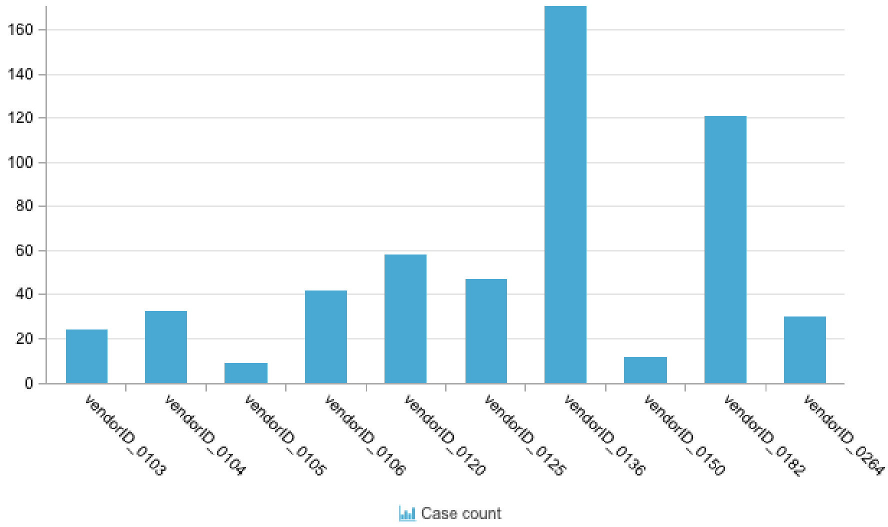


Fig. 11. Number of cases per vendor for cases involving the activity vendor creates debit meme (Own Depiction).

Examination of the top ten vendors

With our research question being “Are there any optimization possibilities in the area of compliance & throughput time regarding the top 10 vendors? “, this chapter focuses on the analyzes of the top ten vendors.

Therefore, we investigated the conformance for each vendor. Hence, we identified the respective happy path and compared it to the as-is process. We chose the happy path as it is the path that is executed most often, assuming the process to be the desired to-be process. The following chapter is sorted from the vendor inheriting the most cases to the vendor inheriting the least cases. We only investigate the invoicing process; therefore, we select only cases where the activities RGR / RIR and CI are included.

To generalize we investigated 50% of the top ten vendors mainly follow the happy path: CPO - RGR - VCI - RIR - CI (Vendor_120, Vendor_182, Vendor_150, Vendor_125, Vendor_264). Additionally, we find that 40% follow almost the same happy path only inserting the activity ROC as the second activity: CPO – “Receive Order Confirmation” (ROC) - RGR - VCI - RIR - CI (Vendor_104, Vendor_106, Vendor_103, Vendor_105). Only the Vendor_136 has a different happy path with the activity VCI following directly the first activity CPO: CPO - VCI - RGR - RIR - CI.

Looking at the top 10 vendors in the packaging spend area, it can be said that all vendors have issues with regard to the invoicing process but the amount is manageable. Vendor_136, Vendor_104 and Vendor_106 follow the happy path to a comparatively low degree. The process after the invoice is received until the invoice is paid takes a long time in almost all cases. However, as suggested in chapter 3.2 this could be due to internal regulations until invoices should be paid. The ten vendors are now briefly described individually with the individual most important findings regarding the payment process.

Investigation Vendor 136

The average throughput time for this vendor is 126 days. 20.99% of the process flows follow the happy path. This is the only vendor where the activity VCI follows directly CPO. The two largest bottlenecks are two activities that occur before CI: RIR (46% of cases) and RPB (52% of cases), which extend the process by 97 and 89 days respectively. 54% of the vendor's cases are not compliant with the Happy Path, as Remove Payment Block (RPB) occurs. This extends the process by three days. A root cause analysis showed that the process will also be extended if RGR occurs after CPO instead of after VCI.

Investigation Vendor 120

The bottleneck increasing the throughput time most considerable is the activity RIR → CI. This process takes 89 days in 76% of the cases. For this process no user can be identified that takes the longest with great distance, while for the whole process user_073 takes the longest with great distance. Opening up the opportunity of an internal investigation for the user_073.

Investigation Vendor 104

The happy path is represented by 30.16 % of the cases. Performing a conformance analysis based on deviations to the category happy path leads to the detection of several activity sets.

For 90% of the cases the step ROC occurs, however while adding on average 1.9 additional steps to the process the throughput time is reduced by 3.9 days on avg. and thus this violation is whitelisted.

For 65% of the cases the process flow deviation of RGR following CPO is detected. This not only increases the process complexity by 1 step on average, but also causes the duration to be 4.3 days longer. Performing a root cause analysis shows, that there are three users which together participate in close to 50% of the deviations. These users are user_065, user_056 and user_062. Because user_065 alone participates in close to 25% of the deviations the corresponding process is analyzed. A time-based process visualization shows that for no apparent reason there is a 62-day delay between receiving the invoice and clearing the invoice when user_065 participates. This affects 532 cases. Opening up the opportunity for an internal investigation.

This observation also holds true for the other two users. A delay of 64 days for 324 cases of user_056 between the steps RIR and CI is found.

The third significant deviation, occurring for 39% of the cases, is the occurrence of the undesirable process step RPB. In most of the cases goods are received before the vendor created the invoice and this invoice is received. Naturally payment can't occur without an invoice. However, looking closer shows that most of the time delay is caused by the 1516 cases that connect RIR and RPB. This step takes 38 days on average. Overall this deviation causes the process to take 7.9 days longer on average.

Investigation Vendor 106

The average throughput time for vendor 106 is 73 days and only 28.56% of all 6,310 cases flow through the happy path. Vendor 106 is the main vendor for metal containers and lids below 30L. In the analysis receiving the order confirmation is identified as part of the process which is mostly executed by user_29. In this process this is the only activity he performs. It decreases the throughput time on average but increases the number of steps by 1.6 per case. Cases without an order confirmation increase the duration of the process by an average of 27 days. Most likely the order handling on the vendors side is not consistent confirming the order. When the order is confirmed the goods are usually received faster. Additionally, it has been noticed that the invoice receipt for different orders is recorded more than once at the same time, sometimes by the same and sometimes by different users. In many cases, when the order is not confirmed by the vendor, first the goods and then the invoice is received. This deviates from the happy path where usually first the goods are received before the vendor creates and sends the invoice.

Investigation Vendor 182

The happy path for vendor 182 is represented by 50.31% of the cases, with an average throughput time of 111 days. The process step 'RIR to CI increases the throughput time most considerably by 81 days in 79% of the cases. An interesting finding is that the step RPB is missing here opening up the opportunity for further investigation why this step is missing.

Investigation Vendor 103

The average throughput time for vendor 103 is 108 days. The happy path is followed by 65.13% of the 3763 cases. As with vendor 182, the bottleneck is the process step RIR to CI but for this vendor increasing the throughput time even more by 99 days in 89% of the cases. Here the step RPB is missing. Another step increasing the throughput time considerably by 93 days for 10% of the cases is the process step RPB to CI. The analyzed violations for this process are, e.g. the undesired activity 'receive order confirmation'. A deeper analysis disclosed many possible root cause violations. User_029 is also accountable for 3,000 violations and therefore could be further investigated.

It appears for 92% of cases increasing the path by 6.3 steps and 108 days. Another relevant violation is that RIR is followed by CI adding 6.3 steps per case and 109 days to the process in 89% of the cases.

Investigation Vendor 150

Vendor 150 mostly sells labels and appears to be a relatively efficient vendor with almost 78 % following the happy path. The average throughput time is 61 days and most of the time is taken by the step from RIR to CI with 40 days. The step slowing down the process by 12 days is RPB 12% of all cases of the vendor. User_15 and user_6 are mostly involved in these cases. In these cases, it could be worth investigating why the payment block is instantiated and kept for several days. Additionally, there are several cases where either user_13 or user_15 record the receipt of invoice again after a long time, after it has already been done by batch_01. This indicates some error in the processing of the order since in almost all cases of these cases the next step is to remove the payment block. They appear to get lost along the process and are "reactivated" later. Vendor_150 waits a long time until reminding for payments since the next invoice is only sent after an average of 103 days and it took the vendor itself a long time until it sent the initial invoice.

Investigation Vendor 125

The happy path is represented by 52.46% of the 2,377 cases. The most frequent violation compared to the categories happy path is that CPO is followed by RGR instead of VCI. This raises the duration on average by 8.6 days and increases the complexity by only 0.6 events. The violation affects 82% of the vendors activities and a root cause analysis shows that over a third, are events in which the user_058 participates. Interestingly while the goods are received before the invoice for most of the violations with this user, barely any payment blocks occur. In general, for this violation the frequency of process steps which take longer than the average seems to be low.

Another violation, similar to the violation of vendor_104 is the occurrence of the undesirable event RPB. The observations regarding the process structure also do not deviate from the observations concerning vendor_104's violations. This causes the process to take 8.3 days longer on average, increasing the complexity by 1.6 events on average, compared to non-violating events, for 26% of the cases.

Investigation Vendor 264

For vendor_0264 65.09% of the cases flow via the happy path. Here, we want to point out the activity RGR being in front of VCI for most of the cases. The average throughput time is 98 days, and just as for almost any other vendor, the most significant bottleneck is the RIR → CI taking 51 days. No relevant violations were found. The most products are labels and Metal Containers & Lids < 30L and packaging other.

Investigation Vendor 105

The average throughput time for this vendor is 64 days, which is about half of the time of the biggest vendor. 76.85% of the cases follow the happy path. This happy path is similar to the one of e.g. Vendor_0136, however, the ROC is missing for Vendor_0136. The two largest bottlenecks are two activities that occur before clear invoice: RIR directly before CI (93% of cases) and RGR before CI (97% of cases), which extend the process by 37 and 24 days respectively. 54% of the vendor's cases that are not compliant with the Happy Path, as RPB occurs. This extends the process by three days. A root cause analysis showed that the process will also be extended if RGR occurs after CPO instead of after VCI.

3.3 Invoice, compliance and deviations investigation

As a first step in the compliance investigation a definition of a compliance deviation is required. In this paper every transaction in which the net value of a case changes for inexplicable reasons is considered a compliance irregularity. Thus, in order to identify problematic invoices, irregularities were identified in three different analysis sections.

For this analysis the otherwise established scope was disregarded. This decision has been done, as the analysis of the entire dataset is not significantly more complex with the methods chosen, than the analysis of a subsection. Consequently, the workload-oriented reasoning for a reduced scope does not apply to the dealt with sub-task.

In the first compliance analysis, we calculated the average of the cumulative net worth per case and compared it to the median worth of the case in Celonis. In case the two values would not be the same it indicates that the cumulative net worth is not the same in all events of the case.

As seen in figure 12, in total these cases account for less than two percent of the total number of cases, while they accumulate to almost 18% of the total number of events. This finding represents that these cases generate a significant amount of extra work.

The vendor with the most activities is vendor_234 with almost 50,000 activities distributed on 207 cases. Looking more closely at vendor_234, no invoice receipts have been recorded. The vendor with the most cases is vendor_233 with more than 300 cases and around 4,000 activities.

However, with regards to compliance, not all of the 4,632 cases do necessarily have a compliance issue, since several goods receipt messages or several invoice clearings could match together for a compliant process. This concludes the first analysis.

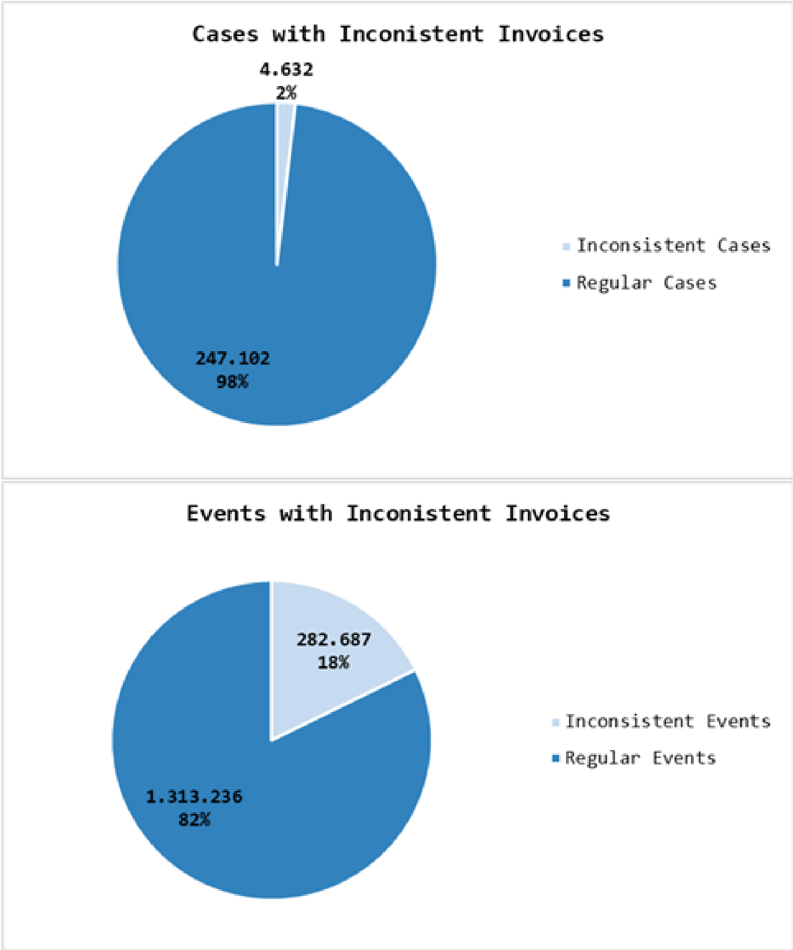


Fig. 12. Comparison of the number of events and number of cases with and without invoice issues (Own Depiction).

The second compliance investigation is conducted with Microsoft Excel and Microsoft Access of the Microsoft Office 365 suite. Before the dataset is analyzed in Excel, Access is used to reduce and compartmentalize the data into excel compatible subsets. Next these subsets are imported into Excel and potential compliance issues are identified with a rough screening of whether all individual events, of a case match in terms of net value. Through this selection step, the initial data volume is reduced by 85% to a set of 242,020 unique events or 4,632 cases.

This screening and subsequent volume reduction enabled the detailed analysis with the available computational power. In the next step of the screening process the minimal net value is identified for each case and then the sum of net values is computed for each case, followed by the calculation of the remainder of the division of the sum of net values by the minimal net value of each case. This step allows the immediate exclusion of all those cases in which e.g. the items are processed separately and then merged together causing the net value of the subsequent event to spike (deviate), even though there is no compliance issue. This step reduces the amount of potential compliance irregularity cases to 4,019, excluding 613 of the candidates. Next the remaining value of the division is divided by the count of events per case. If the result of this is less than 1, the deviation is likely caused by rounding differences and thus the candidate is excluded. This leaves 1,425 cases consisting of 39,199 events for further analysis.

As the next and final step in the selection process the remaining cases are screened for unique combinations of net value and case ID, followed by a calculation between the different events of the cases, sorted in a date based ascending order. Computing this change for cases allows for the evaluation how much deviations occurred between the different values for vendors.

The result is the identification of 28 vendors who are responsible for the possible compliance deviations. Some of these vendors are responsible for comparably many events but few cases, some for few cases but these contain comparably high net value changes. Based on this analysis a definitive categorization of compliance violating vendors is not possible, but it identifies the vendors which need to be analyzed further. A selection of the mentioned data points can be seen in table 1.

It can also be observed, that the identified, potentially violating cases are to more than 99% of the case item category “3-way match, invoice after GR” and to more than 99% concerning the case Spend area text “Logistics”. This might mean that instead of identifying potential compliance violations, the applied selection filters identified characteristics of these particular categories.

As an alternative investigation the cases containing RGR or CI are analyzed.

Table 1: Main vendors responsible for compliance deviations

VENDOR	CUMMULATIVE CHANGE (>0)	CUMMULATIVE CHANGE (<0)	AVG CUMULATIVE NV	COUNT EVENTS
vendorID_0201	2,6788E+12	0	8,92936E+11	3
vendorID_0230	2,89948E+11	-2,89946E+11	1685828475	172
vendorID_0231	410151	-1587588	23522,69712	208
vendorID_0232	5563066	-5335410	99492,82051	78
vendorID_0233	22671608	-60770369	195867,9716	352
vendorID_0234	906441	-742950	198463,2222	18
vendorID_0235	4165972	-3788390	11082,33686	1514
vendorID_0244	523375	0	66064,5	10
vendorID_0330	3,90576E+12	-4,26083E+12	4,84188E+11	11
vendorID_0363	9986920	-6219675	27912,65181	718
vendorID_0364	13173178	-7411052	34740,47342	790
vendorID_0365	7401349	-3344944	17836,12642	617
vendorID_0366	792380	-102440	2662,709877	324
vendorID_0388	0	-1,72138E+12	5,73797E+11	5
vendorID_0470	895930	-989147	7905,188153	287
vendorID_0471	3115	-2256	571,4615385	13
vendorID_0472	4148408	-4020389	28042,66392	848
vendorID_0473	18535150	-18280293	35909,28633	1383
vendorID_0534	25658848	-54941723	194569,5029	1040
vendorID_0535	2262953	-20316230	125991,9216	204
vendorID_0536	5674676	-22298552	139294,9255	443
vendorID_0537	1455558	-9008432	64520,88333	180
vendorID_0538	17399799	-76445802	369888,6975	519
vendorID_0539	10692699	-16927479	77928,1352	429
vendorID_0540	6421962	-23458070	149774,8382	346
vendorID_0541	3174116	-29793713	189900,0261	307
vendorID_0595	170	-100	51,14285714	7
vendorID_0977	300692	-304862	5073,844444	270

In the final analysis the same tools are utilized, however a different approach is chosen. Similar to the previous analysis cases are analyzed towards significant deviations between the net values but in this analysis the net values of the RGR and CI events are compared. As an initial selection we identified 51,672 cases, that have at least one RGR event, but no CI event and 870 cases with the opposite event constellation, both were excluded from further analysis. Next, if the sum and average of RGR net values deviates from the sum and average of the CI net values and the minimum RGR net value does not explain the CI net value sum or average the case is checked for potential rounding deviations. Only cases for which all these tests are conclusive, a final analysis step is conducted. This last step includes the computation of the duration and identification of the corresponding vendor.

This results in the identification of 51 unique vendors responsible for 1,003 cases consisting of 53,682 events that might represent conformance violations. Vendor_0230 is involved in 208 of these cases and 19,272 of the events, in total the deviation of the average RGR and CI net values amounts to an absolute value of 13,180,269,844€. Vendor_0230 dominates the analysis results and thus was excluded from the visualization of the deviations in figure 13.

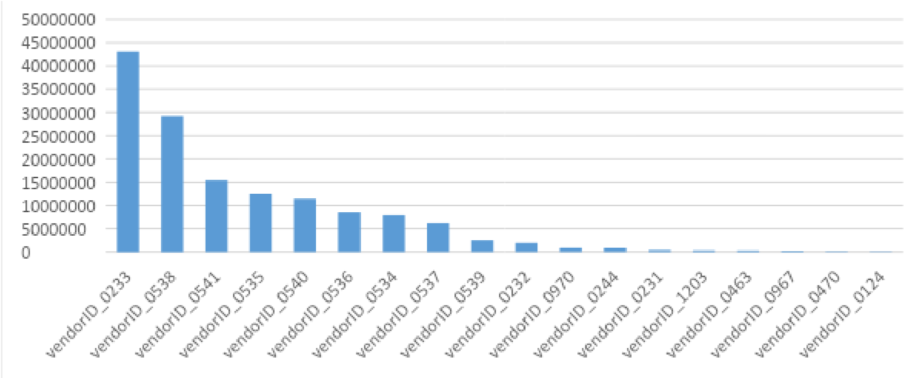


Fig. 13. Total absolute deviation with Create Invoice and Record Goods Receipt (Own Depiction).

It is noteworthy that approximately 99% of the potential compliance deviations identified in this analysis are of the document type "Standard PO" and of the case item category "3-way match, invoice after GR). Also, only 2 cases consisting of 24 events are part, less than 1%, are part of the spend area "Packaging".

In future research it might be worthwhile to conduct a more detailed analysis of these 51 vendors, some data points concerning the potential compliance deviations grouped by vendors can be found in table 2.

Table 2: Vendors with potential compliance deviation with focus on Create Invoice and Record Goods Receipt.

VENDOR	ABS DEV OF AVG NV	AVG DURATION IN DAYS	COUNT CASES	COUNT EVENTS
vendorID_0230	13180269844	1	208	19272
vendorID_0233	43131753,33	36	73	165
vendorID_0538	29226585,29	7	90	514
vendorID_0541	15433107,28	9	48	234
vendorID_0535	12485612,83	13	48	146
vendorID_0540	11420484,64	4	51	282
vendorID_0536	8492909,884	5	47	407
vendorID_0534	7875849,22	1	47	1594
vendorID_0537	6213041,5	24	42	124
vendorID_0539	2510849,718	3	64	1057
vendorID_0232	1953079	43	18	47
vendorID_0970	891915,6923	14	1	24
vendorID_0244	875152	11	11	22
vendorID_0231	395524,4459	8	101	1443
vendorID_1203	303109	18	2	4
vendorID_0463	271586	23	2	4
vendorID_0967	216576	11	3	6
vendorID_0470	133337,9218	8	80	1174
vendorID_0124	107339	40	3	6
vendorID_0812	80645	38	2	4
vendorID_0295	69785	41	2	4
vendorID_0433	60265	36	3	6
vendorID_0183	58050	30	2	4
vendorID_0684	56119	43	2	4
vendorID_0514	42535	40	1	2
vendorID_0279	41093	36	2	4
vendorID_1344	32445,66667	3	1	10
vendorID_0117	31870	28	1	2
vendorID_0658	26025	36	1	2
vendorID_0166	25548,33333	21	2	7
vendorID_0653	23944	18	1	2
vendorID_0271	21733	11	1	2
vendorID_0502	21040,5	29	3	9
vendorID_0483	16550	32	2	4
vendorID_0141	16420	32	1	2
vendorID_0314	11867	34	3	6
vendorID_0171	11601	17	1	2
vendorID_0496	11068	30	1	3
vendorID_0461	8678	45	1	2
vendorID_0195	6826	36	1	2
vendorID_0062	6287	42	1	2
vendorID_0712	6100	10	1	2
vendorID_0467	3880,903656	2	6	101
vendorID_0151	3457,578947	3	1	67
vendorID_0595	2986,5	40	12	32
vendorID_1094	2663	12	1	2
vendorID_0125	1939,75	24	1	7
vendorID_0214	1359	39	1	2
vendorID_0111	555,8571429	12	1	8
vendorID_0119	520	11	4	8
vendorID_0182	440	54	1	2

4 Acknowledgements

All Authors contributed equally to this paper. We would like to express our gratitude to all sponsors for making the BPI Case Challenge 2019 possible, thus facilitating creative approaches for real Process Mining problems. We would especially like to thank Celonis SE for providing us with licenses of their Process Mining Software which was our main tool to conduct our findings.

5 Conclusion

This paper outlines the underlying purchase process flows of a Dutch company operating in the field of coatings and paints. We analyzed and depicted the different processes in order to generate a solid understanding of the company's processes. As a first step irregularities in the data sample were identified and used to define a scope for further analysis. Based on the selected scope and research question, a portion of the data was analyzed with Celonis. The previously defined scope was applied, and the resulting process-categories were documented modeled to establish a baseline. In the further analysis, a specification was performed, as the student category inherits the possibility to focus on particular aspects. To ensure a structured and goal-oriented approach we formulated a research question to orientate and navigate through the upcoming challenges: "Are there any optimization possibilities in the area of compliance, throughput time regarding the top 10 vendors? ".

The investigation of the throughput times of the invoicing process showed, that the activity flow from RGR to CI is with great distance the most often occurring bottleneck. This opens up the possibility for further internal investigation, whether the payment process is supposed to take such a long time.

Throughout the analysis process several challenges had to be overcome. One of these is the uniqueness of most processes and as such the difficulty of abstracting and summarizing without losing sight of important details. The analysis section is based on the assumption, that seemingly inexplicable deviations in the net value can be used to identify compliance violations. In total three analysis methods were used to identify potentially deviating cases and subsequently grouping the results by the corresponding vendors. Especially infrequent or unique cases can be covered by the sheer number of cases. Finally drawing on the established models, the scope and with regard to the research question, the purchase documents of particular vendors were selected in a deep analysis. Here several irregularities were observed. For example, do Vendor_136, Vendor_104 and Vendor_106 follow the happy path to a comparatively low degree. Additionally, a root cause analysis showed that the process will also be extended if RGR occurs after CPO instead of after VCI.

Again, the process after the invoice is received until the invoice is paid takes a long time in almost all cases. Finally, we found for the top ten vendors in the packaging spend area, only minor issues with regard to invoice conformance occur.

Reviewing the entire analysis process leads to the conclusion, that the vast amount of cases from many different spend areas is challenging to oversee and understand. For some cases it could be helpful to have background knowledge making connections between activities easier to understand. Generally, it makes sense to differentiate between the four types of data flows. However, it could be useful to further differentiate and structure according to the most important vendors so that rework especially in the payment process is limited as much as possible.

References

1. Aalst, Wil M. P. & Weijters, A. (2004). Process mining: A research agenda. Computers in Industry. 53. 231-244.
2. Aalst, van der, W. M. P. (2011). Process mining: discovery, conformance and enhancement of business processes. Berlin: Springer.
3. Tiwari, A., Turner, C. J, & Majeed, B. (2008). A review of business process mining: State-of-the-art and future trends. Business Process Management Journal, 14(1), 5-22.
4. Prodel, M., Augusto, V., Jouaneton, B., Lamarsalle, L., & Xie, X. (2018). Optimal Process Mining for Large and Complex Event Logs. IEEE Transactions on Automation Science and Engineering, 15(3), 1309-1325.

Business Process Intelligence Challenge 2019 – a contribute

Lorenzo Botti
lorenzo.botti@logtovalue.com

Abstract

Process mining technology is there but most businesses are still wary of its adoption, because many data-based insights are made available by business intelligence. This contribute aims to show how both approaches are needed and how process-insights can improve decision making and representation of business performance through KPIs. This Challenge is a useful practice to highlight process mining advantages, but it is not a process mining project, so many assumptions will be left unchecked and more questions will be raised than answered given the lack of confrontation with the Client.

1 Introduction

The available dataset[*] is a collection of cases[†] from a Purchase-to-Pay process, that take place - or at least begin - in 2018[‡]. A thorough clean-up and then filtering out of cases is required.

A critical approach about the dataset is followed in order to analyze both visible information (cases, activities etc.) and the set of information that lies behind (assumptions on ETL phase and on the real-life daily routine of the process), as well as "noise" in the event-log.

The analysis encompasses business-intelligence style analytics and process mining, so to show at what extent both approaches can thrive from a synergy between them, and how businesses can fact-check their KPIs set with process-based evidences. While pursuing a suitable set of process models for the most part of traces, peculiar behavior of subsets will be noted and discussed.

For the purpose of the analysis and to the best knowledge of the dataset, it is assumed that data had been collected from an ERP-like system (or a multiplicity of systems) on which both internal users and external ones – vendors - can operate. The latter ones either directly or intermediated by back-office staff.

---

[*] van Dongen, B.F., Dataset BPI Challenge 2019. 4TU.Centre for Research Data. https://doi.org/10.4121/uuid:d06aff4b-79f0-45e6-8ec8-e19730c248f1

[†] in the text terms "case", "process instance" and "trace" and occasionally also "sequence" will address the same meaning

[‡] https://icpmconference.org/icpm-2019/contests-challenges/bpi-challenge-2019/

2 A look at the dataset

A first look at data shows some peculiar features that call for a clean-up. It is however necessary to retain all the useful information and potential insights.

2.1 Calendar

The dataset refers to a Purchase to Pay process taking place in 2018, that is beginning in 2018. Last events should be assumed no further that date of dataset creation (January 2019).

Some activities have been recorded with incongruent calendar dates, spanning from 1948 to 2017, and beyond January 2019 until April 2020 (Table 1).

Traces	251.734
Events	1.595.923
First Event	1948-01-26T23:59:00Z
Last Event	2020-04-09T23:59:00Z

Table 1: original dataset

These activities are almost uniquely related to vendor actions (Vendor creates debit memo, Vendor creates invoice). Just few are due to "Create Purchase" activities that take place in 2017.

It is possible to make two hypotheses: a deliberate or careless behavior on one side, an IT-related glitch on the other side (maybe due to differences in machine set-up of dates and times across different terminals or user interfaces). A third, most probable, hypothesis – induced noise in the event log – cannot be either excluded either investigated, so just the other two are considered.

As for the first hypothesis, involved vendors also participate to other cases where no calendar issue takes place, so it is possible to say that they are aware of how to process information and upload them into the system. For the same reason, not even the terminal set-ups can be associated to the issue. Besides, Vendor-related activities ("Vendor creates invoice" and "Vendor creates debit memo") are all performed only by an user identified as "NONE" and always at 23:59; that is, similar activities related to different instances of the process are uploaded as a batch operation onto the "ERP-like" system at the cut-off time of the day (it is an usual feature of multisite ERP-like systems to go through a periodic planned update of all the tables associated with the underlying database). The fact that even the "out of calendar" vendor activities had been performed at 23:59 tells that either the system failed on recording the date of execution (but not the time) or that the associated date is an invoice-based input and not a system one. The latter hypothesis is not sound with any practice of ERP architecture so only the first one is to be considered. Therefore, an IT-glitch is probably to be investigated for the "out of calendar" events; this glitch happened either at "run time" (when data are recorded by the ERP-like system) or even at the time of data extraction and transformation (the ETL phase for preparation of event log).

The amount of cases is about 0.1 percent (271 line-items) and no recurrent features can be found (spend area, company etc.). It is therefore possible to clean out these cases without fear of any information loss toward the process itself. Few instances originated in the last month of 2017 and then completed the process in the first months of 208, but their small number still allows for them being excluded by the dataset.

Note that very few of those cases (7 out of 271) are due to "create purchase requisition item" (therefore executed by internal users) and span along the year 2017. It could be the case of an "error" or a business accepted process drift (creating the document "in the system" in the afterward of other activities that should be consequent to the document itself)[§].

[§] If this is true, nonetheless any user interface should have a control check on dates associated to an event, with a specific procedure when an input date is different from the current (machine) one.

2.2 Follow the money

An important feature stands out when looking at the very essence of any Purchase to Pay process: what are we paying for? Cases are classified with three attributes that refer to Spend classification categories (Classification, Area text, Sub Spend Area Text). Among the "Area text" there is an "Spend Area Unidentified" to deal with unforeseen or unchecked spend areas (36 traces belong to this classification). It is surprising to see though (Figure 1) that there are 3289 traces associated to a *blank* classification. Not only there is no classification, but at a deeper analysis no peculiar feature can be associated with these cases (vendor, user, item category etc.) apart that 98% of them are referred to as a "Standard" Item Type (which are anyway the most common traces in the set, ca 90%).

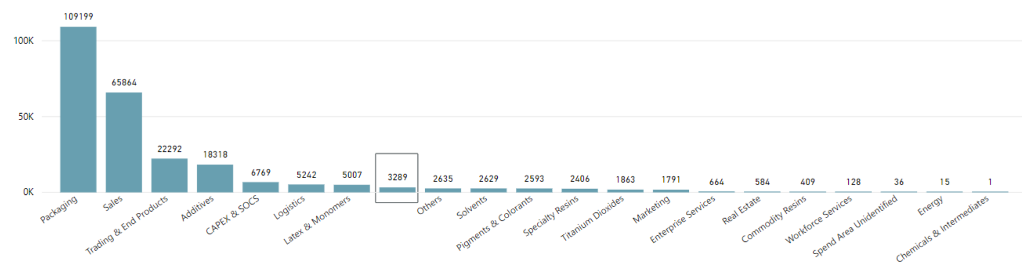


Figure 1 Number of traces associated to a Spend Area

Most of these anonymous traces are "complete" (for 1622 traces there is at least one "Clear invoice" activity). No actionable insight can be found or assumed at this point. If it is not due to a deliberate action (e.g.: to introduce noise in the event-log), then it would be matter for an accounting audit[**]. These cases are better excluded from a first analysis[††] even though other attributes are available to fit those traces into different process categories (Item Category, Item Type etc.).

3 The working set, from intelligence to mining

The clean-up leaves us with a dataset built on 74.851 Purchasing Documents (Vs 76.349) and 248.177 cases (Vs 251.734). The working dataset will be subject to further filtering, but as a first step, a look at the dimensions into which the dataset can be broken down already provides some useful evidence. What follows is a just a partial business intelligence analysis, that can be extended and is surely summed up in dashboards and reports. The focus here is to show how some features can lead us toward a process mining approach. This will support the mining itself and will help to verify assumptions.

Process mining relies on timestamps, the date and time reference of every activity. In the dataset there are some activities, within the same trace, that share unique timestamps. The fact here is not about the precision of the reference (seconds, not microseconds), but to the reality of a "stamp". Activities within a trace that share the same timestamps are to be assumed as executed by "ticking boxes" on a user interface screen-view (or by a batch-user) and then "time-stamped" at the moment of execution of the screen-view (eg.:"save", "next page"…). In such hypothesis, the sequence of these

[**] A rough and inaccurate estimate of the value involved can be given with two figures: the EUR amount of the sum of "Create Purchase Order" is about 25 EUR Million, while the sum of "Clear Invoice" is about 15 EUR Million.

[††] After the clean-up, the count of unique users will decrease by 6 units (out of 627) and that of unique vendors will decrease by 45 (out of 1975); these are anyway just a tiny portion of those involved in the wiped-out cases (262 users, 183 vendors), so no specific actions on user training or vendor contract management would address the situation.

events will be recorded as per process-design and can help explore assumed model (those developed by IT according to business requirements, at least), even if for limited number of activities (such is the case of SRM-likes that belong to E.C. Purchase Order Document Type, and are present in both "3 way Before" and "3 way After" dataflows). Other activities simply share the same timestamp because they are repetitions and related traces will need a deeper investigation in order to understand if those repetitions are due or not in the process.

3.1 Companies, organization and users

The available dataset is supposedly the collection of cases from a Parent Company headquartered in the Netherlands and some of its sixty subsidiaries worldwide. No information is available on the organization, but process instances can reveal some insights.

Traces are associated to four different Companies (Table 2). Most of traces are operated within CompanyID000, while CompanyID_003 shows a specialization ("2_way_match").

	CompanyID000	CompanyID001	CompanyID002	CompanyID003
2_way_match	-	-	-	995
3_way_match_after	14831	2	-	-
3_way_match_before	217874	-	2	-
Consignment	14473	-	-	-

Table 2 Distribution of cases by Companies and Item Category

CompanyID000 is associated to over 99% of instances. Either the Company is an aggregation of some subsidiaries or represent sort of Business Unit (by this approach CompanyID001 and CompanyID002 could be considered "noise"). By breaking down through the Spend Area attribute of the events, CompanyID003 is probably a Service Company to other subsidiaries: in fact, "energy" related purchases are exclusive to this company, just like most of "real estate" ones.

Focusing on users, CompanyID003 has 19 users, among which 4 are exclusive to the company (602, 603, 604, 606) and three of these show a clear specialization: 604 for clearing invoices, 603 in conjunction with 602 for creating and approving Purchase Orders. The relationship between the latter two users (603 and 602) will also have an impact on the discoverable process model, mostly due to the different meaning of an activity they both will perform (*Change approval for purchase order*).

The other companies share 618 users. A sub aggregation of CompanyID000 is still possible through a social analysis based on *handover of work*. In fact, working on "consignment" traces there are three sub-groups of users (out of 155 users) that never interact among each other. Consignment traces have the peculiarity that users probably operate at local companies' plant/warehouse locations (Goods-related activities); besides, these users, even in other Item-category traces (3 way After, 3 way Before) execute only the same set of activities. In the end, the three sub-groups could refer to the number of locations or subsidiaries involved[‡‡].

3.2 Cumulative Net Worth (EUR), matching activities and data check

The dataset provides a "Cumulative net worth" figure for every activity within any process instance[§§]. There can be the case of multiple invoices and good receipts within the same trace, but they all have the same cumulative net worth, which indeed is a cumulative figure not an atomic one; for example, the twelve *goods receipts* of trace 2000000055_01 share the same value (43,312.00 EUR) and this mean that each one would have been on average 3,609.00 EUR. This is verifiable for those spend areas for which multiple invoices are expected (workforce, real estate etc.). The matter is

[‡‡] "Consignment" users do not take part to any "2 way match" traces, as expected, being Goods activities not included in these traces; this confirm "specialization" of these users and the peculiar nature of "2 way match"/CompanyID02

[§§] Exception made for "consignment" traces that each trace should have a unique value of cumulative net worth, as it happens indeed for 98% of traces (Table 3). The compliance assumption is that activities should be matched against the "cumulative net worth" figure (namely, *good receipt* and *invoices*), but for 98% of traces it is only possible to *count* activities not to *match* their values. Looking at the residual 2% of traces with more than one EUR value it is possible to question which activities are matched within the process and raise some questions about event-log preparation and user interface controls.

Total traces (original dataset), units	251.734
Traces with one unique value associated to the process instance (units)	247.102
Traces with 2 to 75 values associated to the process instance, (units)	4.632

Table 3 number of process instances aggregated toward the numerosity of EUR figures through the instance

First, we look *where* traces with multiple EUR values occur, and it is clear they are mostly confined in specific spend areas (Figure 2) and fall into "3 way After" (86%) and "3 way Before" (14%) categories and over 82% are related to *Service* Item Type.

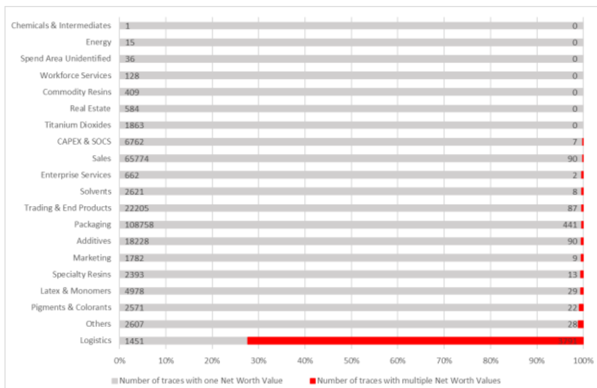


Figure 2 – Number of traces with one or multiple Net Worth values (across Spend Area)

Looking at the process, there are many different sequences and patterns (for example: Figure 3) that require some questioning. Given that the reference is a dataset originated from an ERP-like database, figures could have been affected by data transformation (were original database figures atomic or cumulative?), and matching activities could have been impacted.

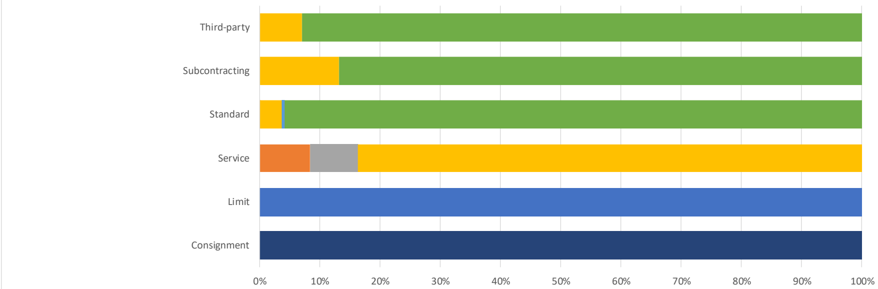
_case_concept_name_	_event_concept_name_	_event_time_timestamp_	_event_Cumulative_net_worth__EUR_	_event_User_
4507006350_00001	Create Purchase Order Item	30/01/2018 13:51:00	276,00	user_200
4507006350_00001	Record Goods Receipt	30/01/2018 13:51:00	552,00	user_200
4507006350_00001	Record Service Entry Sheet	30/01/2018 13:51:00	276,00	NONE
4507006350_00001	Record Goods Receipt	31/01/2018 11:19:00	18.798,00	user_200
4507006350_00001	Record Service Entry Sheet	31/01/2018 11:19:00	276,00	NONE
4507006350_00001	Vendor creates invoice	06/02/2018 23:59:00	276,00	NONE
4507006350_00001	Record Invoice Receipt	12/03/2018 17:36:00	552,00	user_001
4507006350_00001	Clear Invoice	29/03/2018 15:11:00	276,00	user_002

Figure 3 – multiple values within a trace, a case of anomaly for "Record Goods receipt"

"Goods receipt" value apparently is *independent* from other activities, while value matching activities would fall in two distinct sets: first one with "Create P.O."/"Vendor Creates Invoice"/"Clear Invoice", and second one with "Record Service Entry sheet"/"Record Invoice receipt" (the Invoice receipt value being the sum of Service sheet entries). In the end some questions can be raised about: the origin of input for "goods receipt", the reason for manifold and instantaneous repetitions of both "Record service entry sheet" and "Record Goods receipt", the input for matching existing invoices and related "Clear Invoice" activity.

3.3 Dataflows, documents and items: dimensions for modeling

Process instances can follow at least four main data flows and complexity grows according to Document Type and Item Type (Figure 4). While Item Categories, as described, allow some speculation about underlying process model, there are useful insights based on the other two dimensions. For example, some set of activities are specific to document type (*SRM-…* for E.C. Purchase Order). Another example is that some *Document type* traces (Standard P.O.) show a process drift in the period. The complexity of "Service" traces only belongs to "3 way After", as well as unique Item Type and Document Type are peculiar for both "Consignment" and "2 way match" data-flows.



	Consignment	Limit	Service	Standard	Subcontracting	Third-party
2-way match - Framework order		995				
3-way match, invoice after GR - EC Purchase order			481	113		
3-way match, invoice after GR - Framework order			458			
3-way match, invoice after GR - Standard PO			4842	7939	616	384
3-way match, invoice before GR - EC Purchase order				820		
3-way match, invoice before GR - Standard PO				207902	4047	5106
Consignment - Standard PO	14473					

Figure 4 – Distribution of traces across Item Type, Document Type and Item Category

3.4 Users and activities, pattern anomaly

There are over six hundred users that take part to the process in the period and about 1.5 million activities. As shown in *Figure 5*, there is a sharp increase of users in the last months of the year. From an average of 281 users between January and August, the count passes to an average of 396 users between September and December. This increase is apparently not related to an increase of activities performed; by breaking down these figures to Item Categories, the sharp increase in number of users is clearly linked to traces associated with a "3 way Before" data flow and particularly with traces associated to "Standard P.O." document type (*see* 3.3). These are the same traces that will show a higher not-compliance, so it would be interesting to discuss if more resources had been spent to deal with "complexity" or if the more resources involved were the cause of such complexity due to training issues.

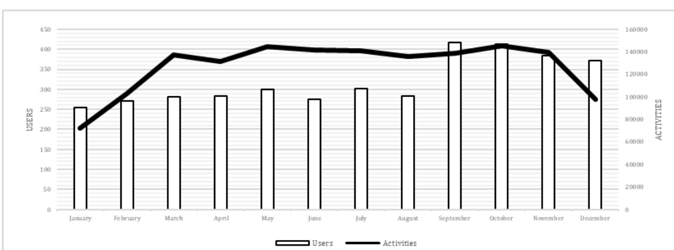


Figure 5: users and activities per month (2018)

Besides, looking at the process - *who does what and when* along the instance - there are some peculiarities with some (1401) of the traces that belong to a subset with a "debit memo" event.

In brief they show a pattern as in in Figure 6 (highlighted and boxed part).

12/01/2018 07:33:00	user_038	Create Purchase Order Item
12/01/2018 23:59:00	NONE	Vendor creates debit memo
12/01/2018 23:59:00	NONE	Vendor creates invoice
23/01/2018 16:35:00	user_007	Record Invoice Receipt
23/01/2018 17:39:00	user_029	Record Goods Receipt
24/01/2018 09:40:00	user_006	Cancel Invoice Receipt
24/01/2018 09:41:00	user_006	Clear Invoice
24/01/2018 09:43:00	user_006	Record Invoice Receipt
01/02/2018 12:01:00	user_002	Clear Invoice

Figure 6 – peculiar pattern on Invoice

The box-highlighted pattern (same user, anomaly sequence in few minutes) occurs in any of the potential dataflows (obviously not for *Consignment* traces). The role of Debit memo would be an interesting matter of discussion with management and IT department[***]. Indeed, any *Debit Memo* is followed by *invoice* related activities (*cancel*, *clear* etc.), with no helpful clues to discriminate these (debit memo) activities from the ones related to "actual" invoices. The fact that boxed activities in Figure 6 are executed by the same user in matter of minutes implies a probable business rule that would need attention.

A smaller subset (377 traces) shows a similar pattern with Goods related activities (Figure 7). In this case, the time interval stretches from minutes to days, but still the same user performs the activities.

Activity	Resource	Date	Time
Create Purchase Order Item	user_079	03.01.2018	15:14:00
Vendor creates invoice	NONE	22.01.2018	23:59:00
Record Goods Receipt	user_094	23.01.2018	12:09:00
Cancel Goods Receipt	user_094	30.01.2018	08:14:00
Record Goods Receipt	user_094	30.01.2018	13:06:00
Record Invoice Receipt	user_015	19.07.2018	09:32:00
Clear Invoice	user_002	19.07.2018	13:16:00

Figure 7 – peculiar pattern on Goods receipt

Some assumptions[†††] can be made to explain the above patterns (operational mistake, intent etc.) and a deeper investigation with the Client is needed. Still, a look at the handling of Logistics P.O. (road packed) is advisable (three case_vendors cover 44% of traces: 534, 388, 472).

3.5 Process drift and forced execution

A process can change along a period for many reasons: workload, seasonality, business changes, etc... When mining process data, a drift is then spotted. In the dataset, at least two changes apparently took place in the process and they can be categorized as it follows:

- Activity-based process change (a change in sequence)
- Resource-based process change (a change in resource for specific activities)

[***] According to accounting practices and evidences in the dataset, a debit memo – formally raised by a Vendor in this case – can be triggered either by the Company or by the Vendor. In the latter case, it acts as an incremental invoicing or a note that there are unpaid invoices/expenses; a Company can trigger it directly or indirectly, as it is with some instances that show "debit memo" occurring after some Goods-adjustment (cancel, change price etc) even in absence of an invoice, probably meaning that the memo is related to a part of expenses (freight…) owed to the Vendor even if goods have been refused (etc.). Cases can be multiple, and they all are supported by the dataset. The handling of any instance could improve by using different names (Cancel Memo/Cancel Invoice, Clear memo/Clear Invoice…) with the advantage of process compliance and KPIs, at the expense of few more tables and relations in the database.

[†††] For cases with adequate time interval, similar behavior can occur in processes where there is a performance-based reward system in place and some activities are performed just to "re-start" the clock; or it could be a sign of a work in progress audit on goods receipt. Where time interval is just a matter of minute, it could be an operational mistake, the cause of which deserve attention

As for the first one, apparently there has been a change in the start events of traces within the category of Standard P.O. document type (Figure 8) between Jan-Aug and Sept-Dec. Within the two sub-periods the total number of Standard P.O. traces are almost equal, but the difference in start events of traces show a change in process. It is certainly not a matter of resource training and it is not associated to specific vendor or item. The scale of the phenomenon can be a consequence of a business decision (procurement mgmt. to monitor the creation of new P.O., Organization acting on performance-based metrics, etc.) or even related to technical requirements (a change in database structure or administration). A discussion with management could help to justify the drift.

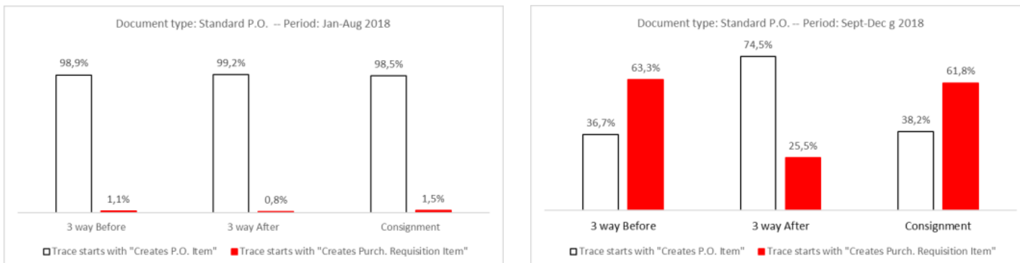


Figure 8 – change in start events distribution along the period

As for the resource-based process change, something peculiar happened with "Clear Invoice" on a specific date. It is not a real process drift and it is better described as a forced action, probably executed in order to "close" cases. If so, it tells that system can be forced and this should happen only under extraordinary circumstances - and under strict supervision.

"Clear invoice" activity can be interpreted as an authorization, not the material payment action, which is probably done by accounting department in a subsequent process. Even as such, "Clear invoice" is mostly executed by few resources with specific authorization profile within the organization. Indeed, for "2 way match" 84% of "Clear invoice" are executed by the same resource (604), whereas for both "3 way Before" and "3 way After" on average over 85% of times "Clear Invoice" is executed by just one resource (002). These "Clear invoice" events happen *in batches*, and this would fit with the hypothesis of a "middle-manager" that periodically releases cases worked-on by others (a typical arrangement of segregation of duties), by *approving* the "Clear Invoice".

Overall, this is constant across the entire period. Then, at the beginning of January 2019 (between Jan 2nd and 7th, with a peak on the 4th), a subset of 7868 traces end with a "Clear invoice" activity executed by "None"[‡‡‡]. In brief, a batch of thousands of process instances supposedly have been forcefully "cleared". The reason is understandable, but these extraordinary measures should be carefully monitored and properly authorized by management.

4 Process models, trace completeness and compliance

Before exploring compliant traces, it is worth mentioning a subset of 1410 traces which all end with a "Record Invoice receipt" activity, but do not show any prior "Vendor creates Invoice/Debit memo". These traces are not complete – no "Clear invoice" – so formally they can't possibly be categorized as not-compliant, nonetheless, the absence of any "Creates invoice" within these traces shows a lack of compliance related to expected sequence. These traces belong to *Sub-contracting* (45%) and *Standard* (42%) item type and what is more important, they have the "Record invoice receipt" activity executed 90% of times by "batch01", while in the entire set, this activity is usually

‡‡‡ During the same period over nine thousand traces end with the same activity executed by actual users (not by "None").

executed (over 90%) by *user_X* (human users). Thus, it could be a case of process change (as in *3.5 Process drift and forced execution*) to be investigated, at least because the last event take place months after the second-to-last event.

The dataset is a snapshot of what happened in a certain period, and as such it is a collection of process instances, either partial or complete. For the discovery of process model(s) only complete traces will be investigated. Assumptions are needed about the meaning of "complete".

The minimal assumption for completeness is about a set of mandatory activities, which can be grouped according to dataflows.

3 way Before	3 way After	2 way	Consignment
Create P.O.	Create P.O.	Create P.O.	Create P.O.
Record Goods Receipt	Record good Receipt		Record Goods receipt
Vendor Creates Invoice	Vendor Creates Invoice	Vendor Creates Invoice	
Record Invoice Receipt	Record Invoice Receipt	Record Invoice receipt	
(Remove Payment Block)			
Clear Invoice	Clear Invoice	Clear Invoice	

Table 4 – Mandatory activities for traces "completeness" according to dataflow (not in sequence order)

Focusing on traces with a minimum set of activities for trace completeness (Table 4), then, compliance must be investigated, along two dimensions: process-compliance and EUR-value compliance.

As per the latter, we have seen (Figure 3) that about a third of "Logistics" traces could be apparently not-compliant based on Goods Receipt EUR value; they are anyway compliant (through "Record Service entry sheet", thus raising questions on where the EUR-value information come from). The total number of these traces allow for them to be excluded from compliance analysis.

Process-compliance can be just about the presence of mandatory activities or even about an expected sequence, with different impact through different dataflows.

The most critical path (and with most of traces in the period) is the "3 way Before". Process instances associated to this dataflow belong to two different Document Types (E.C. Purchase Order, Standard P.O.).

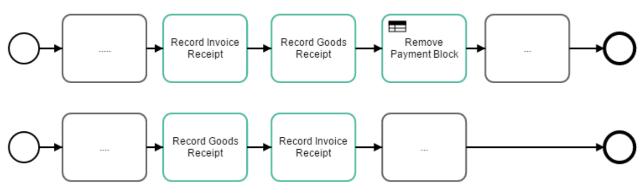


Figure 9 – "3 way Before" traces, process-compliance basic sequences

As in Figure 9, for these traces to be process-compliant there are two chances: if a "Record Invoice Receipt" precedes a "Record Good Receipt", then a "Remove Payment Block" must follow, while the latter is not needed if "Record Goods receipt" precedes "Record Invoice Receipt". Besides, if required by the sequence (as in upper part of Figure 9), "Remove Payment block" must precede any invoice clearing.

Based on these assumptions, not-compliant instances (Figure 10), even if present in small numbers, show severe violations, caused by:

- a lack of business rule (the system neither imposes a strict sequence-compliance neither requires mandatory activities)

- a lack of monitoring (segregation of duties should apply, preventing any invoice-clearing in violation of rules).

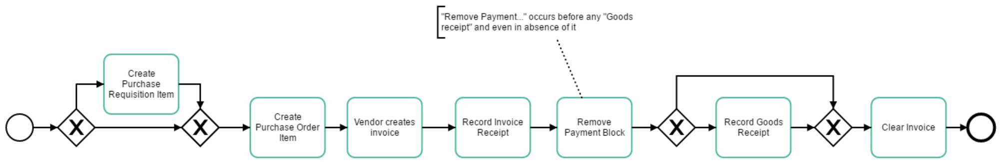


Figure 10 – "3 way before" not compliant traces (358 traces, with 77 traces without "Record Goods receipt")

It's worth mentioning that "Remove Payment Block" is present also in a subset of traces where "Record Goods receipt" precedes (lower part of Figure 9) "Record Invoice Receipt" (it happens in about 30 thousand cases, while it doesn't in about 120 thousand cases). It would be interesting to discuss which business rules apply and if there are more than one: is Remove Payment required or not only according to the sequence (Figure 9), or do other conditional rules apply (vendor, items, EUR-values…)? In case of a set of rules, is there a hierarchy?

As for "3 way After" traces, process compliance requires the presence of few mandatory activities, and with no strict sequence rules (the one in Figure 11 only cover 40% of *invoice-cleared* process instances).

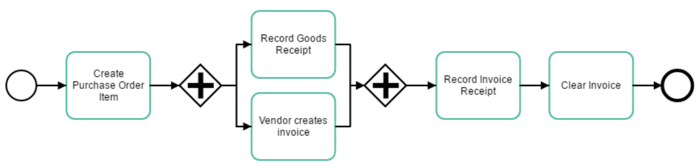


Figure 11 – "3 way After" traces, minimum set of mandatory activities (weak assumption on sequence)

This category of traces mostly shows compliance within the full subset (Table 5).

3 way After	Compliant	Compliant subset	Not Compliant
Standard P.O.	8645	1885	(149)
E.C. Purch. Order	436	20	-
Framework Order	339	56	-

Table 5 – Compliance of traces based on activities and EUR values

There is a subset of compliant traces that include an unexpected activity: "Remove payment block". All traces have the GR-based set at *true*. If this is correct, then, such a subset would require some questioning about the rules behind the payment block (which must be different from that in "3 way Before" instances). On the other hand, the GR-based value could have been wrongly set (if it were *false*, those instances would belong to "3 way Before" and would be fully compliant).

Few traces are not compliant based on Eur-value, but the statement must be proved. It is already known (Figure 3) that "Record Goods Receipt" can differ in values as well as "Record Invoice receipt". On the other hand, "Create P.O." and "Vendor creates invoice" have same EUR value of "Clear Invoice". In the end, value-based compliance is full if referred only to EUR value of Vendor Creates Invoice, Clear Invoice and Create P.O., and mostly full if also Goods receipt and Invoice receipt are considered. The anomalies should be investigated though, and questions be raised about which is the actual input or information that "tells" the operator which EUR values are associated to Goods Receipt and Invoice Receipt.

As for traces that follow a “2 way-match” dataflow, the complete ones (286) apparently are compliant based on presence (not sequence) of activities and EUR values. A process perspective aims also to look at *who does what* and this will reveal a probable *four-eyes principle* in the set. Most of traces are incomplete, even if active since the first months of 2018. This is in line with a Framework Order document type (100% of *2 way match* traces), but still could be a matter for audit, even if it happens that Year1 purchase orders are cleared at the beginning of Year2 (dataset was made available on Jan. 19th).

Consignment traces are mostly complete (13456) and fully compliant. They are mostly affected by the change in process (see 3.5) at the start of any instance, but this doesn’t impact on the model, which is actually a basic one.

5 Process models

Models are investigated for most common set of traces within different dataflows (Figure 4) and main variants, thus, for example, different Start events will be considered only if they have a strong impact on the sequence. Peculiar behavior already mentioned in the text (debit memo, same-user activities etc.) is not included in the analysis, which aims to general schemas or subprocesses.

6 “2 way match – Framework order – Limit”

These set of trace reveals a need for confrontation with the Client and this is mostly due to the fact that the same activity - *Change Approval for Purchase order* - is performed by two users on the basis of an apparent “four eyes” principle (*user 603* as a supervisor, *user 602* as an executive). Therefore, discovered model shows a loop that should be better *untangled* by identifying two different activities in place of one, for example: Change Approval-Supervisor, Change Approval-Executive. Process instances can be triggered by three start events, with different implications for the sequence of activities (Figure 12, Figure 13, Figure 14).

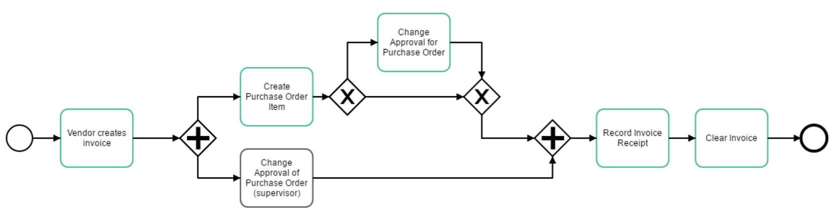


Figure 12 – “2 way match” Process instance starts with *Vendor creates invoice* (basic sequence, no events looked at after “Clear invoice”)

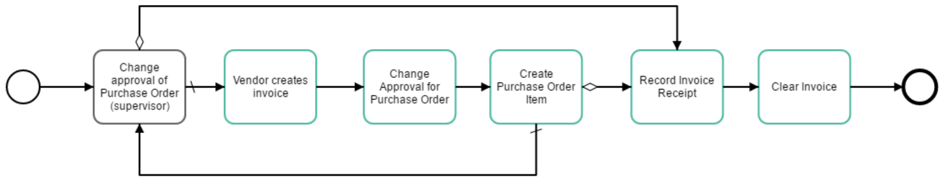


Figure 13 – “2 way match” Process instance starts with *Change Approval … (by supervisor)*, basic sequence

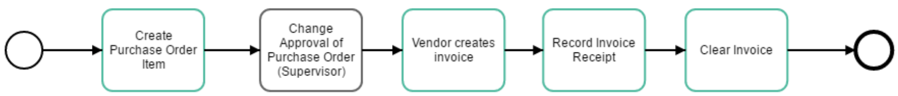


Figure 14 – “2 way match” process instance starts with *Create Purchase Order,* basic sequence

The “four eyes” principle acts also in other phases of the sequence, usually after “Clear invoice” (namely, a sequence of “Change Approval..”, with the last one always executed by the *supervisor*) as if it were an authorization step even after months from Clear Invoice; this can be explained with both Item Type and Document Type associated with these traces. A framework order can have one or more invoices, with intermediate activities that act as periodic “authorization steps” to the payment or update to the payment accounting process. It would help discussing with the Client, given that these traces are mostly related to Real Estate and Energy spend areas.

7 “3 way match – After”

7.1 E.C. Purchase Order, Service (and Standard)

In *3.3 Dataflows, documents and items: dimensions for modeling* we have learnt that some activities belong only to peculiar Document Type, as it is the case of E.C. Purchase Order. There is a set of activities, executed by a *batch* type of user usually in conjunction with Create Purchase Order (this one executed by any *user_X*). The “SRM…” subprocess in Figure 15 can be triggered by a user_X that creates a Purchase order or by a (None) Vendor that creates an Invoice.

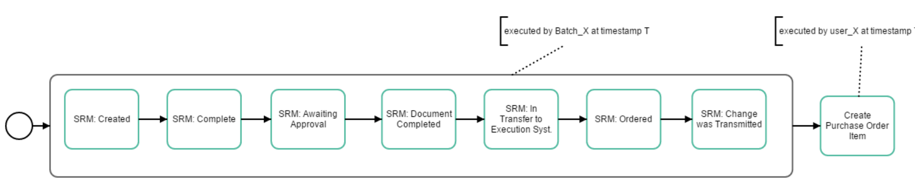


Figure 15 – “3 way After”, EC Purchase Order, Service (Part 1)

the “SRM…” sub process in Figure 16 instead occurs along the instance lifetime in two cases: after a “Vendor Creates Invoice” due to a probable mismatch of values between Invoice and Purchase Order, or after “Clear invoice” due to the fact that “Service” Item type can go through more than one invoice.

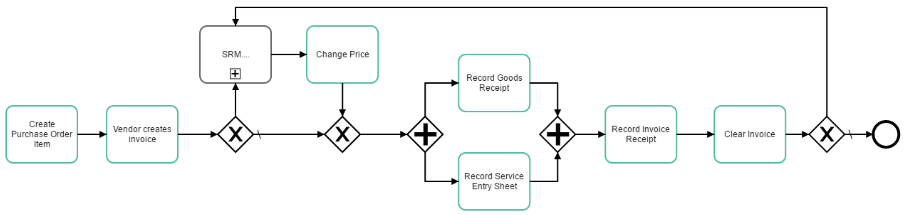


Figure 16 – "3 way After", EC Purchase Order, Service (Part 2)

There is another difference about "SRM…" subprocesses in Part 1 and Part 2 and it is the executor. In Part 1, the subprocess is always executed by a *batch_X*, while in Part 2, the first steps of subprocess (from "SRM: Created" to "SRM: Document Completed") are executed by a *user_X* and then by a *batch_X*. It would be interesting to discuss whether a business rule could be implemented in order to automate the triggering of the subprocess.

Note that the sub-process can have more activities (the depicted ones are just the most frequent) and there is a case for "SRM: in transfer to execution system" that occurs in a small subset of traces after the Clear Invoice activity. What is important to focus in on is the time gap between these activities in such a small subset; in fact whenever the "Clear invoice" happened in 2018, then, at the beginning of 2019 the "SRM: in transfer…" occurs, probably because of a forced update of the system (if not for the event-log preparation), but hardly because of any planned process design.

The Item type is a Service one and this implies the presence of "Record Service entry sheet". It is interesting to point out that this activity enters into the process instance in parallel with Record Goods Receipt (same timestamp). The fact here is that "Record Service entry sheet" is loaded by any vendor and exists *in the system* since, associated with a "None" user (the one for vendors), but it is later loaded *into the process instance* concurrently with the activity of Record Goods receipt which is executed by any user_X. This shows a process design where an event pulls another event and can only be performed if the "pulled" one already exists.

Apart from the "Record Service entry sheet", which won't belong to it, traces within the category "3 way After, EC Purchase Order, Standard" show a similar process behavior as the above one.

7.2 Framework Order/Service and Standard P.O./Service

These traces differ from the E.C. Purchase Order/Service for the absence of "SRM…" activities and sub-process but show a similar pattern for their distinctive events (concurrency of Goods receipt and Service Entry sheet, users etc).

7.3 Standard P.O./Standard (Sub-contracting and Third Party)

Most of these traces show a process model as in Figure 11, in which activities "Record Goods receipt" and "Vendor creates invoice" are just parallel (not concurrent). As discussed about compliance, there is the case of "Remove Payment Block" occurring within the process, but this should not be interpreted as for the case of "3 way match, Before" traces. "Standard" traces also show the mentioned process change at the start of process instances (*3.5 Process drift and forced execution*).

8 “3 way match, Before”

8.1 E.C. Purchase Order/Standard

Traces with Document Type *EC Purchase Order* follow a process model with the above mentioned “SRM…” subprocess, which can occur at the start of the instance or within it in case of a change in price. In Figure 17 and Figure 18, Part 2 of the process is showed for traces where Goods are received before any Invoice and those where Invoice is received before any Goods.

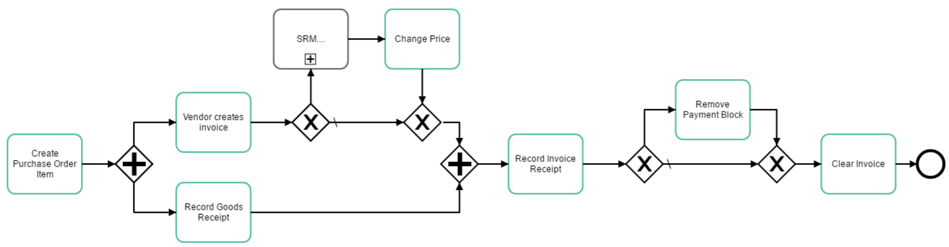


Figure 17 – “3 way Before, EC Purchase Order/Standard (Goods received before Invoice), Part 2

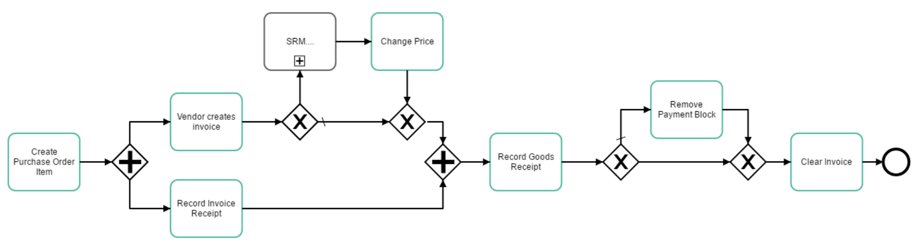


Figure 18 - “3 way Before, EC Purchase Order/Standard (Invoice received before Goods), Part 2

Part 1 of the process is actually the “SRM…” subprocess (model is simplified and doesn’t comprises cases with different Start events). Notes about “SRM…” subprocess (users involved etc) - already mentioned for “3 way After…” traces - apply to “3 way Before…” traces.

8.2 Standard P.O./Standard (Sub-contracting and Third Party)

These traces, the most common in the dataset, abide with the model already seen. The first part shows the change in process (see 3.5 Process drift and forced execution).

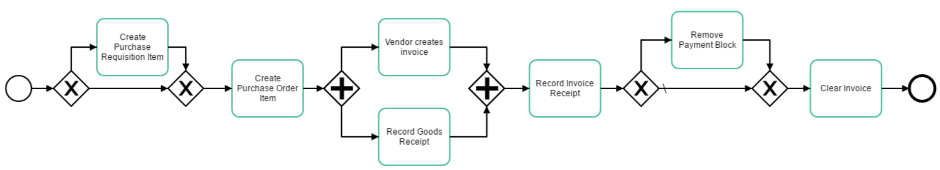


Figure 19 - “3 way Before, Standard P.O. (Goods received before Invoice)

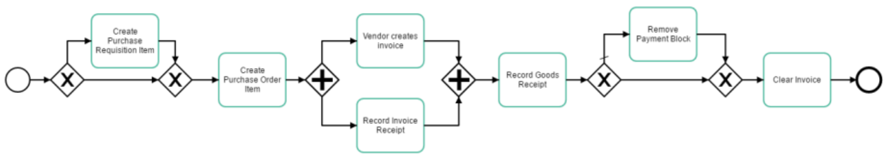


Figure 20 - "3 way Before, EC Purchase Order/Standard (Invoice received before Goods)

It is worth mentioning a peculiar timestamp evidence involving two events ("Record Invoice receipt" and "Vendor creates invoice"). There are cases in which "Record Invoice receipt" occurs just hours before "Vendor creates Invoice" (executed in batches at 23:59). This is not a violation of model-logic, it is instead another example of the fact that events already existing *in the system* are later pulled *into the process* (see last part of 7.1).

9 Conclusions

Process mining can support decision making and enrich those data-related set of performance KPIs any Company would collect and represent in dashboards. This contribute has focused on the added information a process perspective can give to management. A critical approach to data has led to questioning hypotheses and assumptions, discussing about process drift and consequences of process change (forced execution) as well as making assumptions on *four eyes principle* and *segregation of duties* in some instances. Both organization and training issues have been mentioned, while a first step toward process discovery and modeling has raised even more questions than those originally asked for.

Process mining is about information, not IT tools; for such a reason, even if many different tools have been used, none is explicitly mentioned in the text. Besides, many information was originally derived with the aid of data querying languages and data visualization tools and then specifically investigated with mining algorithms.

Procurement data in context: Analysis of the Procurement Process in the area of coatings and paints *

Business Process Intelligence Challenge 2019

Urszula Jessen[1]

akquinet cubit GmbH urszula.jessen@akquinet.de

Abstract. Understanding and managing the processes in a sustainable and optimized way is one of the significant challenges in modern organizations. Process mining offers both methodologies and tools to get unique insights into the process data. The process discovery, visualization, and conformance checking provide support not only for the management or process owners but also for all process users. On the other hand, most companies already possess different tools and paramount knowledge on how to analyze and understand data in their specific domain. In our approach, we try to find synergies between the classic data and process mining to let the users with tools and domain knowledge to deploy it in their daily tasks. In the following paper, we will demonstrate the framework for combining Process Mining Techniques with Data Science and BI Tools to achieve the best data insights for the given context and to utilize existing investment in knowledge and software in the organization. Based on the data from a large multinational company operating in the area of coatings and paintings, we develop a set of tools for exploratory analysis of the purchasing process.

Keywords: BPI Challenge· Process Mining · Event log · Process Intelligence · Descriptive Analytics.

1 Introduction

Business Process Intelligence Challenge is a yearly competition for process mining students, researchers, and professionals. The data provided in BPI Challenge 2019 originates from a large multinational company operating from the Netherlands in the area of coatings and paints and covers parts of its purchase order handling process. Each Process Order in Event log contains one or multiple line items.[1]. This data has been collected for the purchase order handling process for some of companies 60 subsidiaries. The process owner has some questions regarding the process and is generally concerned about compliance issues. The other problems that need to be addressed are generalized process models, process efficiency, process outliers and deviation from the standard process flow.

* Supported by Process Science.

The starting point of our analysis is understanding the processes and the context for the provided data. We will split the data into separate categories that can be described in a general manner as similar processes. We will present some of the "birds-eye" process maps with its most essential characteristics. After establishing the general framework for further research we will focus on three aspects of purchase processes: Efficiency, Effectiveness and Compliance. We will then design measurements and performance indicators for those aspects of the process. Among others, the time, costs and complexity of the process will be examined. We will separately address the questions of the process owner. In the end, we present our conclusions and further recommendation for this project.

Although we gave our best efforts to understand main issues and context of the provided data, we are neither the experts in the given domain nor did we have chance to conduct interviews with the process owner and process users. For the analysis, we made general assumptions, so that some of the outcomes may differ according to individual company policies or rules. The challenge encouraged participants to use different tools, techniques, and methods. To analyze the data, we used various commercial and open-source tools. Apart from process mining techniques, we utilized advanced BI Tools and methods. The process owner has some questions regarding the process and is particularly concerned about compliance issues. The other problems that need to be addressed are generalized process models, process efficiency, process outliers, and deviation from the standard process flow.

2 Understanding provided data

The data presented gives an overview of the transactional purchase process, which covers the activities from creating purchase order till invoice clearance.

Although, the data should contain only the purchase orders from 2018, individual values in timestamps of the event log stretch out to 1948 in the past and 2020 in the future. As we cannot check if such values exist due to conversion error or if there are some exceptional cases in the system, we have decided to filter the timestamps to values from 1/1/2017 till 25/5/2019. We did not filter out the whole traces, but only the timestamps that did not match in the proposed timeframe. The resulted event log contains more than 99% of the original events, and all the traces from original data are preserved. In further analysis, all statements are based on this filtered data.

The provided data consists in total of 251.734 cases. Most of them contain Vendor invoice message (209.889), and goods receipt message (234. 479). Only 73% of the invoices are cleared, and more than half of all process flows contains repetitive tasks. The other characteristics of the process are a relatively low proportion of the Catalogue based purchases; only one per cent of instances include SRM related tasks. The throughput of the process flow is, on average, 64 days.

2.1 Classification of data

In the provided data process owner has specified four different process flows, 3-way matching, invoice after goods receipt, 3-way matching, invoice before goods receipt, 2-way matching (no goods receipt needed) and consignment. In 3 way matching an invoice is usually matched to the corresponding purchase document for quantity and value. Table 1 gives an overview of case statistics for the different flow categories. For better clarity, we applied the abbreviations for each of the flow categories.

Table 1. Different process flow types.

Case Item Category	Abbreviation	# Occurences and (% Occurences)	Description
3-way matching, invoice after goods receipt	3WIAGR	15.148 (6%)	In this type of process flow, the company expects matching values for the goods receipt message, an invoice receipt message and the value that was defined in the first item creation .
3-way matching, invoice before goods receipt	3WIBGR	220.814 (88%)	A goods receipt message is required, but there is no need for GR-based invoicing.
2-way matching	2W	891 (0.4%)	For these items, the value of the invoice should match the value at creation, but there is no separate goods receipt message required.
Consignment	CONS	14.494 (6%)	For these items, there are no invoices on PO level as this is handled fully in a separate process.

Apart from that event log contains multiple attributes like document type, spend area text, source, or spend classification. The purchase documents can also be divided into product related and non-product related goods and services. The NPR, also known as indirect goods and services include all goods and services that are not directly involved in the production, such as capital equipment, marketing, legal assistance or telecommunication. The NPR related process variants should be considered separately as they usually have some different characteristics from typical product-related processes [2]. This kind of purchase is generally time-consuming as the items are typically non-standardized and purchased in small bulks. The NPR usually needs much more user attention and not only from the purchasing department but also all different parts of the company. Figure 1 shows the connection between spend classification, spend are and flow type. In the provided data almost 70% of all orders belongs to product-related category.

The other feature dividing the purchase document can case Document Type, where there are three types of processes, Framework order, EC purchase order, and Standard PO. Especially Framework orders which are processes that have some specific validation time (with a start- and end- date) and item limits, differ in their characteristics from standard PO. This kind of purchasing process is usually used for services, travel expenses or utilities.

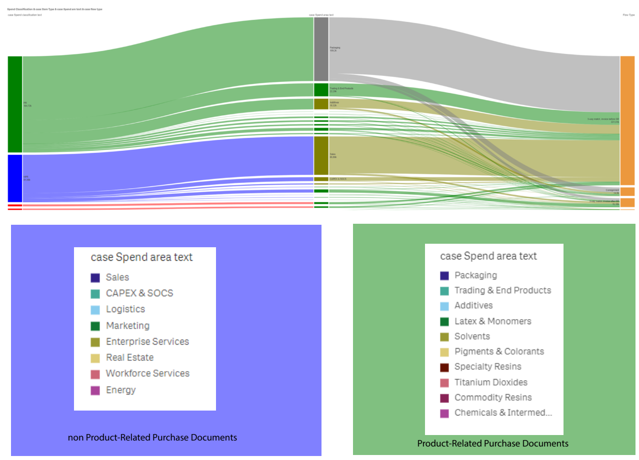


Fig. 1. Spend classification connection with spend areas and process flow types.

2.2 Data characteristics

Besides the typical data classification, it is essential to understand what kind of attributes should be considered as critical indicators for further inspection. In our evaluation, we want to focus on three aspects of the process: efficiency, effectiveness and compliance. Considering those aspects and available data, we concentrate on two distinct indicators: purchase volume and process throughput.

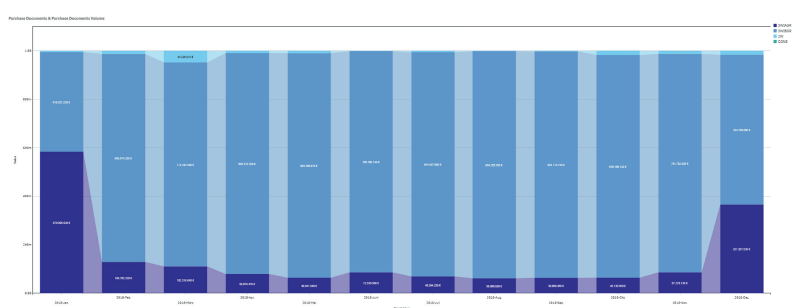


Fig. 2. Purchase volume for different type categories.

Figure 2 depicts the distribution of purchase volume throughout the year and for the different category types. For this analysis we used the average value of all documents in each process instance. Although the documents of the kind “3-way match, invoice before GR” have the most purchase documents line items and the most significant amount, there are two distinct value peaks for the document of the type “3-way match, invoice after GR” in January and December. It can be explained through service or line items, that are usually paid yearly (at the beginning or the end of each year). Similar amount peak can be observed for “2-way match” documents in March. Overall is the purchase value stable and constant over the year. Consignment documents are not shown as they have no amount value.

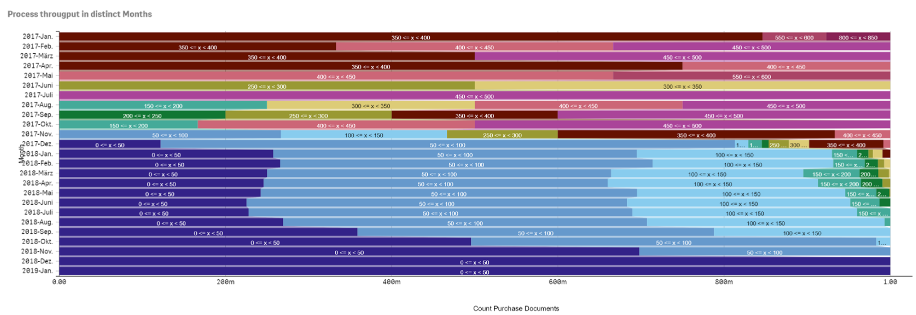


Fig. 3. Throughput of process instances(by year/month).

In the following analysis, we have divided throughput times into buckets of 50 days. Based on this classification, figure 3 depicts the throughput times of process instances from January 2017 till January 2019. There can be a distinct trend observed, in which process flow is getting steadily lower. Whereas most of the process instances in January 2017 have cycle times of 350-400 days, in January 2019 most of the duration of the case is not higher than 50 days. The most significant difference can be observed from January 2018.

3 Analyzing control flow

Process mapping is one of the most important parts of process mining projects. The literature describes multiple algorithms and tools for this task. In our analysis, we wanted to find the most generalized model of the given data and figure out how well does it fit with the whole data. After understanding the “big picture,” we would like to find more models that could better fit different categories of process flow.

From the provided event log, we extracted unique paths and calculated the frequency with which is each path present in the data. Figure 4 depicts the most general process flow of the provided data. The primary process flows start with the purchasing document, creating the invoice by vendor, goods receipt, invoice receipt message and in the end, clearing the invoice. To measure how well the depicted process generalizes the actual process flow, we have counted all process instances and calculated the percentage of the cases that are going through each

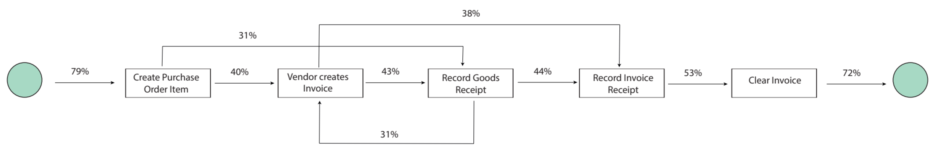


Fig. 4. Generalized process flow.

path. As shown in the figure, even though the first and last event can be good generalized, with over 70% of instances going through them, the other process activities are describing only the smaller part of the process flow.

The process flows that are generated from the provided data have an unproportionally high amount of process variants. In total, there are 11 824 different flows the process can follow. That means that on average, every 21 process instances have different process flow. Consequently, a "spaghetti-like" process model results from the full, unfiltered event log. This unstructured process visualization is incomprehensible and cannot be used in operational process analysis.

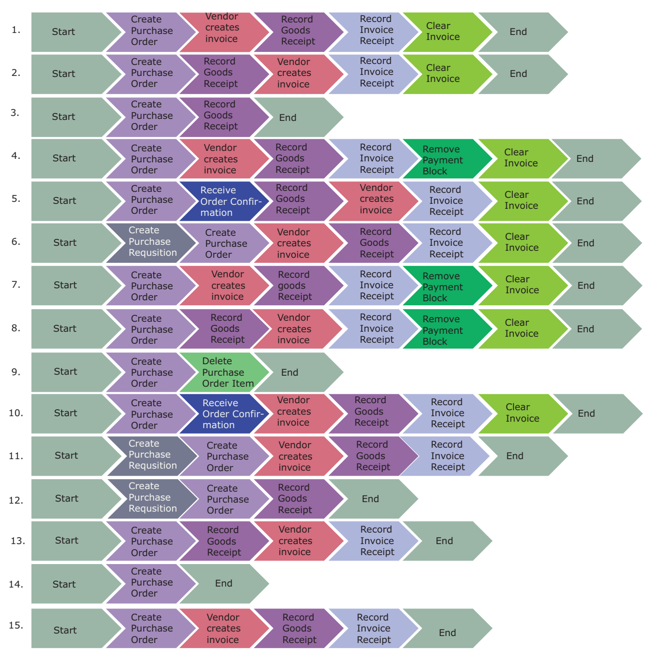


Fig. 5. Top 15 most frequent process variants.

To get a better understanding of the process, we have utilized the L* approach for Spaghetti processes [3]. Consequently, we have identified 15 most frequent paths and calculated some of the performance metrics. Those most frequent sequences are depicted in Fig. 5. The variants are grouped according to the count of instances that go through those flows. In the following chapters, we will use the numbers of paths for describing their properties. The generalized process flow shows that although the process map fits more than 66% of all process cases, it shows only about 50% of all events and less than 1 per cent of all process variants. Moreover, even if it shows the "birds-eye" view of the company, it does not help to answer questions under which conditions do the bottlenecks, or undesired paths take place.

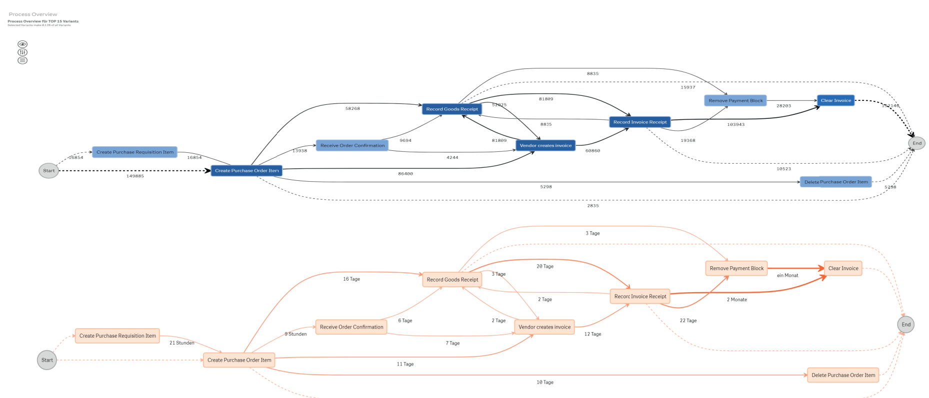


Fig. 6. Process Flow with frequency and duration for top 15 most frequent process variants.

Figure 6 depicts the process models extracted from the variants described in table 2. The paths between most of the tasks take from 1 day to maximal two weeks. The biggest bottlenecks are between invoice receipt and removing of the payment block and invoice clearing activities. We can also see that there are 158. 606 Goods Receipt Messages, 142.669 Vendor Invoices but only 132. 146 cleared invoices. The analysis of the processes for singular months shows that whereas the general process model fitted from 70 - 80% of data in January to August 2018, the same model in September - December 2018 was applicable for only about 50% of all cases. In those months, almost 50% of all processes started with creating a purchase requisition item instead of purchase item. Likewise, the average cycle time of the process drops to 42 days instead of the 67 days.

Table 2 shows the top 15 process variants and some of the performance indicators for them. The top 15 Variants can describe ca. 65% of all process instances. So although the primary process flow variant fits every fifth purchase document; still, all the lower variants explain only the small percentage of the given data. The most straightforward process variants with the smallest amount of activities are also the ones with the shortest throughput.

Table 2. Process flow and performance indicators.

Pattern	# and % of Occurences	Average (days)	St. dev(days)
1.	50.286 / 20%	70d 13h	33d 20h
2.	30.798 / 12%	84d 13h	33d 10h
3.	12.214 / 5%	22d 8h	21d 4h
4.	11.383 / 5%	91d 13h	41d 5h
5.	9.694 / 4%	93d 8h	20d 7h
6.	8.921 / 4%	60d 16h	23d 23h
7.	8.835 / 4%	77d 14h	39d 10h
8.	7.985 / 3%	102d 9h	34d 15h
9.	5.298 / 2%	9d 23h	25d 4h
10.	4.244 / 2%	76d 10h	23d 2h
11.	4.210 / 2%	38d 2h	19d 21h
12.	3.723 / 1%	26d 9h	20d
13.	3.548 / 1%	33d 8h	21d 23h
14.	2.835 / 1%	0d	0d
15.	2.765 / 1%	43d 3h	33d 12h

4 Challenge questions

Up until that point, we have described the general characteristics of the data and discovered the most widespread process model. In the next chapters, we want to answer the challenge questions and try to build the framework of performance indicators.

4.1 Q1: Collection of process models

Is there a collection of process models which together properly describe the process in this data. Based on the four categories above, at least 4 models are needed, but any collection of models that together explain the process well is appreciated. Preferably, the decision which model explains which purchase item best is based on properties of the item.

In the previous chapter, we have discovered and described model, which corresponds with ca. 65% of all cases. However, the generalized process model fits only the purchase documents that were created between January and August 2018. The standard deviations of throughput for different process variants were also relatively high, and the cycle times between them differed substantially.

In the next step, we compared the general process model created in the previous chapter with all four process flows described in the challenge. As shown in table 3, the 3WIAGR and 3WIBGR document flow generally fit with the primary process model, 2W and especially CONS (Consignment process) does not many similarities with our previous process overview. Consequently, as in the previous chapter, we have separated the most common variants for all four case item types to discover the best fitting process models.

Figure 7 depicts the process model for 3WIAGR documents. The main difference between the general process and this picture are the additional tasks: record service entry sheet and change price. The throughput of the purchase

Table 3. Most frequent process variants for main flow types.

% of Occurences					
Pattern	All cases	3WIAGR	3WIBGR	2W	CONS
1.	50.286 / 20%	22%	15%	0%	0%
2.	30.798 / 12%	13%	9%	0%	0%
3.	12.214 / 5%	1%	3%	61%	0%
4.	11.383 / 5%	5%	3%	0%	0%
5.	9.694 / 4%	4%	0%	0%	0%
6.	8.921 / 4%	4%	2%	0%	0%
7.	8.835 / 4%	4%	0%	0%	0%
8.	7.985 / 3%	3%	2%	0%	0%
9.	5.298 / 2%	2%	1%	2%	0%
10.	4.244 / 2%	2%	0%	0%	0%
11.	4.210 / 2%	2%	0%	0%	0%
12.	3.723 / 1%	1%	1%	12%	0%
13.	3.548 / 1%	2%	2%	0%	0%
14.	2.835 / 1%	1%	1%	2%	0%
15.	2.765 / 1%	1%	1%	0%	0%

orders of 3WIAGR type is 70 days, and each process has, on average, 21 activities. 42% of all SRM Orders are included in that category. The documents are created only by companyID_0000 and companyID_0001. About 10% of all invoices of that type are not cleared.

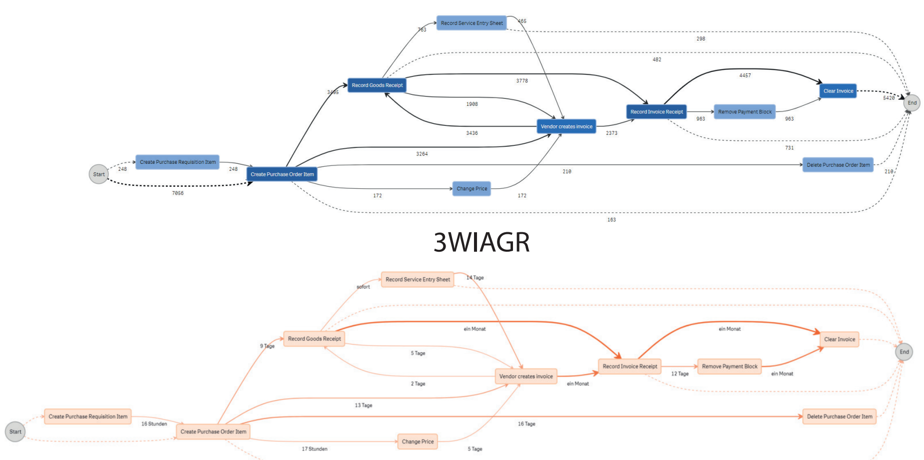


Fig. 7. 3-way match, invoice after GR.

Figure 8 depicts the process model with 15 most common variants for 3WIBGR documents. About 88% of all purchase orders belong to that category. T. The throughput of the purchase orders of 3WIAGR type is 73 days. In opposite to 3WIAGR documents the typical process in this category has only 5 activities.

58% of all SRM Orders are included in that category. The orders are created exclusively by companyID_0000 and companyID_000. About 10% of all invoices of that type are not cleared.

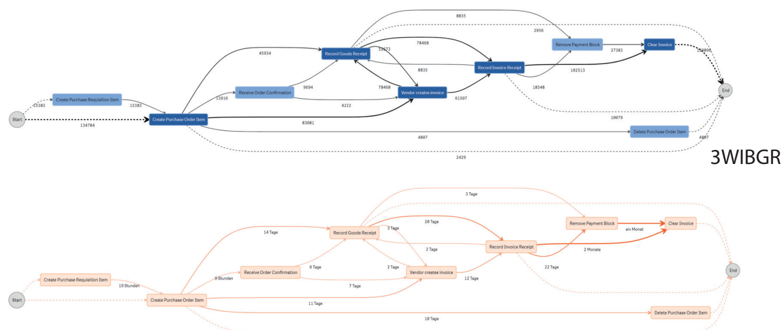


Fig. 8. 3-way match, invoice before GR.

About 6% of all cases involve consignment processes. With 24 days cycle time, it is the shortest process category. There is typically only 2 activities pro process. Also, there is no amount for the documents of that type. The typical case spends areas are packaging and additives. In comparison to other process types, there are proportionally more repetitive tasks and changes like "change delivery indicator".

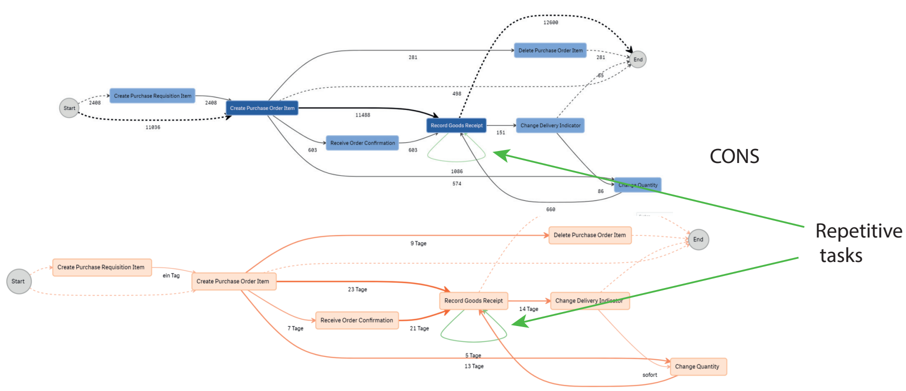


Fig. 9. Consignment.

2W is the smallest category in the provided event log. With 57 days of throughput are 2-way match process much shorter than those belonging to

3WIAGR and 3WIBGR category. The typical process has 5 activities. There are no goods receipt messages in that process type and no SRM tasks. The biggest group of case spend area are Real estate and others purchases. There are no product-related documents in this process flow.

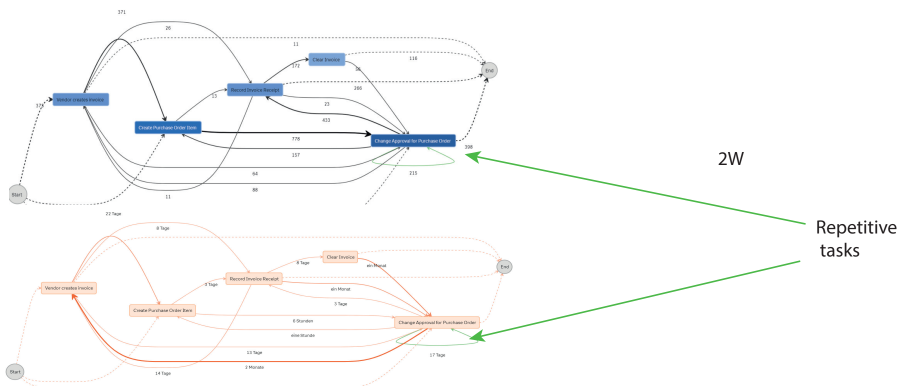


Fig. 10. 2 way match.

With the general model, we could answer the question of what happens in the process but not why and under what circumstances does it happen.

4.2 Q2: Invoicing process throughput

What is the throughput of the invoicing process, i.e. the time between goods receipt, invoice receipt and payment (clear invoice)? To answer this, a technique is sought to match these events within a line item, i.e. if there are multiple goods receipt messages and multiple invoices within a line item, how are they related and which belong together?.

In our general analysis, we used the compliance flag in to compare the first amount of purchase document (value at the creation) with the sum of values connected to goods receipt items and the same for the invoice receipt messages. That value matches for almost 85% of all cases and 61 of the total purchase volume. Similarly, for the second question, we have loaded only the events from the invoicing process. From the filtered cases, only about 20% (51.497) of purchasing documents do not have matching values. Figure 11 shows the attributes for non-matching and matching values. Figure 11 depicts the typical attributes of the purchasing documents. Both categories have mostly 3WIBGR type documents. The most significant difference is by not matching documents the proportion of product-related purchases, which is much bigger than by matching documents.

Figure 12 shows the comparison of performance metrics for both matching- and not matching the category of purchase documents. The documents that do

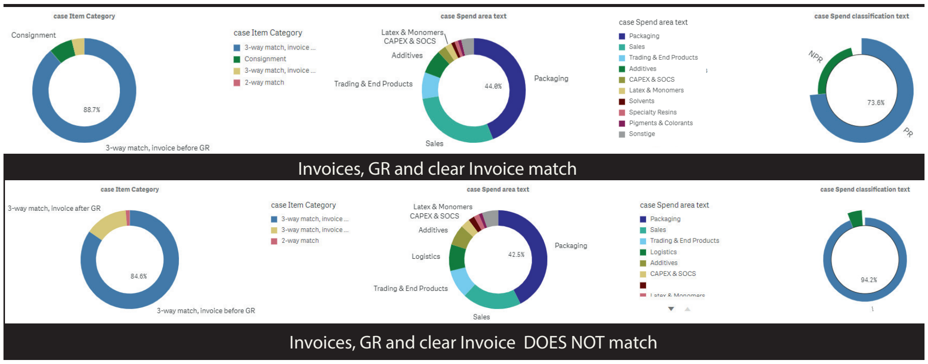


Fig. 11. Comparison of matching and not matching Purchasing Documents.

not match have usually more activities pro process (more repetitive tasks) but at the same time, the cycle time is much shorter. The amount for each purchase documents is for that category bigger than for those with matching values.

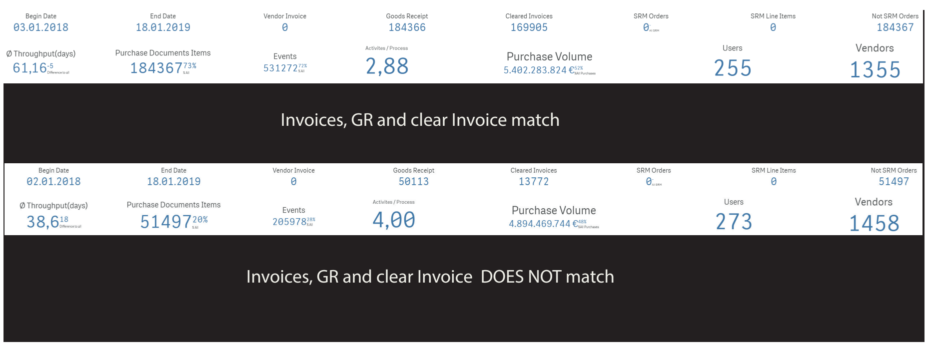


Fig. 12. Comparison of matching and not matching Purchasing Documents.

Figure 13 shows the top 10 vendors for which the values do not match. All the documents for those vendors belong to the non-product related category. Most of the purchases are involving logistics or marketing. Although there are only 843 purchase document items, they add up to 8% of the total purchase volume. All of those documents belong to the 3WIAGR category.

Figure 14 compares process models for both matching and not matching categories. The process models show much more repetitions for not matching documents. The repetitive tasks are on average longer than in those matching cases. To compare the two categories, we have classified all documents in cat-

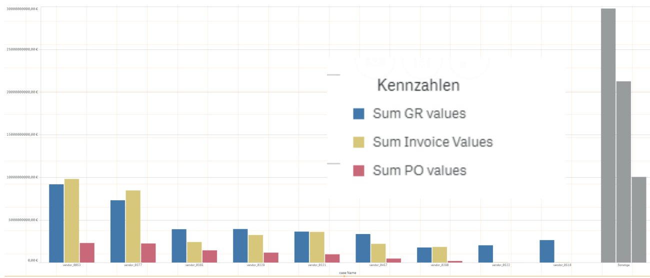


Fig. 13. Top 10 Vendors for not matching documents.

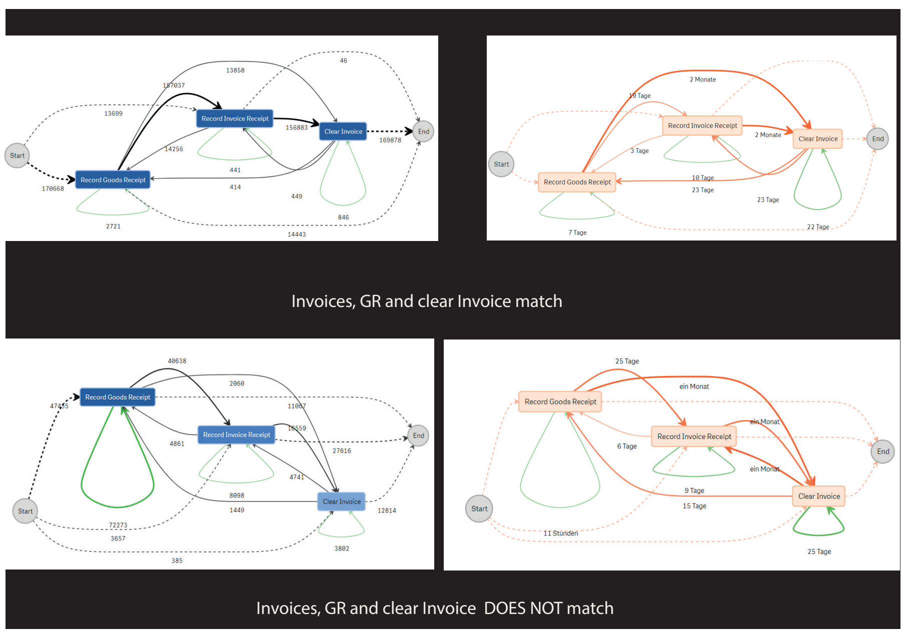


Fig. 14. Comparison of process models for matching and not matching Purchasing Documents.

egories comparing their throughput time. Figure 15 shows the comparison of those two types of documents based on throughput clusters. The most frequent cases have cycle times from 0 to 50 days (dark blue) . Especially at the beginning of 2018, there is a clear difference in throughput times between those documents.

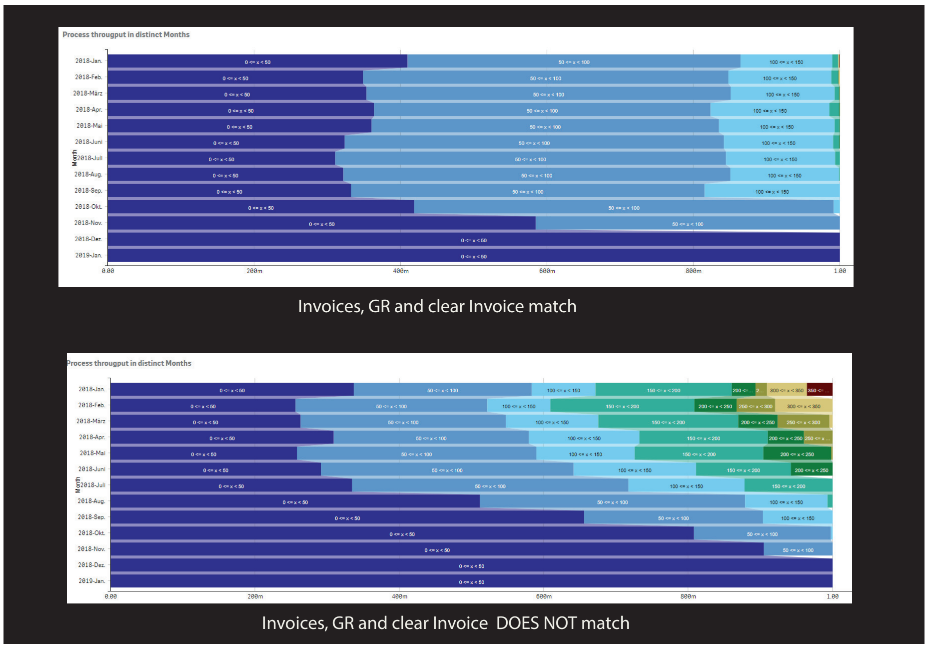


Fig. 15. Comparison of cycle times for matching and not matching Purchasing Documents.

Similarly to throughput times, we have divided all documents according to the number of events for each process. Figure 16 shows a comparison of those groups. The cases in which there are matching values have at most 5 activities pro process, and those with outstanding amounts have more activities and repetitive tasks.

Figure 17 shows the spend areas for which values in the invoice, goods receipt and payment does not match. Especially in the logistics process, the amount of goods receipt is much higher than invoice and payment values. Other typical categories for which the singular documents do not match are marketing, workforce services and packaging.

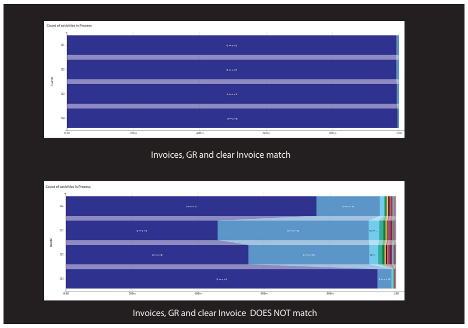


Fig. 16. Case spend areas for not matching Purchasing Documents.

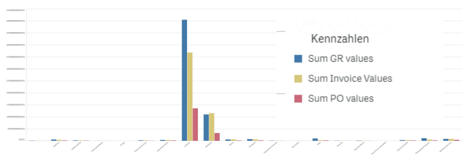


Fig. 17. Comparision of matching and not matching Purchasing Documents.

4.3 Q3: Outliers and deviations

Finally, which Purchase Documents stand out from the log? Where are deviations from the processes discovered in (1) and how severe are these deviations? Deviations may be according to the prescribed high-level process flow, but also with respect to the values of the invoices. Which customers produce a lot of rework as invoices are wrong, etc.?

As a part of the data load, we have created different flags for the cases that are, in some way deviating from the normal process flow or could cause in some way compliance or rules violation. We have found that about 34% of all cases have different deviations or suspect behaviour. The most frequent process deviation is rapid goods delivery. There are about 6 thousand cases where the receipt of the goods follows other tasks only after a few minutes. This kind of behaviour is typical for vendors 0458, 0522,0228 and 0525. The affected spend areas are logistics, capex & socs and sales.

The other suspected behaviour are the cases where purchase order item is created three or more weeks after the vendor invoice. This late Purchase Order creation is typical for about 3 thousand documents. This behaviour is found mainly at Trading & End products. In about 2400 cases the price of the purchase order is changed after an invoice has been issued. It is usually the case for packaging, sales and additives purchases.

5 Conclusion

As a part of BPI Challenge 2019, we have analyzed the purchasing process for a large multinational company. The decentralized process was not only complicated but did involve many different subprocesses. On the one hand, the big part of purchasing involved non-product related purchases. Especially the processes involving marketing or logistics contained many deviations and untypical behaviour.

References

1. van Dongen, B.F., Dataset BPI Challenge 2019. 4TU.Centre for Research Data. https://doi.org/10.4121/uuid:d06aff4b-79f0-45e6-8ec8-e19730c248f1
2. Telgen, J. and de Boer, L. 1995. Developments in purchasing of non-production items. Proceedings 4th IPSERA Conference, Eindhoven, 1-8.
3. W. M. P. van der Aalst, "Process mining: discovering and improving Spaghetti and Lasagna processes," 2011 IEEE Symposium on Computational Intelligence and Data Mining (CIDM), Paris, 2011, pp. 1-7. https://doi.org/: 10.1109/CIDM.2011.6129461

BPI Challenge 2019 Report: a Purchase-to-Pay Process Analysis

Adriano Augusto, Volodymyr Leno, and Daniel Reissner

University of Melbourne
{aaugusto, vleno, dreissner}@student.unimelb.edu.au

Abstract. Every year, since 2011, the Business Process Intelligence Challenge (BPI Challenge) allow students and practitioners to test their process mining skills: analysing process data sourced from real operating companies, addressing specific stakeholders questions, and trying to derive valuable insights. This year (2019), the process data captures the purchase-to-pay process of a company operating in the Netherlands in the area of coatings and paints. This paper reports the results of our process mining analysis, which revolved around the three main questions set by the stakeholders. Precisely, we (i) discovered a set of as-is process models, and a set of to-be process models representing the purchase-to-pay process; (ii) developed a Java application to run a throughput analysis of the process execution; (iii) identified the process deviant behavior and highlighted the most interesting and valuable insights.

1 Introduction

Business processes shape the way organizations operate in order to provide services and products to their customers. Given their importance, business processes are most of the times supported by information systems, which record data about individual executions of the processes. This data is stored in the form of event logs, where each event represents the execution of an activity in the context of a case. However, this raw process data would be useless if not analysed. In the last two decades, numerous research studies considered the problem of process data analysis, proposing a range of techniques that help organizations to gain useful information and knowledge about their business processes, with the ultimate goal of assessing and (where possible) improving them. These techniques belong to the discipline of *Process Mining* and they can be categorised into three macro groups: (i) automated process discovery (generating a process model from an event log); (ii) conformance checking (identifying and diagnosing mismatches between reference behavior and recorded behavior); and (iii) process enhancement (enriching a process model using the information recorded on the event log) [3].

Given the rising interest in Process Mining both in research and industry, in 2011, the Business Process Intelligence Challenge (BPI Challenge) was established, with the goal of giving an opportunity to students, academics, and practitioners to test their process mining knowledge, techniques, and skills: analysing process data sourced from real operating companies, addressing specific stakeholders questions, and trying to derive valuable insights.

This year, for the BPI Challenge 2019, we analysed the process data recorded during the execution of the Purchase-to-Pay (P2P) process of a company operating in the Netherlands in the area of coatings and paints[4]. This report summarises the results of our analysis, which was driven by three main questions proposed by the company stakeholders. Precisely, the company is interested in:

1. obtaining a collection of process models that clearly explain their P2P process.
2. analysing the throughput of the P2P process, with focus on its time performance.
3. identifying compliance issues, comparing the process behavior recorded in the process execution data with the expected behavior captured in the process models and its high-level textual description.

The structure of this report follows the three main questions of the company. In Section 2, we focus on process identification and discovery, describing the steps we applied (and tools we used) to analyse and understand the P2P process execution in order to produce a set of process models capturing the recorded process behavior (as-is processes) and a set of process models representing the expected process behavior (to-be processes). In Section 3, we focus on the throughput of the P2P process, with a great focus on its time performance, we describe the application we implemented for analysing this latter, and we report a wide range of measurements. In Section 4, we discuss our conformance checking analysis, focusing on the compliance issues we identified and highlighting the most interesting mismatches between desired process behavior and recorded process behavior. Finally, the conclusion summarises the output of our process (mining) analysis.

2 Process Discovery

In this section, we first provide a textual description of the P2P process, as a reference to understand in depth this report and our analysis. [1] Then, we analyse the process behavior recorded in the event log. Finally, we show how we generated the *as-is process models*, combining automated process discovery with ad-hoc refinements driven by our process map analysis. We conclude the section proposing a set of *to-be process models*, which capture in a more precise way the process description and leave out infrequent (and senseless) behavior that we identified in the event log during this first step of our analysis.

2.1 Process Description

The P2P process handles purchase orders (POs), from their creation to their clearance. Each PO is recorded into a *PO document* containing several lines. Each line of a PO document refers to a *PO item*, specifying its details (e.g.

[1] The process description we provide can be found also in the BPI Challenge Manifest, available at: https://icpmconference.org/icpm-2019/contests-challenges/bpi-challenge-2019/

item type, category, etc.) and its value (i.e. its cost). Each PO item is processed separately and according to a specific data flows. The data flows are characterized by three main steps and their execution order, these are: (i) *Record Goods Receipt*, (ii) *Record Invoice Receipt*, (iii) *Clear Invoice*. Here, we summarise the four possible data flows.

DF1. *3-way matching, invoice after goods receipt.* In this type of data flow, *Record Goods Receipt* must occur before *Record Invoice Receipt*, and this latter before *Clear Invoice*. The value of the Goods Receipt message must match the value of the Invoice Receipt message and the value at the creation of the PO item.

DF2. *3-way matching, invoice before or after goods receipt.* In this type of data flow, *Record Invoice Receipt* can occur either before or after *Record Goods Receipt*. However, if it occurs before *Record Goods Receipt*, the data flow is blocked until the Goods Receipt message is received. Also in this case, as the previous, *Clear Invoice* can occur only if the value of the Goods Receipt message matches the value of the Invoice Receipt message and the value at the creation of the PO item.

DF3. *2-way matching, no goods receipt needed.* In this type of data flow, no Goods Receipts message is received. Instead, the Invoice Receipt message is received and its value must match the value at the creation of the PO item. However, the entire value of the PO item could be consumed partially, such that multiple Invoice Receipt messages are received, each of them covering a sub-value of the total PO item value (until the total value is covered).

DF4. *Consignment.* This data flow is the simplest one, it requires only to *Record Goods Receipt*, whilst it does not need to *Record Invoice Receipt*, nor to *Clear Invoice*.

The data flows give an overview of what are the main steps of a PO item processing and their execution order, however, a PO item could follow a certain data flow multiple times. For example, a PO item processed with the data flow DF1 could have multiple Goods Receipt messages, each of them followed by an Invoice Receipt message and its clearance (i.e. *Clear Invoice* step).

In the following, we refer to this process description to interpret the process execution data recorded in the event log, as well as to provide the final to-be process models that capture both the behavior in the event log and the behavior description.

2.2 Process Analysis

We started our process analysis by visualising the data recorded in the original event log [4]. To do so, we uploaded the event log into Disco[2], which allows us to visualise it as a process map, as well as to apply filters on events, activities, paths, timeframe etc. Figure 1 shows the unfiltered process map. We can easily understand that the data recorded in the event log is noisy, and filters are necessary to interpret it, this is common when dealing with real-life data.

[2] A commercial process mining platform, available at: https://fluxicon.com/disco/

Given that each case in the event log records the data flow of a PO item within a PO document, and that each data flow belongs to one of the four types described above (i.e. DF1-DF4), we decided to split the original log into four sublogs, each capturing one of the four data flows. Then, we took into account the timeframe of the events recorded for each data flow, and we observed that several events where recorded in doubtful dates and times (see Table 1). Consequently, we applied a filter on the timeframe, retaining only the data flows recorded between the *31 Dec 2017* and the *18 Jan 2019*. We believe that the events falling outside this timeframe are due to either recordings errors or data corruption, the company may be interested in revising the robustness of its process recording software and/or the security of its databases.

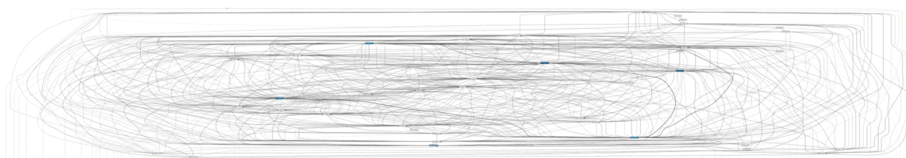


Fig. 1. Process map of the original log, unfiltered.

Data Flow	Recordings Timeframe	
	Start	**End**
DF1	*24 Jan 2001*	*10 Apr 2020*
DF2	*27 Jan 1948*	*06 Dec 2019*
DF3	*26 Jan 2017*	*02 Feb 2019*
DF4	*31 Dec 2017*	*18 Jan 2019*

Table 1. Data flows timeframes observed in the event log.

At this stage, we obtained four different event logs (one per data flow type) having a restrained timeframe. We proceed with analysing each of them separately.

Data Flow 1. The DF1 recordings show that up to 38 unique activities can be executed during the processing of the PO items. However, having a closer look at the activities, we notice that some of them are likely to be *external*. We consider external those activities that do not relate to the PO item processing because either belonging to another context or executed by external entities.
Assumption-1. All the *SRM*-tagged activities (e.g. *SRM: Created*, *SRM: Ordered*) are external, since they clearly relate to the *Supplier Relationship Management* (SRM) process.

It is not easy to understand why the SRM activities are recorded in the event log of the P2P process, and unfortunately, there is no mention of the SRM process in the details provided by the company. For the remaining, we resort to Assumption-1 when analysing the SRM activities. In fact, it is likely that they follow a separate workflow, being performed most of the times by *Batch Users*, and often in between different (random) activities not relating to the SRM process.

Assumption-2. The activities performed by the *Vendor*[3] are external, which are the following two: *Vendor Creates Debit Memo* and *Vendor Creates Invoice*. Furthermore, these activities are always recorded by the user *NONE*, in the event log.

Following assumption-2, we analysed the user *NONE*. This latter performs only four activities in the event log: (i) *Vendor Creates Debit Memo*; (ii) *Vendor Creates Invoice*; (iii) *Record Service Entry Sheet*; (iv) *Clear Invoice*. This finding led us to formulate the next assumption.

Assumption-3. Tthe user *NONE* represents the *Vendor* within our event log.

Assumption-3 associates the activity *Record Service Entry Sheet* to the Vendor as well, since this activity is always and only performed by the user *NONE*. More complicated is the case of the activity *Clear Invoice*, being this latter performed 4.1% of the times by *NONE*, but the remaining by company users. This fact was remarkable, indeed, according to the process description the *Clear Invoice* activity is one of the most important of the P2P process, how come such activity could be recorded by a user without leaving traces?[4] Consequently, we did not associate *Clear Invoice* to the Vendor, but we considered outliers those instances of *Clear Invoice* which are recorded in the event log by the user *NONE*.

After identifying the external activities, i.e. SRM and Vendor activities, we filtered them out from the event log, being left with 23 activities. Among these 23 activities, we identified 5 which are rework activities (see Table 2), yet the total number of activities is very high to have an overview of the DF1 processing procedure. At last, we decided to filter the activities with frequency less than 1%. This last filter highly reduced the total number of activities to five: *Create Purchase Order Item*; *Record Goods Receipt*; *Record Invoice Receipt*; *Clear Invoice*; *Remove Payment Block*. Except for the latter, we can immediately identify the DF1 provided in the textual description. Therefore, no more filters were applied, Table 3 summarises the number of activities identified in each data flow.

Rework Activities
Change Quantity
Change Price
Change Delivery Indicator
Change Storage Location
Change Currency
Change Payment Term

Table 2. Rework activities.

Later in this section, we show how we discovered the as-is process model of the DF1 starting from this highly filtered log.

Data Flow 2. The recordings for the DF2 are very similar to the ones of DF1. The DF2 counts up to 39 distinct activities, 13 of which are external (11 SRM and 2 Vendor activities). Of the remaining 26 activities, 6 are rework activities, and 9 occur with a frequency rate over 1%. The latter ones are the

[3] According to their activity name.
[4] A *NONE* user cannot be tracked.

Data Flow	**Activities**					
	Unique	**External**	**SRM**	**Vendor**	**Rework**	**Freq. > 1.0%**
DF1	38	15	12	3	5	5
DF2	39	13	11	2	6	9
DF3	11	2	0	2	1	4
DF4	15	0	0	0	6	7

Table 3. Actvities per data flow.

following: *Record Goods Receipt*; *Create Purchase Order Item*; *Record Invoice Receipt*; *Clear Invoice*; *Remove Payment Block*; *Create Purchase Requisition Item*; *Receive Order Confirmation*; *Change Quantity*; *Change Price*. Compared to the most frequent activities identified for the DF1, in DF2 we have almost double, some of which are not mentioned in the process description, e.g. *Create Purchase Requisition Item*.

Data Flow 3. The DF3 is the simplest among the four data flows. Its total number of distinct activities is only 11. There is no trace of SRM activities, but we find again the 2 Vendor activities. Removing the latter ones and focusing on the activities with frequency rate over 1%, we are left with the following: *Change Approval for Purchase Order*; *Create Purchase Order Item*; *Record Invoice Receipt*; *Clear Invoice*. Compared to the most frequent activities identified for the DF1 and DF2, in DF3 it is already very clear the full process behavior, which matches (almost) straightforward with the process description for the DF3.

Data Flow 4. The DF4 recordings contain up to 15 activities, 6 of which are rework, and none of them external (this was expectable, given that no invoice clearance is necessary in this data flow). The most frequent activities (freq. > 1%) are: *Record Goods Receipt*; *Create Purchase Order Item*; *Create Purchase Requisition Item*; *Change Quantity*; *Receive Order Confirmation*; *Change Delivery Indicator*; *Delete Purchase Order Item*.

Having identified external activities, rework activities, and most frequent activities, we further filtered the four event logs. First, we removed all the external activities: (i) we discarded the activities relating to the Vendor tasks, and (ii) we generated a fifth event log comprising only and all the SRM activities (we will use this latter to discover the SRM process, keeping it disjointed from the P2P process). Then, we filtered out from each of the four event logs the infrequent activities (absolute frequency less than 1%), and the rework activities (we postpone their analysis). These five event logs are the input for the next phase of our analysis, which is the discovery of the as-is process models. Table 4 briefly summarises the five (filtered) event logs key statistics.

Filtered Event Logs	Cases Total	Cases Unique	Events	Duration (days) AVG	Duration (days) MAX	Duration (days) MIN
DF1	15,129	881	122,427	66.6	379.6	<1
DF2	220,810	1,539	960,838	70.8	365.0	<1
DF3	1,027	104	5,038	46.5	348.3	<1
DF4	14,498	37	33,923	23.4	209.7	<1
SRM	1,426	58	11,415	23.7	379.6	<1

Table 4. Filtered event logs statistics.

2.3 As-Is Process Models

To discover the as-is process models, we uploaded the five event logs to Apromore,[5] which allows us to visualise the event logs as process maps, to easily turn them into BPMN models applying Split Miner [2], and (if needed) to manually edit the discovered models for enhancing them.

Figures 2 to 6 show the BPMN models discovered on Apromore (using the Process Map Discovery Plugin and its embedded Split Miner [6]) from the four (filtered) event logs capturing the four data flows and the SRM event log containing only the SRM activities. We note that these process models highly match the four data flow descriptions, however, we need to mention a few remarks.

Remark-1. In the DF1 process model, the *Remove Payment Block* activity is unexpected, given that for the DF1 the *Record Invoice Receipt* activity is always following the *Record Goods Receipt* activity, the payment block should never be required. Also, if a payment block is applied, it is done in a transparent way, since the payment block removal is frequently observed in the event log, but not its setting (i.e. the activity *Set Payment Block* is rare).

Remark-2. In the DF2 process model, the *Record Invoice Receipt* activity is rarely observed before the *Record Goods Receipt* activity, despite it is allowed according to the process description. Often, an order confirmation is required for the PO items processed with the DF2, though we were not able to determine when or why. Finally, another highly frequent activity not mentioned in the DF2 description is: *Create Purchase Requisition Item*. The name of this latter, unfortunately, does not give us any hint in determining the purpose of the activity, but we believe it is worth mention its frequency (42564 occurrences, 4.17%), the company stakeholders may find it interesting or unusual.

Remark-3. The DF3 process model is the simplest out of the four, as well as the rarest type of data flow, with only 1027 cases (see also Table 4). The model clearly captures its description except for the activity *Change Approval for Purchase Order*, which occurs at least once in all the cases. This activity seems to be necessary (or even mandatory) for this type of data flow.

Remark-4. The DF4 process model presents some characteristics similar to the DF2 one, for example, the execution of the activities *Create Purchase Requisition Item* and *Receive Order Confirmation*. However, also in this case, we do not

[5] Apromore is the web-based process analytics platform maintained by the BPM research group of the University of Melbourne, more info at: apromore.org. Apromore is free and publicly available at: apromore.cis.unimelb.edu.au

[6] With parameters: 100(*activities*), 20(*arcs*), and 40(*parallelism*)

have enough information to explain their recurrent appearance. It is interesting to note that the DF4 is the only type of data flow where the deletion of the PO item occurs frequently (422 times out of 14,498).

Remark-5. Regarding the SRM process model, we have no information nor reference behavior to claim or reject its correctness. However, its activities names are self-explicative and can be easily understood. Accordingly, we believe the behavior captured in the SRM process model is either correct or very close to the correct one, being its control flow reasonable and having applied no filters on the SRM activities.

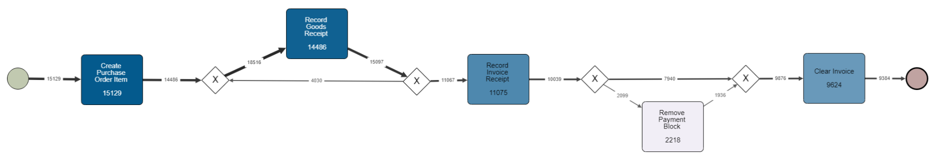


Fig. 2. As-is BPMN process model of the DF1, filtered infrequent behavior.

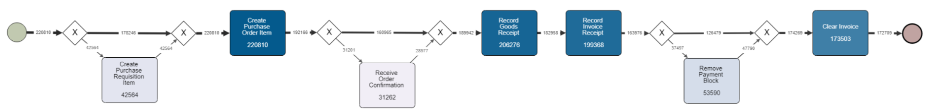


Fig. 3. As-is BPMN process model of the DF2, filtered infrequent behavior.

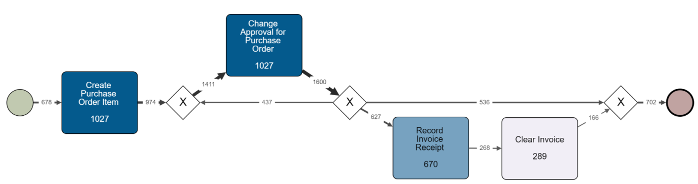


Fig. 4. As-is BPMN process model of the DF3, filtered infrequent behavior.

The filtered as-is process models are a good representation of the most frequent behavior observed in the four data flows. Exception made for the above remarks, the company should be relieved knowing that the majority of the times the workflow adheres to the prescribed one. Nevertheless, as we highlighted in the process analysis, several activities are observed in the original event log that generate noise (infrequent or deviant behavior). Given that the noise recorded in the event log may conceal interesting insights, which the company may find valuable, we tried to incorporate that information into the models showed in

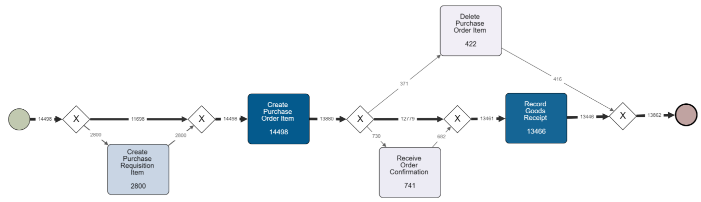


Fig. 5. As-is BPMN process model of the DF4, filtered infrequent behavior.

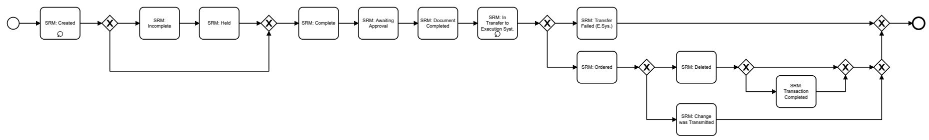


Fig. 6. The SRM process model.

Figures 2 to 5. To do so, we did the following. For each of the four data flows, we considered all the infrequent activities that were filtered out (i.e. activities with frequency $< 1\%$). For each infrequent activity, we analysed only the event log cases containing it, we detected its most frequent preceding activity and its most frequent succeeding activity, and we placed the infrequent activity between the former and the latter.

The process models we obtained after this procedure were then enhanced using the complete BPMN elements (to reduce complexity and increase understandability), as an example we reported in Figure 7 the DF2 complete and enhanced model (we suggest view on screen, for zooming in). It is easy to note that when trying to capture all the infrequent behavior recorded in the event log, the complexity of the process model explodes. This should not discourage the company stakeholders, but rather become the starting point for a deeper analysis of the P2P process behavior. Indeed, reaching this stage in our analysis allowed us to interpret further the process behavior and generate the to-be process models.

2.4 To-Be Process Models

Starting from the as-is process models (the complete and unfiltered versions), we decided to *redesign* them in what we think could/should be the actual process models. [7] The redesign did not follow a structured approach, instead, we integrated our knowledge and understanding of the process behavior into the models.

Figure 8 shows the root process, we assumed that a PO document is processed as a whole, whilst its line items are processed by subprocesses, each referring to a specific data flow (DF1-DF4). Furthermore, we assumed that any rework

[7] All the process models discovered in our analysis are available at: https://www.dropbox.com/s/j83l5m5y9gkiluo/bpic2019models.zip?dl=0

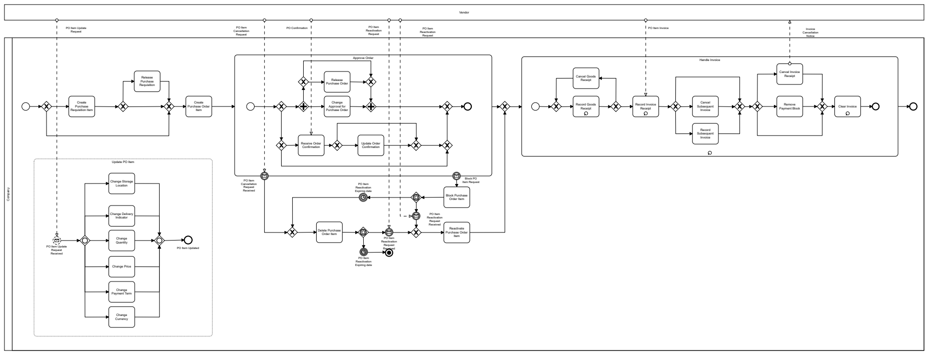


Fig. 7. As-is BPMN process model of the DF2, unfiltered and enhanced with complete BPMN elements.

activity can be required at any time and it must be triggered by an external message/notification, this is captured by the event subprocess in the root process.

The subprocesses modeling the behavior of the four data flows are shown in Figures 9 to 12. The behavior represented matches the reference behavior, but at the same time allows for the extra (infrequent) behavior identified in the original event log. Describing in details the behavior of the process models is out of the scope of this report, since the BPMN representation is straightforward and self-explicative. We recommend to visualise the models on screen, or to download the original PDF files and print them full scale on paper.

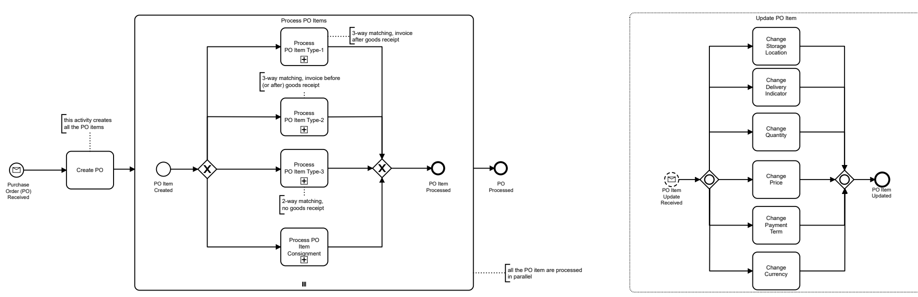


Fig. 8. Root process model.

3 Process Performance

In this section, we compute and analyse the throughput time and the throughput of the P2P process. First, we calculate the throughput time of the process, which we defined as the time between the main steps of the data flows (i.e. *Record Goods Receipt*, *Record Invoice Receipt* and *Clear Invoice*), and we propose a control-flow based technique to match these steps when recursively executed during a

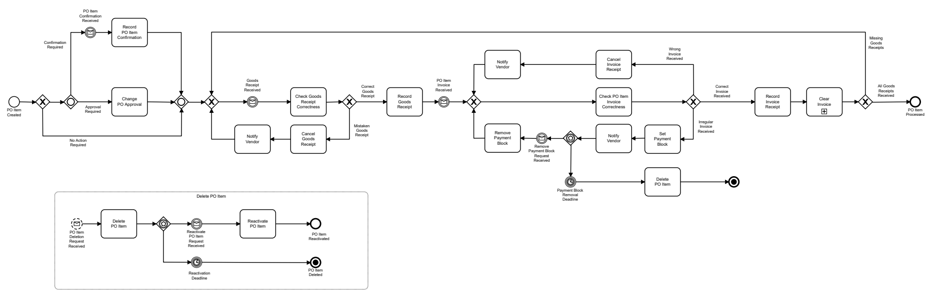


Fig. 9. To-be BPMN process model of the DF1.

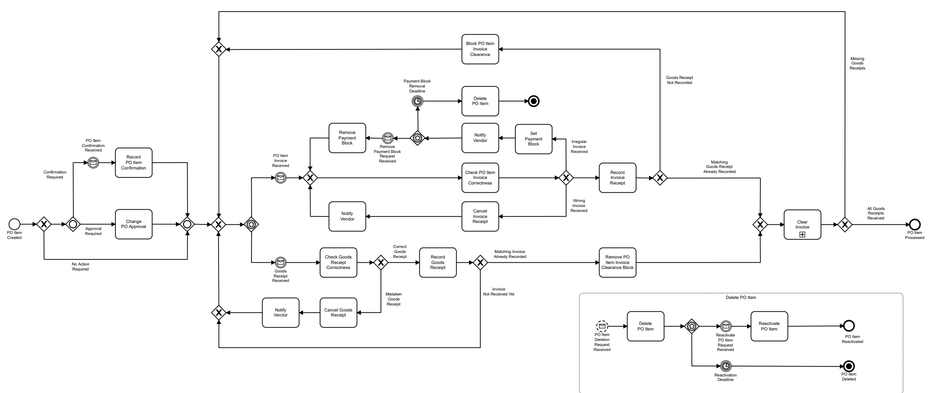


Fig. 10. To-be BPMN process model of the DF2.

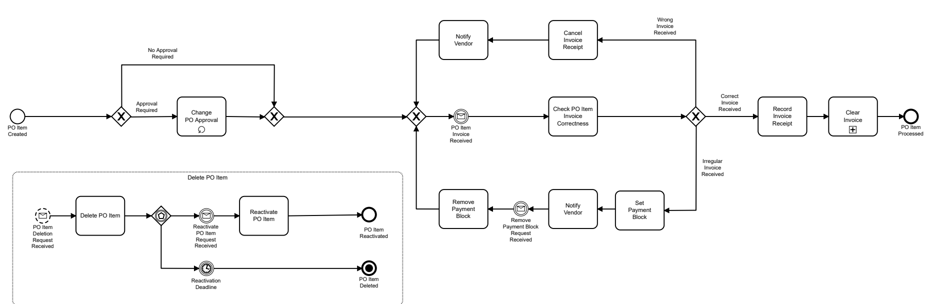


Fig. 11. To-be BPMN process model of the DF3.

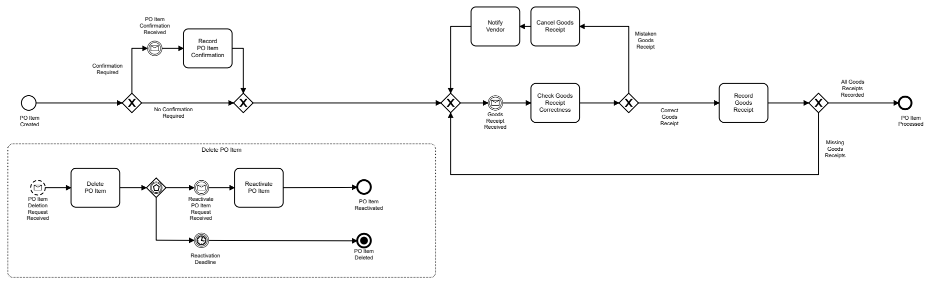


Fig. 12. To-be BPMN process model of the DF4.

PO item processing. Then, we calculate the throughput of the process in terms of net worth flow and payments rate (average number of payments per day).

3.1 Throughput time

In the context of our P2P process analysis, we define the *Throughput Time* as the amount of time required to handle single invoice, with reference to the times between the three main steps: *Record Goods Receipt*, *Record Invoice Receipt* and *Clear Invoice*.

We performed our throughput analysis using the four event logs described in the previous section (see Table 4), one per data flow, except for the one capturing case DF4. We held out this latter because it does not contains any invoicing information. The main challenge of computing the throughput times between the three main steps of the data flows, is the presence of recurring events representing multiple times the same steps for a given PO item. For example, given a PO item, we can observe in the event log multiple *Record Goods Receipt* activities, followed by multiple *Record Invoice Receipt* activities, followed by multiple *Clear Invoice* activities. Therefore, we asked the following question: can we match each of the *Record Goods Receipt* activity to the corresponding *Record Invoice Receipt*, and each of the latter ones to the matching *Clear Invoice* activity? Unfortunately, the information recorded in the attributes of the events recorded in the logs is not sufficient to efficiently cluster the three main steps into triplets. Therefore, we designed an ad-hoc matching technique to compute the throughput time.[8] Our technique is based on the First In First Out (FIFO) principle, meaning that the earliest invoice will be closed first, e.g. the earliest *Record Goods Receipt* activity will be matched with the earliest *Record Invoice Receipt* activity, which will be then matched with the earliest *Clear Invoice* activity.

Given a data flow event log, e.g. DF1, first, we create three vectors: idx_1, idx_2, and idx_3, representing (respectively) the observations of *Record Goods Receipt*, *Record Invoice Receipt* and *Clear Invoice*. Then, we scan all the cases in the event log searching for the aforementioned events and we save their positions in the corresponding vectors. For example for the case presented in Table 5 we

[8] The application and its sources are available at: https://github.com/volodymyrLeno/BPIC2019

will have $idx_1 = (1, 2, 7, 8, 9, 10, 16, 17, 18)$, $idx_2 = (3, 4, 6, 11, 12, 13, 14, 19)$ and $idx_3 = (5, 15, 20)$.

	Case ID	*Activity*	*Timestamp*
1	2000000015_00001	Record Goods Receipt	2018-01-25T11:16:00.000+10:00
2	2000000015_00001	Record Goods Receipt	2018-02-05T21:46:00.000+10:00
3	2000000015_00001	Record Invoice Receipt	2018-02-07T02:46:00.000+10:00
4	2000000015_00001	Record Invoice Receipt	2018-03-14T03:03:00.000+10:00
5	2000000015_00001	Clear Invoice	2018-03-23T17:24:00.000+10:00
6	2000000015_00001	Record Invoice Receipt	2018-03-23T17:26:00.000+10:00
7	2000000015_00001	Record Goods Receipt	2018-03-26T18:32:00.000+10:00
8	2000000015_00001	Record Goods Receipt	2018-03-26T18:34:00.000+10:00
9	2000000015_00001	Record Goods Receipt	2018-03-26T18:36:00.000+10:00
10	2000000015_00001	Record Goods Receipt	2018-03-26T18:39:00.000+10:00
11	2000000015_00001	Record Invoice Receipt	2018-03-26T19:01:00.000+10:00
12	2000000015_00001	Record Invoice Receipt	2018-03-27T01:05:00.000+10:00
13	2000000015_00001	Record Invoice Receipt	2018-03-27T01:06:00.000+10:00
14	2000000015_00001	Record Invoice Receipt	2018-03-27T02:28:00.000+10:00
15	2000000015_00001	Clear Invoice	2018-04-05T23:51:00.000+10:00
16	2000000015_00001	Record Goods Receipt	2018-04-30T02:26:00.000+10:00
17	2000000015_00001	Record Goods Receipt	2018-04-30T02:28:00.000+10:00
18	2000000015_00001	Record Goods Receipt	2018-04-30T02:30:00.000+10:00
19	2000000015_00001	Record Invoice Receipt	2018-04-30T23:19:00.000+10:00
20	2000000015_00001	Clear Invoice	2018-05-09T22:12:00.000+10:00
..	..	..	..

Table 5. First 20 events of case 2000000015_00001

Afterwards, we can match the events into corresponding triplets. This procedure differs for each data flow as we have to consider their unique requirements. For example, DF1 and DF2 require the execution fo the *Record Goods Receipt* activity, whilst DF3 does not; in DF2 the *Record Invoice Receipt* activity can occur before the *Record Goods Receipt* activity, whilst in DF1 the former always follows the latter.

In DF1, for each element $idx_{1i} \in idx_1$, we take the first element $idx_{2j} \in idx_2$ such that $idx_{2j} > idx_{1i}$, and the first element $idx_{3k} \in idx_3$ such that $idx_{3k} > idx_{2j}$. Together, idx_{1i}, idx_{2j}, and idx_{3k} create a triplet of activities representing the main steps of the P2P process. One *Record Goods Receipt* can be matched only with one *Record Invoice Receipt*, and one *Clear Invoice* activities. Thus, after matching the activities into triplet, their corresponding indexes are removed from the vectors. Continuing the example given in Table 5, we can identify three triplets: (1, 3, 5), (2, 4, 15) and (7, 11, 20). Note that, the other events representing the *Record Goods Receipt* and *Record Invoice Receipt* activities (e.g. events 6, 8, 9) cannot be associated with any *Clear Invoice* activity, meaning that they belong to *incomplete invoices.* In DF1, we consider an invoice incomplete when it is missing one of the three main steps. For the computation of the throughput time we consider only completed invoices.

In DF2, we identify the first activity in the case and then select the corresponding matching order: *Record Goods Receipt* $\rightarrow$ *Record Invoice Receipt* $\rightarrow$ *Clear Invoice* or *Record Invoice Receipt* $\rightarrow$ *Record Goods Receipt* $\rightarrow$ *Clear Invoice*. In the former case, the procedure is the same as for DF1. In the latter case, we match each element of idx_{2i} with the first element $idx_{1j} \in idx_1$ such that $idx_{1j} > idx_{2i}$, and then select the first element $idx_{3k} \in idx_3$ such that $idx_{3k} > idx_{1j}$.

In DF3, we use only the vectors idx_2 and idx_3, given that the *Record Goods Receipt* activity is not observed in DF3. For each element $idx_{2i} \in idx_2$ we select the first element $idx_{3k} \in idx_3$ such that $idx_{3k} > idx_{2i}$, and we match them as couples.

After obtaining the triplets (couples in the case of DF3), we can compute the time between *Record Goods Receipt* (GR), *Record Invoice Receipt* (IR) and *Clear Invoice* (CI) activities simply subtracting their timestamps:

$$time(GR, IR) = IR.timestamp - GR.timestamp \quad (1)$$

$$time(IR, CI) = CI.timestamp - IR.timestamp \quad (2)$$

$$time(GR, CI) = CI.timestamp - GR.timestamp \quad (3)$$

Figure 13 shows the throughput times between the GR and IR activities. Since DF3 does not contain GR activities, we held it out, and we report only the results for DF1 and DF2. As we can observe, in more than 50% of the cases, the throughput time between GR and IR does not exceed 10 days. Whilst, for 80% of the cases the throughput time is in the range 0 to 40 days. Table 6 reports the minimum, maximum, average, and median throughput times, which complement the histograms chart in Figure 13. Although, at a first glance, the throughput time distributions look very similar for DF1 and DF2, we note that (on average) the DF1 case is slower than the DF2 case to *Record Invoice Receipt* after a Goods Receipt is recorded (29.37 days vs. 17.47 days). On the other hand, the maximum throughput time between GR and IR is much greater in the case of the DF2, 325.06 days against the 226.85 days for DF1. The minimum throughput time seems less reliable, especially in the case of DF2 which is equal to 0. This was observed because in some cases (of the DF2) GR and IR are recorded in the event log with identical timestamps. We were not able, though, to understand whether this was a recording error, a compliance issue, or just a normal execution.

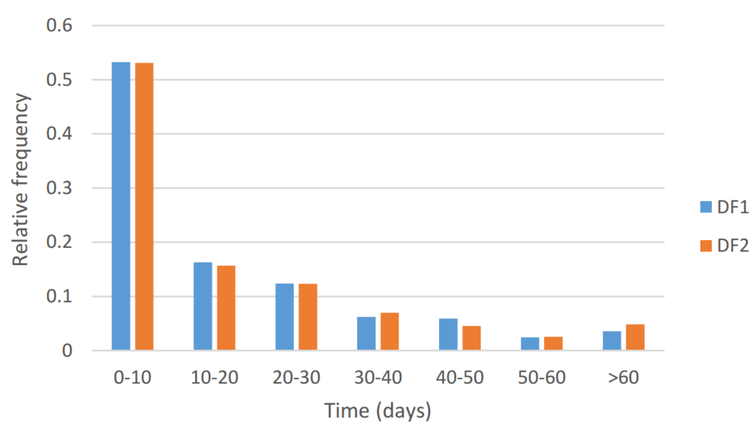


Fig. 13. Throughput times between *Record Goods Receipt* and *Record Invoice Receipt* activities, histograms chart.

	DF1	DF2
MIN	1 minute	0 minutes
MAX	226.85 days	325.06 days
AVG	29.37 days	17.47 days
MEDIAN	20.82 days	8.77 days

Table 6. Throughput times between *Record Goods Receipt* and *Record Invoice Receipt* activities, statistics.

Figure 14 and Table 7 summarises the throughput times between IR and CI. We notice that these throughput times increase substantially with respect to (w.r.t.) the throughput times between GR and IR. Indeed, more than 50% of the cases takes more than 40 days to *Clear Invoice* after the Invoice is recorded. If we consider the average throughput time (in the case of DF2), we can see that it more than doubles w.r.t. the average throughput time between GR and IR (17.47 days against 49.06 days, see Table 7). Even worse it is the case of the median throughput time which in the DF2 increases of 400%, and in the DF1 increases of 40%. On the other hand, the throughput times between IR and CI remain low for the DF3 case.

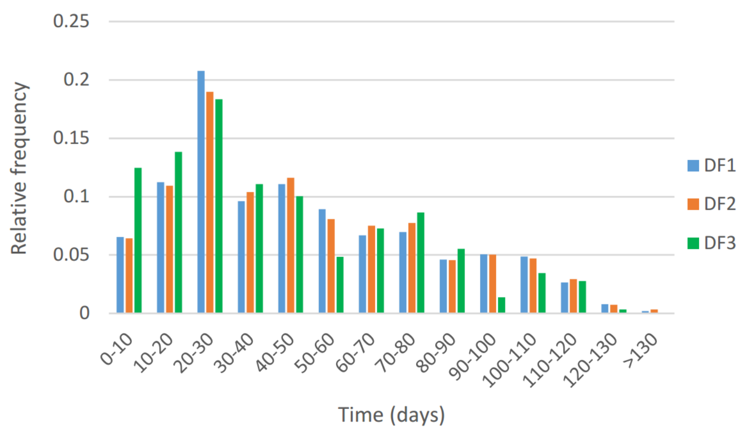


Fig. 14. Throughput time between *Record Invoice Receipt* and *Clear Invoice*, histograms chart.

	DF1	DF2	DF3
MIN	1 minute	0 minutes	2 minutes
MAX	317.46 days	341.94 days	202.1 days
AVG	37.61 days	49.06 days	10.04 days
MEDIAN	27.81 days	42.93 days	5.25 days

Table 7. Throughput time between *Record Invoice Receipt* and *Clear Invoice*, statistics.

Finally, Figure 15 and Table 8 report the throughput times between GR and CI, for the DF1 and the DF2 (not applicable to DF3). In contrast with the throughput times between GR and IR, and IR and CI, in this case both the DF1 and the DF2 throughput times are in line, with average close to 66 days.

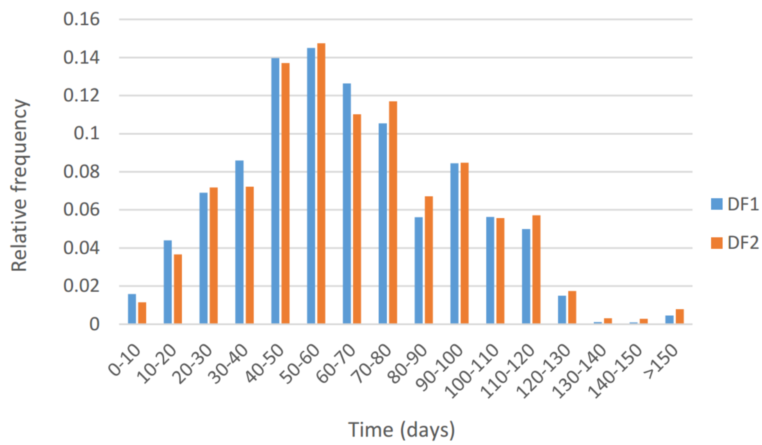


Fig. 15. Throughput time between *Record Goods Receipt* and *Clear Invoice*, histograms chart.

	DF1	DF2
MIN	2.13 hours	2.2 hours
MAX	330.51 days	345.22 days
AVG	66.99 days	65.98 days
MEDIAN	63.16 days	63.01 days

Table 8. Throughput time between *Record Goods Receipt* and *Clear Invoice*, statistics.

Data Flow	Cleared Invoices	Payments per day (on average)	Total net worth (in millions)	Avg. net worth per day (in millions)
DF1	13150	34.55	1243.623	3.267
DF2	180156	473.74	624.145	1.641
DF3	291	0.78	2.022	0.005
Total	193597	508.575	1869.79	4.912

Table 9. Throughput statistics

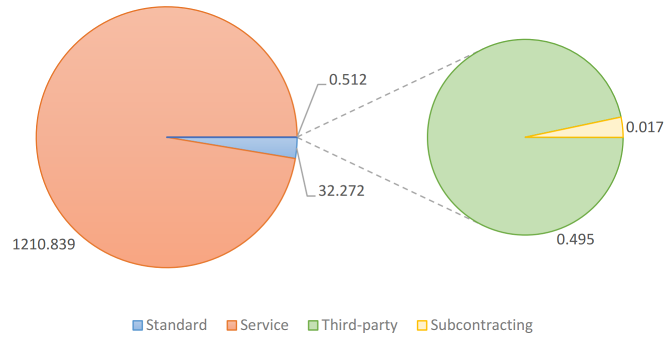


Fig. 16. Net worth per item type for DF1 (in millions).

3.2 Throughput

We define the *throughput* of the P2P process as the average amount of payments (i.e. invoices cleared) performed per day. Consequently, to compute the throughput, we consider only the cases where the *Clear Invoice* activity is observed. Also, we report on the total net worth, and average net worth per day, to compute the net worth we referred to the attribute *Cumulative Net Worth* of the *Clear Invoice* events recorded in the event log. The results are shown in Table 9. Although the number of payments performed in DF2 is the highest, their net worth is less than the net worth of the payments performed in DF1 (being this latter almost double the net worth of DF2).

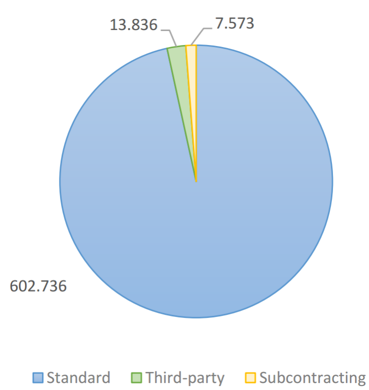


Fig. 17. Net worth per item type for DF2 (in millions).

On average there are around 508 payments performed every day across all data flows. The total net worth is equal to 1.8 billions euros, 66.5% of which is coming from DF1 and 33.4% from DF2 correspondingly. DF3 accounts only for the 0.1% of the total net worth. Furthermore, we broke down the net worth distribution into the different items types. Figure 16 and 17 show the net worth distributions for Df1 and DF2, resp. In DF1 most of the net worth is allocated to the *Service* item type, accounting for 97% of the total. By contrast, DF2 does not handle any payment for *Service* item type, and the most dominant item type is *Standard* with net worth over 602 millions (around 96% of total net worth of DF2). In DF1, the *Standard* item type covers only about 32 millions of net worth, around 2% of the total net worth. In both DF1 and DF2, the *Third Party* item type is characterized by higher net worth than *Subcontracting* (almost 30 times higher in DF1 and 2 times higher in DF2). In DF3, all the payments involve only one item type which is *Limit*.

4 Process Compliance

In this section, we analyse the compliance of the processes w.r.t different objectives, i.e. completeness of cases, exceptional cases, cases with compliance issues and cases with re-work. We exclude data flow DF4 from our compliance analysis, i.e. items of the consignment type, since this process is handled externally and activities from the external company are not recorded in the event log. Thus, an analysis will not provide inside into company internal compliance issues. For the analysis we then consider the BPI challenge log 2019 with a time-frame filter from 31st Dec 2017 - 18 January 2019, as discussed in section 2.2, and an attribute filter removing cases with item type 'consignment. This filtered log maintains 94% of the cases and 97% of all events. For finding the compliance issues, we applied the tool Disco provided by the company Fluxicon.

As a first step, we extract a set of incomplete cases from the event log and categorize them according to different characteristics. For the following analysis, we then remove this set of incomplete cases from the event log. As a next step, we define criteria for classifying cases as exceptional and investigate some example cases. We divided the analysis of compliance issues into two categories: compliance issues resulting from (1) high level control flow issues and from (2) mismatching values of invoice receipts. Finally, we will identify the customers of the process that cause the most rework.

4.1 Incomplete cases

In this section, we consider a case to be incomplete, if it does not start or end with specified start or end activities. For the BPIC19 log, we consider 'Create Purchase Order Item' or 'Create Purchase Order requisition item' and 'Vendor creates invoice' to be relevant starting activities and 'Clear invoice' to be the only relevant end activity of a given case. Both activities 'Create Purchase Order requisition item' and 'Clear invoice' are mentioned in the high-level description for all data flows one to three as necessary activities, i.e. the goods receipts and invoice values need to be matched against the value at the creation of an item and only if these values match, the payment is issued. During the analysis of the data, we identified that sometimes activity 'Create Purchase Order Item' is preceded by either activity 'Create Purchase Order requisition item' or 'Vendor creates invoice' and thus we included this activity as starting activities as well. To filter the BPIC19 log to only maintain complete traces, we apply an endpoint filter with the option to discard traces that have not one of the aforementioned starting or end activities. As a result, we retrieve an Event log that maintains 71% of the cases and 72% of the events. In the following, we will investigate the 23% of the cases that were filtered in this step and try to categorize them according to different criteria.

To maintain only the set of incomplete cases, we apply the endpoint filter with the option to discard cases that have 'Clear Invoice' as an end event. As a result we retrieve an event log with 22% of the cases. We noticed several cases with deletions and cancellations that should be separated from the set of open cases. For that purpose, we used the Attribute filter that requires the all activities with 'Cancel' or 'Delete' in their activity names as mandatory. 4% of all cases were deleted or cancelled after their creation. When we change this filter from the option mandatory to forbidden, we can receive an event log of currently open cases (18%) that are not cancelled or deleted. As there are strict payment deadlines for invoices, usually within one moth, we want to apply a Timeframe filter to divide the set of open cases into cases younger (8%) and cases older than one month (9%). For open cases younger than one month, no action is usually required and the cases can just proceed naturally. For cases that are older than one month usually a follow up or an action is required. For that purpose, we distinguish the set of older cases into cases that never received an invoice (4%) and cases that received and invoice but no payment (5%). Since a relatively high amount of cases is open longer than one month, the data provider could consider to implement payment reminders for cases that received invoices or send follow-up requests for invoices to the vendors for cases without invoices.

Table 10 summarizes the filter options applied and table 11 summarizes the resulting event log sizes. Moving forward, we will focus the analysis of compliance issues on the event log of complete cases, i.e. after applying filter (1).

4.2 Exceptional Purchase Order Documents

We deem a PO document to stand out from the event log, if it has a high number of events, i.e. is in the top 1% of cases when sorting the cases according to their number of events. Alternatively, it is possible to use case durations for the identification of exceptional cases. In this section, however, we focus more on the analysis of compliance issues and a high number of events is usually an indicator for compliance issues. To identify exceptional cases, we apply a performance filter by the number of events. We identify the cut-off for the 1% of cases with the highest number of events for cases that have a minimum number of 20 events. The 1% of cases cover 5% of the events of the BPIC19 log. Fig. 19 shows the top 5 exceptional cases. These cases have more than 300 events and some cases have a duration of over one year. We also discovered a high-level process map with the tool Apromore for 50% of the activities and 5% of the arcs. It is noticeable that these cases contain a lot of rework, i.e. several 'Change' activities and also cancelled goods receipts or invoices. These cases can be used for further analysis with domain knowledge to identify causes for these rework activities.

Case ID	▲ Events	Variant	Started	Finished	Duration
4507000436_00010	513	Variant 69	03.01.2018 01:33:00	11.01.2019 01:45:00	1 year, 8 days
4507000449_00060	500	Variant 73	03.01.2018 02:20:00	18.01.2019 00:41:00	1 year, 14 days
4507013395_00001	458	Variant 348	02.03.2018 08:45:00	07.06.2018 21:13:00	97 days, 13 hours
4507000684_00010	399	Variant 78	04.01.2018 00:23:00	18.01.2019 00:41:00	1 year, 14 days
4507036789_00001	345	Variant 676	15.06.2018 06:44:00	06.09.2018 22:09:00	83 days, 15 hours

Fig. 18. Top 5 exceptional cases.

Filter ID	Description	Filter type	Options	Filter parameters
(1)	Filter for complete cases	Endpoints Filter	Discard cases	**Start event values:** 'Create Purchase Requisition Item', 'Create Purchase Order Item', 'Vendor creates invoice' **End event values:** 'Clear Invoice'
(2)	Filter for incomplete cases	Endpoints Filter	Discard cases	**Start event values:** any **End event values:** any besides 'Clear Invoice'
(3)	Filter for cancelled or deleted cases	Attribute Filter	Mandatory	**Event values:** 'Cancel Goods Receipt' 'Cancel Invoice Receipt' 'Cancel Subsequent Invoice' 'Delete Purchase Order Item'
(4)	Filter for open cases	Attribute Filter	Forbidden	**Event values:** 'Cancel Goods Receipt' 'Cancel Invoice Receipt' 'Cancel Subsequent Invoice' 'Delete Purchase Order Item'
(5)	Filter for cases with invoices	Attribute Filter	Mandatory	**Event values:** 'Record Invoice Receipt'
(6)	Filter for cases without invoices	Attribute Filter	Forbidden	**Event values:** 'Record Invoice Receipt'
(7)	Filter for cases before 18th of Dec.18	Timeframe Filter	Keep cases: Completed in Timeframe	**Start value:** 31/12/17 **End value:** 18/12/18
(8)	Filter for cases after 18th of Dec.18	Timeframe Filter	Keep cases: Completed in Timeframe	**Start value:** 18/12/18 **End value:** 12/01/1

Table 10. Applied Filters for identifying and classifying incomplete cases

Category	**Filters**	**%Cases**	**%Events**
Complete cases	(1)	71%	72%
Incomplete cases	(2)	22%	24%
Cancelled or deleted cases	(2) & (3)	4%	3%
Open cases younger than one month	(2) & (4) & (8)	8%	6%
Open cases without invoice older than one month	(2) & (4) & (6) & (7)	4%	9%
Open cases with invoice older than one month	(2) & (4) & (5) & (7)	5%	4%

Table 11. Event log sizes after filtering for incomplete cases

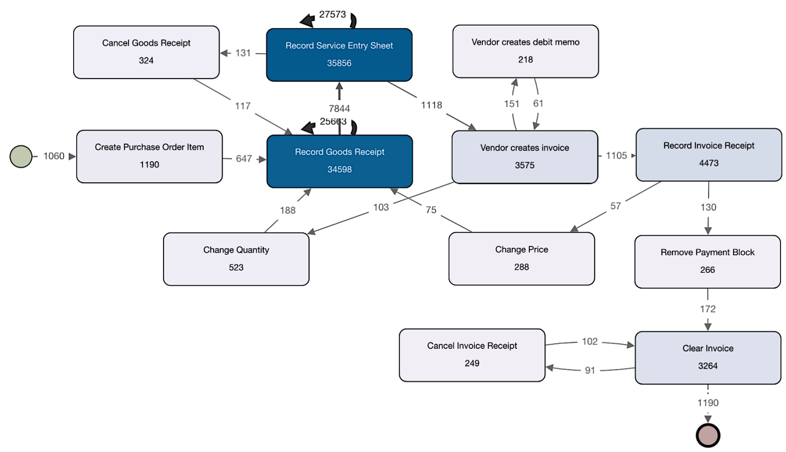


Fig. 19. Exceptional cases process map discovered with the tool Apromore.

4.3 Compliance issues from a control flow perspective

Typically, compliance issues from a control flow perspective when comparing event logs with process models can be assessed with techniques such as trace alignments [1]. The process models from section 2, however, support a wide range of BPMN elements that are not currently supported by these trace comparison techniques, i.e. events, sub-processes or message flows. Since removing these elements of the process models would not properly represent the processes developed in section 2 any more, we decided to not apply common conformance checking techniques and rather fall back to high level analysis of compliance issues with filters.

Activity	Resource
Create Purchase Order Item	user_038
Vendor creates invoice	NONE
Record Invoice Receipt	batch_00
Remove Payment Block	user_006
Record Goods Receipt	user_029
Clear Invoice	user_005

Fig. 20. Example Case for preemptive removal of payment block (ID 4507000254_00010).

At first, we checked high level constraints of data flows one and two given in the description of the challenge. In particular, in DF1 an invoice can only be cleared, if an invoice is received after the goods are received. Hence, if an invoice would be received for this item category before the goods are received, this case would cause a compliance issue. We tested for these cases by applying an attribute filter for the corresponding item category and a follower filter, i.e. filtering for cases where activity 'Record Invoice Receipt' is eventually followed by activity 'Record Goods Receipt'. When investigating the resulting 306 cases, however, we realized that most cases just resolve around several goods and invoice receipts. When filtering these cases with another follower filter, i.e. 'Record Goods Receipt' is never eventually followed by 'Record Goods Receipt', no cases remained, i.e. this compliance rule is never violated.

Another possible compliance issue is described in DF2: If an invoice is received before the goods receipt, then an automatic payment block needs to be removed before the invoice can be cleared. It would be a compliance issue,

Activity	Resource
Create Purchase Requisition Item	user_292
Create Purchase Order Item	user_136
Record Invoice Receipt	user_012
Vendor creates invoice	NONE
Record Goods Receipt	user_136
Clear Invoice	user_015

Fig. 21. Example Case of missing activity 'Remove Payment Block' (ID 4508069895_00010).

if the activity 'Remove Payment Block' is missing. We check for this compliance issue by applying an attribute filter for item category of DF2, a follower filter, where activity 'Record Invoice Receipt' is eventually followed by activity 'Record Goods Receipt', and another follower filter, where activity 'Record Goods Receipt' is never eventually followed by 'Remove Payment Block'. In addition, we apply an additional follower filter to remove cases with duplicate invoices as those cases interfered with the analysis.

As a result, we found 291 cases with compliance issues. When investigating these cases, we found that activity 'Remove Payment Block' was sometimes executed before the goods were received, which can also be observed in Fig. 20. We can filter for these cases with another follower filter, i.e. where 'Remove Payment Block' was eventually followed by activity 'Record Goods Receipt'. Cases with pre-emptive removal of the payment block make up 263 out of 291 cases. The remaining 28 cases, where activity 'Remove Payment Block' was missing, can be identified by choosing the option 'never eventually followed' for the last mentioned filter. Fig. 21 shows a sample case, where activity 'Remove Payment Block' is missing.

Activity	Resource
Create Purchase Order Item	user_142
Vendor creates invoice	NONE
Record Goods Receipt	user_080
Vendor creates invoice	NONE
Record Invoice Receipt	user_007
Clear Invoice	user_005
Clear Invoice	user_005

Fig. 22. Example Case of duplicate activity 'Clear Invoice' (ID 4507002162_00010).

We also found other compliance issues: When applying a follower filter, when 'Clear Invoice' is directly followed by itself, we can find cases where an invoice was paid twice. We again filter out multiple invoices since two cleared invoices could also be correct for cases with two invoice receipts. As a result, we identified 626 cases of duplicate payments. Fig. 22 shows an example case with a duplicate 'Clear Invoice' activity for a single invoice.

Activity	Resource
Create Purchase Order Item	user_103
Cancel Goods Receipt	user_104
Vendor creates debit memo	NONE
Cancel Invoice Receipt	batch_00
Clear Invoice	user_111

Fig. 23. Example Case of a payment without a valid invoice (ID 4507002401_00010).

Another possible compliance issue revolves around cancelled invoices that are still cleared. We can identify these cases with a follower filter, where activity 'Cancel Invoice Receipt' is directly followed by the activity 'Clear invoice'. In a total of 2,492 cases the invoice was cleared directly after a cancel event occurred. We identified two different kinds of cases with different problems: Fig. 23 shows a sample case, where a cancelled invoice is cleared without ever receiving a valid invoice are goods receipt. We can filter for these cases by applying another follower filter, where 'Clear Invoice' is never eventually followed by another 'Record Invoice Receipt' event. 769 out of 2,492 cases represent this case. The other kind of cases eventually receives a correct invoice that is cleared later on, for example shown in Fig. 24. This kind

Activity	Resource
Create Purchase Order Item	user_000
Vendor creates invoice	NONE
Vendor creates debit memo	NONE
Record Goods Receipt	user_000
Record Invoice Receipt	user_007
Cancel Invoice Receipt	user_006
Clear Invoice	user_006
Record Invoice Receipt	user_006
Clear Invoice	user_002

Fig. 24. Example Case of receiving a correct invoice after clearing a cancelled invoice (ID 2000000185_00001).

of cases make up the remaining 1,723 out of 2,492 cases. The second type of wrongly cleared invoice is less severe since the error is later corrected in the case. However, it is possibly beneficial to implement security procedures to prevent clearing invoices right after an invoice cancellation.

4.4 Compliance issues regarding wrong invoice values

Compliance issues resulting from wrong invoice values can be identified with a Follower filter. Particularly, we filter for cases, where one of the events 'Create Purchase Order Item', 'Record Goods Receipt' or 'Record Invoice Receipt' are eventually followed by one of the events 'Record Goods Receipt', 'Record Invoice Receipt' or 'Clear Invoice' and that require a different value for the cumulative net worth of the invoice. As a result, we retrieve 1,529 cases (<1%), where invoice values are wrong. We further investigated different case variants, when mismatching invoice values occurred. The most common case is depicted in Fig. 25, where the values of several goods receipts are summed up in one invoice receipt. The invoice for these cases, however, is then cleared for the value of a single goods receipt, which either indicates a wrong invoice value or a wrong payment amount.

Activity	Resource	Cumulative net worth (EUR)	User
Create Purchase Order Item	batch_06	595.0	batch_06
Record Goods Receipt	batch_06	595.0	batch_06
Record Goods Receipt	batch_06	595.0	batch_06
Record Goods Receipt	batch_06	595.0	batch_06
Record Goods Receipt	batch_06	595.0	batch_06
Record Service Entry Sheet	NONE	595.0	NONE
Record Service Entry Sheet	NONE	595.0	NONE
Record Service Entry Sheet	NONE	595.0	NONE
Record Service Entry Sheet	NONE	595.0	NONE
Vendor creates invoice	NONE	595.0	NONE
Record Invoice Receipt	user_007	2382.0	user_007
Clear Invoice	user_002	595.0	user_002

Fig. 25. Example Case for multiple goods receipts with a summary invoice (ID 4507001872_00001) .

Another common case can be found in Fig. 26, where the goods receipt value is exactly twice the amount of the corresponding invoice and payment of the case.

Activity	Resource	Cumulative net worth (EUR)	User
Create Purchase Order Item	user_036	88.0	user_036
Vendor creates invoice	NONE	88.0	NONE
Record Goods Receipt	user_029	177.0	user_029
Record Invoice Receipt	user_024	88.0	user_024
Clear Invoice	user_002	88.0	user_002

Fig. 26. Example Case for wrong goods receipt (ID 4507000855_00080) .

One last common case with wrong invoice values is shown in Fig. 27, where multiple goods receipts show values unrelated to the PO item during creation of the case. The invoice then similar to the case in Fig. 25 sums up the goods receipt values, but assumes that the goods receipt values are equal to the value during creation of the PO. The payment, however, is then again equal to the amount during the creation of the invoice. While the amount of compliance issues from wrong invoice values is low, further analysis should be conducted with more domain knowledge.

Activity	Resource	Cumulative net worth (EUR)	User
Create Purchase Order Item	batch_06	185.0	batch_06
Record Goods Receipt	batch_06	1295.0	batch_06
Record Goods Receipt	batch_06	9067.0	batch_06
Record Service Entry Sheet	NONE	185.0	NONE
Record Service Entry Sheet	NONE	185.0	NONE
Vendor creates invoice	NONE	185.0	NONE
Record Invoice Receipt	user_008	370.0	user_008
Clear Invoice	user_002	185.0	user_002

Fig. 27. Example Case for unrelated multiple goods receipt values (ID 4507001538_00001) .

4.5 Identifying Customers with the most rework cases

One question of the challenge was to identify the customers that cause the most amount of rework. From a compliance perspective, we consider a case to cause rework, if it contains at least one of the rework activities from Table 2. We can filter for all cases with rework by applying an attribute filter with the option to maintain only cases that contain at least one of the six mandatory rework activities. 8% of the cases contain rework activities. We consider the vendors to be the customers of the process. Thus, we can identify the customers causing the most amount of rework by using the statistics view of Disco for the attribute 'Vendor'. Fig. 28 shows the summary statistics after applying the filter and especially the top 5 vendors with the highest amounts of frequencies. The amount of rework cases of the top 5 vendors (>3000 cases) is rather significant considering an average of only 162 cases per vendor with a standard deviation of 559 cases.

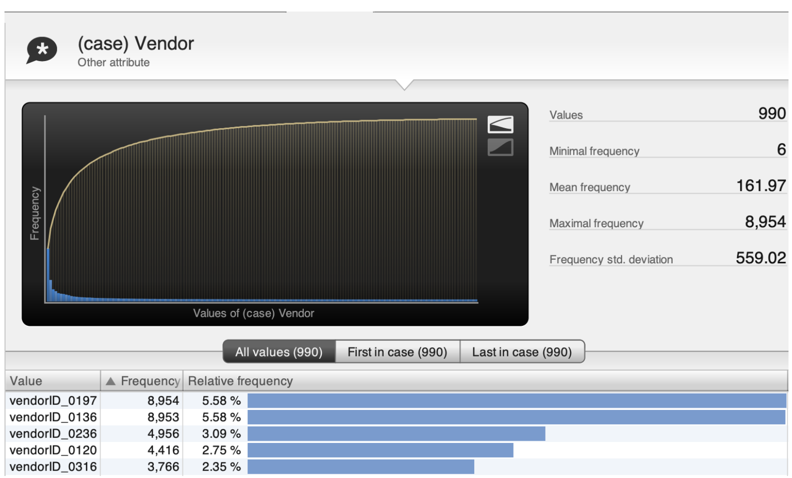


Fig. 28. Top 5 vendors with rework cases.

5 Conclusion

In this paper, we analysed the event log of a Purchase-to-Pay (P2P) process from a large multinational company operating from The Netherlands, provided for the BPI Challenge 2019. Our aim was threefold; (i) to derive a collection of process models explaining the P2P process recorded in the even log; (ii) to analyse the throughput of the P2P process; and (iii) to identify compliance issues present in the event log that conflict with the high-level control-flow descriptions.

First, we analysed the event log and the activities recorded within. We identified four different data flows, and discovered the respective as-is BPMN process models. From our understanding of the data, we proposed a set of to-be BPMN process models that capture not only the reference behavior but integrate also the extra behavior observed in the event log.

Then, we developed a technique to match each *Goods Receipt* with the corresponding *Invoice*, and this latter to its *Clearance* event. Based on these matchings, we estimated the throughput times and the throughput in terms of total invoices cleared per day and total net worth. We learned that the throughput times between *Goods Receipt* and *Invoice Receipt* in more than 50% of the cases did not exceed 10 days. Whilst the throughput times between *Invoice Receipt* and *Clear Invoice* exceed most of the times the 40 days. Regarding the throughput, we found that 508 payments are processed with an average net worth of 4.9 millions per day.

Lastly, we analysed compliance issues with respect to the high-level process control flows descriptions. We provided a categorization of incomplete cases and recommended to implement follow-ups for open cases that go beyond the usual payment deadlines of invoices. For the set of completed cases, we gave a more detailed analysis of compliance issues by finding a set of exceptional cases and by identifying the top five customers causing the most amount of rework. When analysing compliance with regards to the control flow, we found cases with double payments, payments for previously cancelled invoices and wrongly removed payment blocks for invoices. Finally, we identified a small set of cases with wrong invoice values and gave some characteristics of the problematic cases.

References

1. A. Adriansyah, J. Munoz-Gama, J. Carmona, B. van Dongen, and W. van der Aalst. Alignment based precision checking. In *BPM*. Springer, 2012.
2. A. Augusto, R. Conforti, M. Dumas, M. La Rosa, and A. Polyvyanyy. Split miner: automated discovery of accurate and simple business process models from event logs. *KAIS*, 2018.
3. W. van der Aalst. *Process Mining - Data Science in Action*. Springer, 2016.
4. B.F. van Dongen. *Dataset BPI Challenge 2019. 4TU.Centre for Research Data.* https://doi.org/10.4121/uuid:d06aff4b-79f0-45e6-8ec8-e19730c248f1.

Process Mining in the Coatings and Paints Industry: The Purchase Order Handling Process

Business Process Intelligence Challenge 2019

Peyman Badakhshan, Sam Gosling, Jerome Geyer-Klingeberg, Janina Nakladal, Johannes Schukat, Jennifer Gsenger

Celonis SE, Theresienstraße 6, 80333 Munich, Germany
{p.badakhshan, s.gosling, j.geyerklingeberg, j.nakladal, j.schukat, j.gsenger}@celonis.com

Abstract. The Business Process Intelligence (BPI) challenge is an annual competition in process mining that is co-located with the International Conference on Process Mining (ICPM). BPI 2019 is providing participants with a real event log from the Purchase-to-Pay process of a Dutch company in the field of coatings and paints. The process owner is interested to gain insights into the process from three perspectives. First, discovering various process models referring to different use cases. Second, focusing on throughput time of the invoicing process. Third, detecting deviations based on the expected process model. The aim of this paper is to report the insights and results derived from a comprehensive process analysis using the Celonis Intelligent Business Cloud (IBC). The report also discusses limitations due to process and data characteristics and outlines recommendations for additional data collection and analysis.

Keywords: Process Mining, Process Improvement, Process Discovery, Conformance, Purchase-to-Pay, BPI Challenge, Celonis.

1 Introduction

The Purchase-to-Pay (P2P) process is a challenge for all organizations. This challenge starts with the creation of purchase requisitions and their time-consuming approval processes. It follows with supplier relationship management and ends with handling the invoices and their respective related subprocesses. Thereby, it is crucial for organizations to deal with lack of ownership and maintain good master data on suppliers, enable a proper process and data governance, provide suppliers an automated ordering process, better contract management, and better visibility of payment processing times.

Process mining has proved to be an innovative and efficient method to discover, analyze, and predict business process behavior [1]. For the analysis of the BPI process data, we structure our work by two perspectives. On the one hand, we consider the three main types of process mining including discovery, conformance checking, and enhancement. The process mining types, as described in Figure 1, enable us to discover the as-is process, to conduct process analytics, and to recommend necessary actions for process improvement.

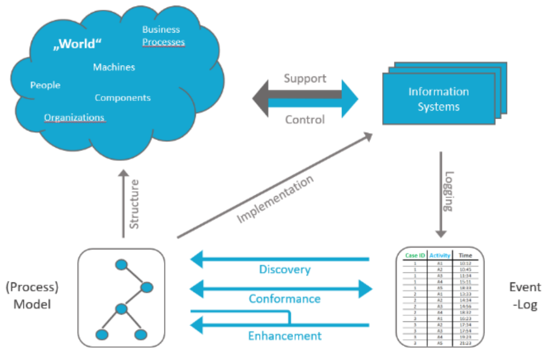


Fig. 1. Process mining overview [1]

On the other hand, and in order to handle the big data provided for the challenge, we consider four types of analysis: Descriptive, Diagnostic, Predictive, and Prescriptive. The descriptive analysis supports us to understand what happened, while diagnostic analysis supports us to understand why it happened like that. The predictive analysis points to what is likely to happen, and prescriptive analysis provides guidelines on what can be done to achieve expected results from the process. The second perspective is adopted from [2] and illustrated in Figure 2.

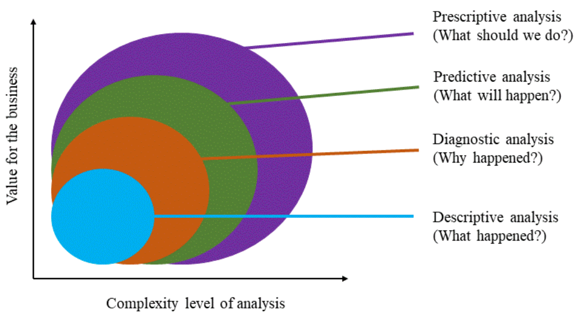


Fig. 2. Four types of big data analytics

The aim of this paper is to analyze the provided process data in order to develop intelligible insights and to answer the process owner's questions. Based on the two mentioned perspectives, we will address these questions as described in Table 1.

Table 1. Analysis overview

No.	Question	Our approach		Conducted Action/Analysis
		First perspective	Second perspective	
1	What is the best collection of process models? - At least four process models needed for four process flows - Best explaining model for each item depends on its properties	Discovery	Descriptive analysis	Process discovery with Celonis discovery algorithm, filtering, and PQL (Process Query Language) functions
2	What is the throughput of the invoicing process?	Enhancement	Predictive analysis, Prescriptive analysis	Cycle time analysis, advanced invoice matching
3	Which Purchase Documents stand out from the log? AND Where are deviations?	Conformance checking	Descriptive analysis, Diagnostic analysis	First-time-right analysis, Celonis features including Conformance checking [7], Process AI and root cause analysis

The remainder of this paper is structured as follows. Section 2 starts with a management summary of the applied analyses, their main findings, and the overall recommendations for process improvement. In Section 3, we explain our understanding of the challenge including the data and the process to show the logic behind our analysis approach. Afterwards, we comprehensively explain the analysis and the results regarding the process owner's questions in Section 4. Section 5 concludes with the limitations.

2 Management summary

Using various process mining methods, we answer the process owner's questions and provide several suggestions with the aim of giving more insights to the process owner. Regarding process discovery, we report one general model that represents the full process along with four additional models representing in-depth views. This process model satisfies four different scenarios including "3-way matching, invoice after goods receipt", "3-way matching, invoice before goods receipt", "2-way matching (no goods receipt needed)", and "Consignment". We also design the models using BPMN notation and use them as a reference model for conformance checking.

Regarding conformance checking, we identify violations to the reference process using root cause analysis and the Celonis Process AI feature along with detecting deviations from the expected process model. We also take advantage of the First-Time-Right (FTR) analysis for further detection of compliance issues. Finally, we approach the enhancement of the process using cycle time analysis.

2.1 Analysis overview

In the following section, we explain the applied analyses as well as the main findings.

Cycle time analysis. With the recent focus on cost reduction, many organizations are forced to identifying opportunities for long-term savings. One way to cut costs is by decreasing the cycle times associated with procuring materials and services. Cycle times provide critical information on an organization's procurement efficiency. With the help of process mining, purchasing organizations can realize shorter cycle times (i) by making their and their suppliers' procurement efficiency transparent, and (ii) by suggesting measures to improve this efficiency

The cycle time analysis on the provided data set and specific activities is summarized as the following:

1. The average throughput time from "Record Goods Receipt" to "Record Invoice Receipt" is 19.9 days; the median throughput time is 9.8 days.
2. The average duration between the creation of the invoice in the source system (Record Invoice Receipt) and the clearing of the invoice is 47.9 days; the median is 41.9 days.

First-Time-Right (FTR) analysis. Procurement quality means that the procurement organization ensures to procure the right product, in the right quantity and quality, in the right place, at the right time, and at the best price. The goal of the FTR analysis is to evaluate how well the purchasing process really works - by calculating how often purchase orders run through the process without being touched by rework to change incorrect orders or recurrent activities. Importantly, those cases that require additional effort are analyzed and their root causes and/or potential drivers are identified.

A main finding from this analysis is that many cases are labeled with their respective classification, however, they are used to clear debit memos created by a vendor. The suggestion here is to create a specific memo clearing process, which enables an accurate conformance calculation for these types of process flows.

Another finding is that for those cases where rework activities occur, the process flow conformance ratio increases. This is interesting as the ROI for additional manual effort can be tracked, i.e. when an employee makes a price change after receiving the invoice, it is done so to comply with the process flow process. It is advised that the company investigates further what the root causes of these changes are, which has already been started with the analysis below.

Conformance checker. The conformance checker allows checking the expected/ideal process model against the process discovered from the recorded data. Conformance checking enables organizations to get an overview of cases that performed correctly as well as detecting the violations that appeared in execution. Conformance checker of Celonis is empowered by the Process AI feature for identifying the root causes assigned to each violation and it supports process analysts to decide on where they should focus on their process improvement initiatives.

Conformance checker and FTR are complementary analyses to answer compliance related questions. The main finding of the conformance checker is designing the expected process model and aligning with the discovered process to detect violations. Additionally, the potential root causes (e.g. EBELP, user, LIFNR, etc.) of process violations are investigated.

2.2 Recommendations on dealing with violations

A main interest of the process owner is to adhere to process compliance. Therefore, we identify dominant violations in the process affecting the performance of each process step. In the following, we explain these violations along with their causes and business impact.

Price changes. Price changes in the procurement process slow down business efficiency, as they are typical rework activities. At the same time, the manual effort associated with price changes significantly increases process costs. Price changes also impact a company's revenue forecast and cash flow and may also indicate unauthorized or inappropriate discounting. Therefore, it is in the corporate interest to detect the root cause of price changes and to take actions against it [8]. Price changes caused by manual users are of highest impact, often resulting from errors in master data or manual entry errors of free-text orders. Manual price changes usually slow down the process and increase the possibility of manual input errors. Free-text-orders refers to Purchase Order (PO) items that have little or no material master data, requiring manual entry of names, codes and prices, which increases the potential of price changes and debit memos. We recommend checking various price change types including multiple price changes as well as price changes that are happening after receiving the invoice, receiving goods, after order confirmation, and before sending the purchase order.

Quantity changes. Quantity changes occur before the goods receipt, e.g. due to an incorrect entry in the purchase order, or after goods receipt, for instance, due to frequently occurring over-runs. The problem with quantity changes in the purchasing process is that they slow down the process, which reduces planning predictability. At the same time, the significant manual effort involved in changing quantities also significantly increases process costs. Therefore, it is recommended to focus on reducing the ratio of orders with quantity changes.

Remove payment block. A smooth and fast process is important in order to pay incoming invoices in time and avoid overdue notices as well as the associated costs of the dunning charges. Setting unnecessary payment blocks, e.g. due to discrepancies within the invoice positions, disturbs the procurement process. It causes longer throughput times, which can result in payments after the cash discount due date or even after the overall due date. In addition, unnecessary payment blocks increase the amount of manual rework activities, which is also associated with higher costs. Therefore, minimizing the amount of unnecessary payment blocks is a desirable objective. For the case of the "3-way match, invoice before goods receipt" however, the removal of the payment block is recognized as a natural part of the process, used for compliance checking.

Cancel invoice receipt. Liquidity is a measure of the ability of a company to meet its current liabilities with its current assets. An important aspect in process efficiency is the avoidance of rework activities on invoice position level as these are often the root of manual effort and slow the process down. In order to ensure a smooth and fast invoice position processing, it is of great interest to identify invoice positions that were correctly processed at the first time and adapt the handling applied in these cases. One way of increasing liquidity is to investigate the root causes of cancelling invoice receipts and decrease them as much as possible. Some recommendations in dealing with this activity can be described as follows:

- Create a guideline showing the different steps of invoice positions and how to evaluate them.
- Organize training sessions and demonstrate how to implement the best practice and how to evaluate current invoice positions.
- Continuously identify First-Time-Right rates and send out automated reports to all sales and/or accounting employees.

Delete purchase order items. Deleting purchase order items is another rework activity and a barrier preventing a fluid process. It is important to identify and reduce the deleted purchase order items and contact the suppliers. Four important causes of this activity are described as follows. First, the absence of a performance-based reward/penalty scheme with various network vendors. Second, a high dependency on different internal functions as well as suppliers to bring the orders to completion. Third, a high lead time between the receipt of a customer order and the order entry into the respective Order Management System. Fourth, a high number of orders with incomplete or incorrect information. We recommend taking the following actions in order to decrease the occurrence of deleted purchase order items:

1. Assign a responsible team/person, set a response time limit, and set a deadline for resolution.
2. Optimize the work processes across sales, production and logistics to avoid organizational silos.
3. If necessary, reorganize existing processes or add additional ones in order to further reduce error probability.

3 Understanding of the challenge

Business Process Intelligence Challenge 2019 (BPI 2019) has collected data from a Dutch organization and its sixty subsidiaries in the business of coatings and paints regarding their P2P process [3]. In this section, we present our understanding of the provided data and the process. This understanding is the base for the analysis and interpretation of the process in order to better respond to the process owner's questions.

3.1 Understanding of the data

The event logs are provided by the process owner in XES- and as CSV-format [4]:

- XES: https://data.4tu.nl/repository/uuid:d06aff4b-79f0-45e6-8ec8-e19730c248f1
- CSV: http://icpmconference.org/wp-content/uploads/BPIChallenge2019CSV.zip

The data refers to the P2P process excluding approval steps for purchase orders and invoices. The data covers purchase orders submitted in 2018 and includes various categories of services, goods, and different types of vendors. Table 2 presents the attributes of the event logs and their descriptions.

Table 2. Overview of the data attributes

Attribute	Description
Concept:name (Case Key)	A combination of the purchase document ID and the item ID
Purchasing document	The purchasing document ID
Item	The item ID
Item type	The type of the item
GR-Based Inv. Verif.	Flag indicating if GR-based invoicing is required
Goods receipt	Flag indicating if 3-way matching is required
Source	The source system of this item
Document category name	The name of the category of the purchasing document
Company	The subsidiary of the company from where the purchase originated
Spend classification text	A text explaining the class of purchase item
Spend area text	A text explaining the area for the purchase item
Sub spend area text	Another text explaining the area for the purchase item
Vendor	The vendor to which the purchase document was sent
Name	The name of the vendor
Document Type	The document type
Cumulative Value	The cumulative value of the item at each step in the process
Item Category	There are four categories including 3-way with GR-based invoicing, 3-way without GR-based invoicing, 2-way, and consignment

The event log contains 76,349 purchase documents, which includes 251,734 items (251,734 cases). This dataset is suitable to perform various explorative and descriptive analyses of the P2P process.

3.2 Understanding of the process

The purchasing process is a business process that covers activities of requesting, purchasing, receiving, as well as paying for goods and services. There are various processes that are inter-related to perform the full purchasing process. To this end, we can address purchase requisitions and accounts payable as two important processes that are handling the purchasing process from an end-to-end perspective by connecting procurement and invoicing.

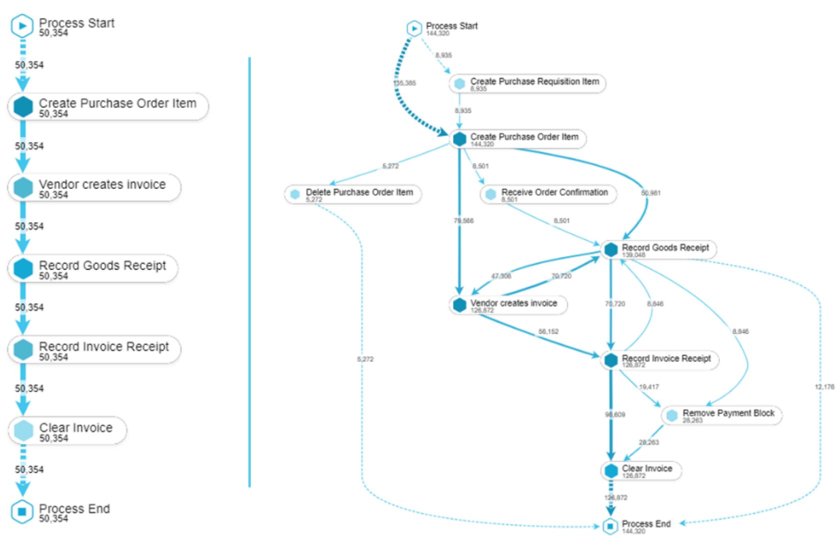
The process typically starts when there is a purchase request. The purchase requisition (PR) is a request from employees when they need to make a purchase on behalf of their company. After creating the PR, several approval steps are conducted, and the decision is made whether to accept the requisition. If the requisition is accepted, it will go through the purchasing process.

Various roles are participating in the PR processes. The "Preparer" creates purchase requisitions for himself or other users. "Approvers" authorize requisitions based on different rules. "Procurement agents" approve and finalize the requisitions from undefined categories within the supply chain and perform collaborative requisitioning. "Administrators" maintain approval rules and provide basic technical support. PRs are playing an important role from different perspectives. First, they are starting points of the purchasing process and can always serve as evidence of communication among employees as well as suppliers. Second, PRs are used as control tools in order to avoid frauds, such as personal usage of items, as well as controlling if the requested material is needed for the organization. Third, PRs support the centralization of the procurement process through the purchasing department. This makes the management of the process much easier. When the requisition is approved, it changes from PR to purchase order (PO).

The PO is sent to the vendor, where it is investigated for legitimacy and accuracy, and if it is accepted, the goods are sent to the purchasing company. The invoice of the purchase is created by the vendor either after or before the buyer receives the goods. After receiving the goods, there are two main activities to be performed. The invoice is booked, and the payment is cleared. This part of the process, which is concerned with receiving invoice and payment is also known as Accounts Payable (AP) process. This part of the process is mainly concerned with avoiding late payments to prevent penalties and extra fees.

Using the event logs and connecting the process with Celonis, the main process model (most frequent process flow) is discovered which is shown in Figure 3.

Fig. 3. General process model: most frequent variant (left); nine most frequent variants (right)



According to the models in Figure 3, the activities of the application are perceived as described in the subsequent Table 3.

Table 3. Overview of the process activities

Attribute	Description (our understanding)
Create purchase requisition item	Internal request for items to be purchased.
Create purchase order item	The accepted purchase requests become purchase orders and are sent to the vendor.
Delete purchase order item	Deleting purchase orders. This can be caused by various reasons.
Receive order confirmation	The vendor confirms the receipt of the order.
Vendor creates invoice	Vendor creates an invoice for the ordered items.
Record goods receipt	When the orders are delivered, a receipt will be recorded that proves the order is received.
Record invoice receipt	The invoice of the order is recorded in the system.

Remove payment block	Invoices might become blocked for various reasons. When the issues are solved the blocks should be removed.
Clear invoice	Clear invoice is referring to the complete payment of the orders.

4 Challenge analysis and results

Before conducting detailed analyses, we must understand the main questions of the process owner. Therefore, in this section, we first reflect our understanding of the process owner's questions and identify the most relevant analyses to respond to the questions. These analyses will be presented afterwards along with the results and suggestions for improvement. At the end of this section, we recommend further analyses that could beneficial for this process beyond the challenge.

4.1 Process owner's questions

The provided data covers holistic information on the P2P process; however, the process owner is particularly interested in compliance questions. Accordingly, there are three main questions proposed in the challenge. In this section, we explain each question, reflect our understanding, and suggest the relevant analysis.

Question 1:

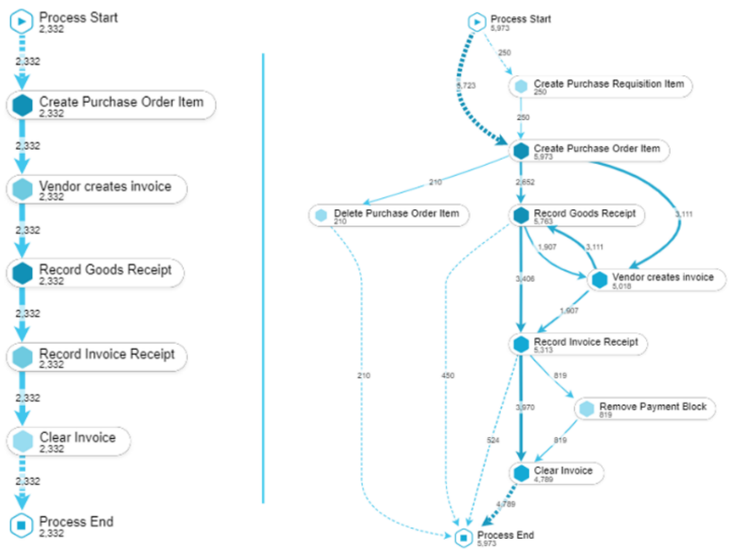
Question. "Is there a collection of process models which together properly describe the process in this data. Based on the four categories above, at least 4 models are needed, but any collection of models that together explain the process well is appreciated. Preferably, the decision which model explains which purchase item best is based on properties of the item."

Our understanding. This question is mainly concerned with discovering the best process model that describes the process accurately. Four categories are recommended by the process owner to be considered in discovering the process as outlined in the following.

1-) 3-way matching, invoice after goods receipt. *"For these items, the value of the goods receipt message should be matched against the value of an invoice receipt message and the value put during the creation of the item (indicated by both the GR-based flag and the Goods Receipt flags set to true)."*

We filter the process on the cases with both "EKPO.GR_VERIF" and "EKPO.GR" values equal to "1", which indicates cases with "GR_based Inv.Verif." flag and "Goods Receipt". This filtering results 15.2k cases out of 252k and the process is discovered as shown in Figure 4.

Fig. 4. Process models (with GR_VERIF & GR): Most frequent variant (left); nine most frequent variants (right)



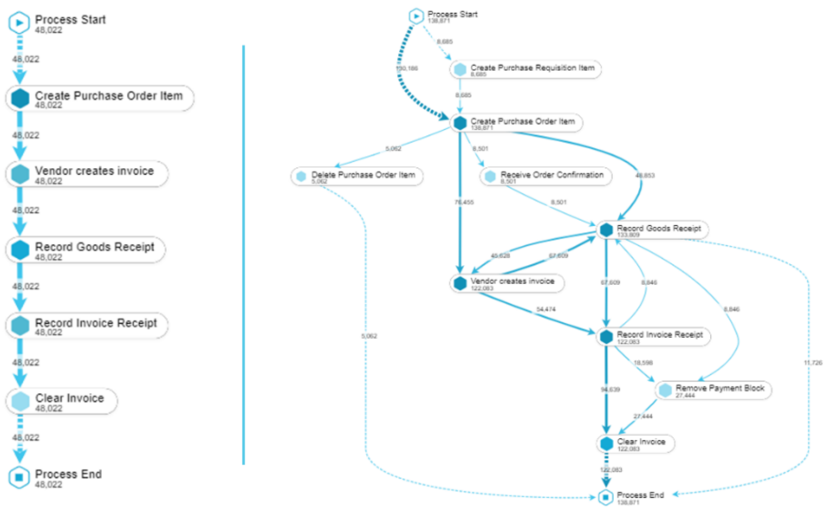
As presented in Figure 4, the first nine variants of the process model cover 6k cases (referring to 39% of related cases) that show the high level of inefficiencies within the process execution. In addition to the main activities, the model includes extra activities that are not violations such as “Remove Payment Block” (820 cases), which is necessary to be done for the clearing of the invoice. The activity “Delete Purchase Order Item” that happened in 210 cases is considered as an undesired activity.

2-) 3-way matching, invoice before goods receipt. *“Purchase Items that do require a goods receipt message, while they do not require GR-based invoicing (indicated by the GR-based IV flag set to false and the Goods Receipt flags set to true). For such purchase items, invoices can be entered before the goods are a receipt, but they are blocked until goods are received. This unblocking can be done by a user, or by a batch process at regular intervals. Invoices should only be cleared if goods are received and the value matches with the invoice and the value at creation of the item.”*

We filter the process on the cases with “EKPO.GR_VERIF” value of “0” and EKPO.GR value of “1”, which indicates cases without “GR_based Inv.Verif.” flag but includes “Goods Receipt”. This filtering results 236k cases out of 252k and the process is discovered as shown in Figure 5.

According to Figure 5, the first nine variants of this process model cover 139k cases (referring to 59% of related cases), which shows that there are plenty of inefficiencies within the process execution. Like the process model in Figure 4, the model includes extra activities that are not violations such as “Receive Order Confirmation” (8k cases) and “Remove Payment Block” (27k cases), which are necessary to be done in order to clear the invoice. However, the activity “Delete Purchase Order Item” is considered as an undesired activity and happens in 5k cases.

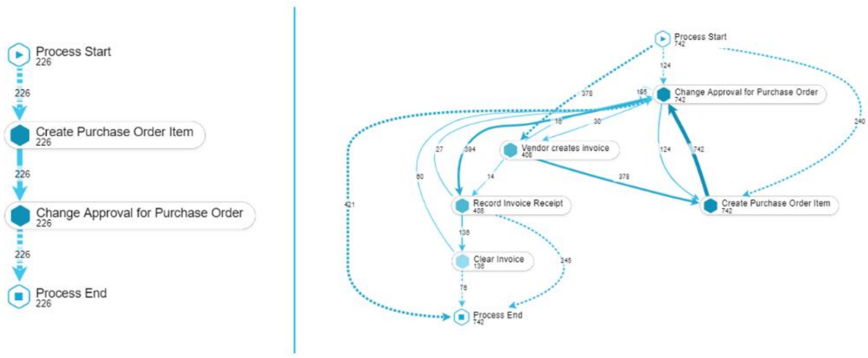
Fig. 5. Process models (without GR_VERIF, with GR): most frequent variant (left); nine most frequent variants (right)



3-) 2-way matching (no goods receipt needed). *“For these items, the value of the invoice should match the value at creation (in full or partially until PO value is consumed), but there is no separate goods receipt message required (indicated by both the GR-based flag and the Goods Receipt flags set to false).”*

We filter the process on the cases with both “EKPO.GR_VERIF” and “EKPO.GR” values of “0” that indicates cases without “GR_based Inv.Verif.” flag and “Goods Receipt”. This filtering results in 1.04k cases out of 252k and the process is discovered as shown in Figure 6.

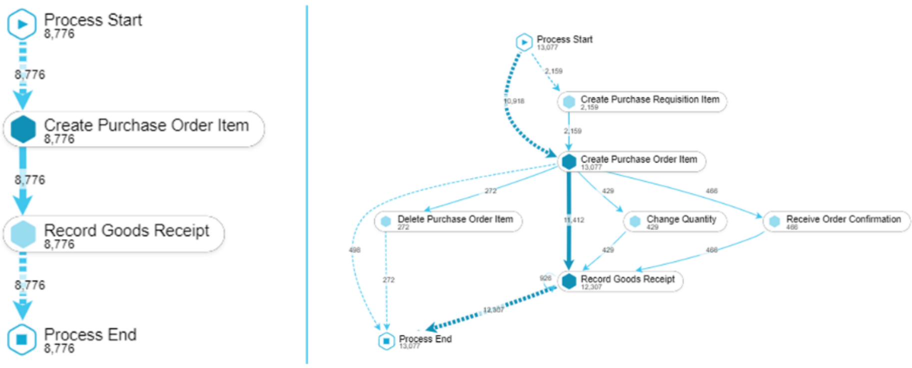
Fig. 6. Process models (without GR_VERIF & GR): most frequent variant (left); nine most frequent variants (right)



4-) Consignment. *"For these items, there are no invoices on PO level as this is handled fully in a separate process. Here we see GR indicator is set to true but the GR IV flag is set to false and also we know by item type (consignment) that we do not expect an invoice against this item."*

We filter the process on the cases with "EKPO.GR_VERIF" value of "0" and "EKPO.GR" value of "1" that indicates cases without "GR_based Inv.Verif." flag but includes "Goods Receipt". Additionally, we filter the process on the material class named "Consignment". This filtering results 14.5k cases out of 252k and the process is discovered as shown in Figure 7.

Fig. 7. Process model (consignment material class, (without GR_VERIF, with GR): most frequent variant (left); nine most frequent variants (right)

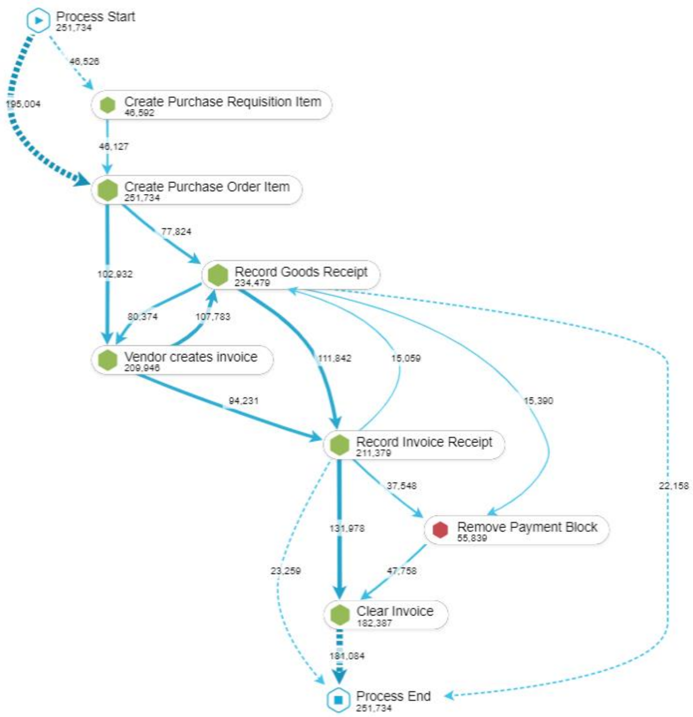


According to Figure 7, the first nine variants of this process model cover 13k cases (referring to 90% of related cases). This process model also includes an extra accepted activity of receiving order confirmation as well as two undesired activities including deleting purchase order item (269 cases) and changing the quantity (431 cases).

Considering the discovered process models, we design a model that describes dominant process instances. This process model is considered as an ideal process model to be used in conformance checking as well in order to detect process violations. The discovered process model is presented in Figure 8.

The process starts with either creating a purchase requisition or creating a purchase order item. Three activities including "Record Goods Receipt", "Vendor Creates Invoice", and "Record Invoice Receipt" can be performed in three sequences. The purchasing organization might either first receive the goods and then the invoice that is created by the vendor or the vendor first creates invoice and then sends the goods to the organization. Additionally, the process can follow to the end point after receiving goods in case of dealing with consignment material group. The activity "Remove Payment Block" is not a desired activity in an ideal process model, however it is a necessary step for clearing the invoice if a block has been set for the conformance reasoning laid out by the "3-way match, invoice before goods receipt" process flow.

Fig. 8. Process model summary of four scenarios



Question 2:

Question. "What is the throughput of the invoicing process, i.e. the time between goods receipt, invoice receipt and payment (clear invoice)? To answer this, a technique is sought to match these events within a line item, i.e. if there are multiple goods receipt messages and multiple invoices within a line item, how are they related and which belong together?"

Our understanding. The second question is concerned with throughput time, mainly of the second part of the process, which relates to invoicing. For instance, the throughput time of a sub-process that starts with the activity goods receipt ("Record Goods Receipt") and ends with the payment activity ("Clear invoice"). Accordingly, we conduct two groups of analyses that are directly referring to this question: "Cycle time" and "Advanced invoice matching".

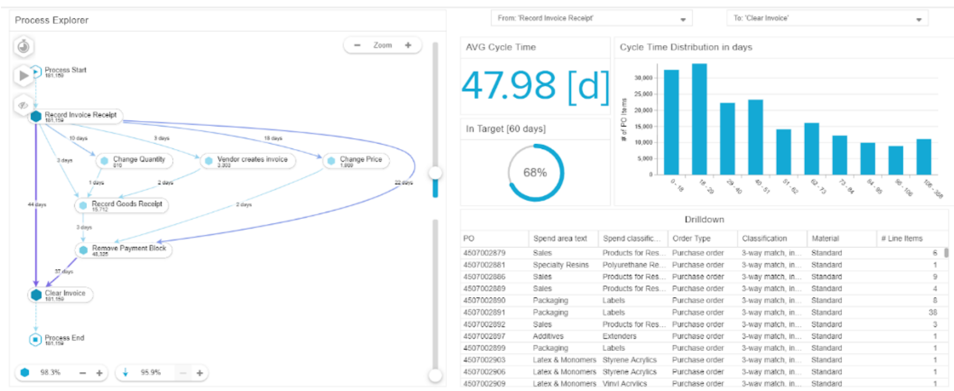
Invoice processing is a sub-process of the P2P process. Considering the invoicing process, the cycle time analysis provides critical information on the efficiency of incoming invoice management. With the help of process mining, these cycle times can be monitored and analyzed. Bottlenecks and inefficiencies can be recognized easily and quickly. Provided with these insights, accounting departments can optimize their accounts payable and thus realize faster processes and reduced turnaround times. This analysis focuses on the time between the entering of incoming goods and the payment of their related invoices. This segment is particularly important as, e.g., paying dunning charges for late payments can lead to a considerable decrease in the liquidity. Therefore, the analysis of the cycle times between the "Record Goods Receipt" and the invoice payment, as well as the avoidance of so-called "long-runners", are essential.

Accordingly, we initially transform the data into a usable format through the Celonis Event Collection. We then start by deleting "Clear Invoices" that are not clearings as the corresponding invoices have been cancelled. Having a look at the process, we noticed that in most cases the clearing took part almost immediately after the invoice was cancelled. As we do not have sufficient information, we decide that clearing invoice right after the invoice cancellation would be deleted, as either the case contained more than one or there were legitimate clearings in it. This is achieved by self-joining the activity table and looking at the time differences between the stated activities per case. Having removed the false clearings, we need to change the activity name of receipts.

In order to calculate cycle times, we exclude goods and invoice receipts that are cancelled later in order to avoid mismatches. As mentioned previously, due to available information, we assume that the cancellation of received goods always relate to the most recent goods reception. Therefore, using a window function, we are able to determine which receipts are cancelled by looking at preceding receipts and ordering them. Afterwards, we rename the first activity of the list. This is repeated until the number of cancellations is identical to the number of receipts to be cancelled.

After having all the prerequisites, we create a formula that is used to calculate the cycle times. To achieve this, we utilize the Celonis built-in Process Query Language (PQL) functions, namely the Pull-Up-Functions. This function allows us to aggregate a column based on another table. We define the parent table to which the child-tables entries are pulled, and explicitly define on which basis calculations are executed. Using the activity table as both parent and child table, we are then able to pull up the event times for each activity ordered by their occurrence in time, to determine the time that is passed.

Fig. 9. Cycle time analysis



As output of the described steps, we can build an analysis to tackle the questions asked by the process owner. It has been mentioned in the management summary that the average throughput time from “Record Goods Receipt” to “Record Invoice Receipt” is 19.9 days; the median throughput time is 9.8 days. Also, the average duration between the creation of the invoice in the source system (Record Invoice Receipt) and the clearing of the invoice is 47.9 days; the median is 41.9 days.
Taking a payment target of 60 days into account, 68% of cases are paid within this timeframe (from “Vendor creates invoice” to “Clear invoice”). Next to the possibility of supplier dissatisfaction and risk of facing non-favorable supplier conditions, the late fee rate of 3% is the industry standard.

Invoices, for which the payment needs 30 months or longer (90 days or more), can be specified as long runners. These invoices account for 12% (or 48,960) of all invoices and take on average 107.8 days to be completed. Especially the scenario of “3-way match, invoice before goods receipt” is a main root cause for these long runners, which is accounting for 99%. Additionally, another root cause for long runners is the material group “Labels”, 85% of invoices are due to this spending, while in total they account for 23%, which falls under the spend area “Packaging”.

Question 3:

Question. “Finally, which Purchase Documents stand out from the log? Where are deviations from the processes discovered in (1) and how severe are these deviations? Deviations may be according to the prescribed high-level process flow, but also with respect to the values of the invoices. Which customers produce a lot of rework as invoices are wrong, etc.?”

Our understanding. Using the Celonis Conformance Checking feature [7] along with FTR analysis allows us to address the third question that is mainly concerned with business process deviations.

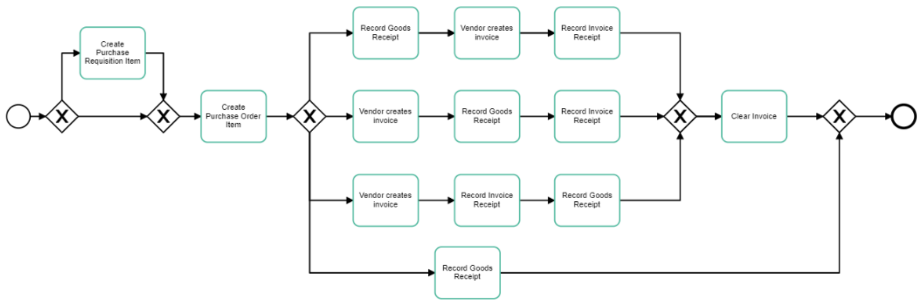
Conformance checking analysis is very useful to find deviations from the discovered model and point to the impact of each deviation. This analysis satisfies the high-level process flow as well as the four main process scenarios. The conformance checking analysis allows us to automatically check a reference process model (modeled by the organization or captured from standard frameworks) against the discovered process model from the extracted data. This analysis provides various information by comparing the two models. Some of the important findings can be addressed as follows:

- Statistics on conforming and non-conforming cases according to the reference model,
- The number of violations from the reference model and violation root causes
- The influence of violations on predefined KPIs (e.g. throughput time)
- Conforming cases trend over time

Conformance checking is a promising analysis for finding the inconsistencies between expected process model and the executed event logs. This analysis provides good insights on what can be changed within the process in order to fit the expected process model [5].

The expected process is modeled using BPMN language (Figure 10) that represents the ideal process. We use it as a reference model for conformance checking and detecting deviations.

Fig. 10. The expected process model



After modeling the expected process model and launching the conformance checking analysis, we must carefully check the list of violations and whitelist the possible incorrectly detected violations. This refers to activities that were considered as undesired steps but are instead accepted, such as “Receive Order Confirmation” and “Record Service Entry” or potential accepted sequences of activities such as “Vendor creates Invoices” followed by “Record Goods Receipt”.

Aligning the event data and the reference model results in 155k conforming cases against 96.7k non-conforming cases and 56 violations can be detected in the process model. In the ideal process, we expect the invoices to go smoothly from “Record Invoice Receipt” to “Clear Invoice”. However, there are issues appearing for invoices and in many cases, they have been blocked. Setting unnecessary blocks, e.g., due to discrepancies within the invoice positions, disturbs the process. This causes longer throughput times which can result in payments after cash discount due date or after the overall due date. In addition, unnecessary payment blocks increase the number of manual rework activities which is also associated with higher costs.

An overall investigation of violations for the general discovered process model using Celonis Process AI feature shows that root causes are mainly related to specific dimensions such as “Item Number of Purchasing Documents (EBELP)”, “User_Name”, “Vendor (LIFNR)”, “Purchasing document type (BSART)”, and “Material group (MATKL)”. After conducting the conformance checking analysis, the top-ranked violations are reported as several undesired activities including “Remove Payment Block”, “Change Quantity”, “Change Price”, “Cancel Invoice Receipt”, and “Delete Purchase Order Item”.

Using Celonis Process AI, we further discover that in the dominant violations, the main purchasing document type (BSART) is “standard PO” and the assigned material group (MATKL) is mainly from three categories including “Standard”, “Subcontracting”, and “Consignment”.

Out of all violations, “Remove Payment Block” as an undesired activity is a dominant violation with 22% of cases (58,839 cases). This activity affects the throughput time for 26 days longer. Accordingly, we investigate the root causes and find that most of these violated cases are assigned to the “Item Number of Purchasing Documents” 30 and 40 with 4k violations. Table 4 summarizes the conformance checking results and three related root-causes including “Item Number of Purchasing Documents (EBELP)”, “USER_NAME”, and “Account Number of Vendor or Creditor (LIFNR)”.

Table 4. Overview of the process violations

Violations	#Cases	Potential root causes
‘Remove Payment Block’ is an undesired activity	22%	• EBELP = 30, 40 - 4K violations • USER_NAME = user_015, user_006, user_023, batch_02 - 20K violations • LIFNR = vendorID_0136, vendorID_0104 - 7K violations
‘Change Quantity’ is an undesired activity	7%	• EBELP = 10, 20, 30 - 6K violations • USER_NAME = user_084 - 2K violations • LIFNR = vendorID_0136, vendorID_0197 - 2K violations
‘Change Price’ is an undesired activity	4%	• EBELP = 10, 20 - 4K violations • USER_NAME = user_071, user_037, user_038, user_039 - 3K violations • LIFNR = vendorID_0236, vendorID_0197 - 1K violations
‘Delete Purchase Order Item’ is an undesired activity	4%	• EBELP = 10, 40 - 2K violations • USER_NAME = user_252, user_060, user_158 - 825 violations • LIFNR = vendorID_0104, vendorID_0106 - 1K violations
‘Cancel Invoice Receipt’ is an undesired activity	3%	• EBELP = 10 - 2K violations • USER_NAME = user_015, user_004, user_013, batch_01 - 2K violations • LIFNR = vendorID_0118 - 389 violations

Additionally, we apply the conformance checker for the four discovered process models to be compared with the expected process.

First, taking the process model of ***“3-way matching, invoice after goods receipt”*** into consideration results in 9.13K conforming cases against 6.05K non-conforming cases. In addition to “Remove Payment Block”, “Change Quantity”, “Change Price” as undesired activities, another violation is referring to cases that “Clear Invoice” is followed by “Record Goods Receipt”. The later sequence might happen if the invoice is being cleared partially after receiving a good partially. However, the lack of information in the data does not allow us to separate partial payments from the rest.

Second, taking the process model of ***“3-way matching, invoice before goods receipt”*** into consideration results in 147k conforming cases against 88.6k non-conforming cases. Along with the main introduced violations, 2% of the non-conforming cases in this process model are violated by “Vendor creates debit memo” as an undesired activity. Additionally, the process owner might consider that "Remove Payment Block" is not necessarily an undesired activity in all scenarios. For instance, in this process flow, a payment block is created at the time of invoice receipt, to ensure that the goods are not paid for until they arrive. This exception is considered in the FTR analysis.

Third, taking the process model of ***"2-way matching (no goods receipt needed)"*** into consideration results in 1.01k conforming cases against 37 non-conforming cases. Along with the main introduced violations, "Set Payment Block" is also as an undesired activity. Decreasing the occurrence of setting payment blocks will cut costs and increase the speed of the process.

Fourth, taking the process model of ***"Consignment"*** into consideration results in 149k conforming cases against 86.7k non-conforming cases. Like in the process model of "3-way matching, invoice before goods receipt", "Vendor creates debit memo" is an undesired activity along with the main violations.

Along with conformance checking, we use FTR analysis in order to drill down the process and get insights about invoices and customers. FTR combines rework (activities that should not happen in the process even once, e.g. "Removing payment blocks" or "Change quantity") and recurrent activities (i.e. those activities that occur more often than they should, e.g. more occurrences of "Record Invoice Receipt" than of "Record Good Receipt".

Having assessed the most common rework activities and then the data flow conformance, this section starts by filtering on non-conforming data flows and then gains insights from there. Most of these non-conformances come from the spend area of logistics (514 from 532 cases, 96.6%), of which 444 cases were in the sub-spend areas of "Road Packed" (86%), 50 cases in "Other Logistical Services" (9.7%) and 19 cases in "Sea" (3.9%). The most frequent vendors involved are: vendorID_0538 (77 cases, 14.5%), vendorID_0233 (76 cases, 14.3%), and vendorID_0535, vendorID_0540 & vendorID_0541 (48 cases each, 9% each). The most frequent users, i.e. those that appear in the most number of cases, are: user_002 (458 cases, 86%), user_200 (357 cases, 67%), user_001 (140 cases, 26%), user_012 (128 cases, 24%), user_013 (114 cases, 21%).

The main root cause of this non-conformance is that the goods receipt value does not match that of the PO creation value or invoice receipt value, with all non-conforming cases resulting from this, across both FTR and not-FTR cases. Interestingly, the value from the activity "Vendor creates invoice" is more often not correct, i.e. not equal to that of the clear invoice activity.

The FTR principle is a quality management concept that states that designing a process that minimizes defects is more cost-effective than one that includes defect detection and associated corrective rework efforts [6]. An important aspect of purchase efficiency is the avoidance of rework activities as these are often the root of additional manual effort, longer process throughput times and lost opportunity. In order to ensure a smooth and fast purchase order processing, it is of great interest to identify the ratio of purchase orders that were correctly processed the first time compared with those that were not, using this as a main KPI to improve over time. With the help of process mining, orders not considered as FTR can be spotted easily and quickly, and their root causes can be identified. Provided with these unique insights, businesses can optimize their P2P process to save time and money by effectively reducing causes of process inefficiencies.

In order to produce a meaningful analysis, the following filters and classifications are applied to the data set:

1. Not considering consignment cases, only cases that flowed through the "Clear Invoice" activity are considered in the FTR analysis. This reduces the potential of early-stage open cases to skew the FTR results, e.g. cases where only a Purchase Order has been created would be considered as FTR, which is a poor indicator of its performance as many other process activities are still yet not done.
2. Only cases created during the year of 2018 are considered in the analysis.
3. The following activities are added to the "Whitelist", i.e. a list of the activities that will not trigger a violation to FTR, should only occur once per case: "Clear Invoice", "Create Purchase Order Item", "Create Purchase Requisition Item", "Receive Order Confirmation", "Record Goods Receipt", "Record Invoice Receipt", "Record Service Entry Sheet", "Vendor creates invoice".

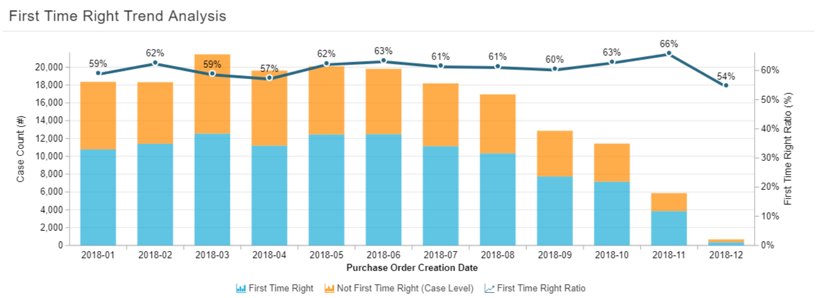
Note, there are some special cases where the whitelist is altered slightly, however these will be explicitly stated in the respective analysis section.

All activities not included in the whitelist are considered as rework activities, and thus if they occurred even once per case, the case is not considered as FTR. Additionally, if those activities in the whitelist occur more than once per case, the case is also not be considered as FTR. Furthermore, whilst classifying cases into either FTR or not, we count the amount of rework activities and recurrent activities per case, allowing the severity of an FTR violation to be set. This has two benefits: (i) It enables the analyst to identify irregularities that require filtering or data cleaning, and (ii) it ultimately enables the client to identify the most problematic activities in their process, thus allowing for data-driven transformation initiative prioritization. With FTR as defined above, and the respective predefined filters also set, the data yields 111,470 FTR cases from a total of 183,677 cleared cases, meaning the FTR ratio of 60.7%.

The FTR ratio is plotted over time, enabling a visual understanding of any trends that may be present over time. As becoming apparent in Figure 11, there is a relatively consistent ratio of FTR cases distributed throughout the year. An interesting observation is that even though there are few cleared cases towards the end of the year, 17,976 in the last quarter compared with 58,120 in the first quarter, their FTR ratio differs by only 3%, 63.2% compared with 59.8% respectively. This provides an indication that purchase order volume may be only slightly correlated with FTR.

When considering only those cases that are not-FTR (72,207), it is obvious that rework activities are much more significant to their classification than recurrent activities, with 94.7% of not-FTR cases including rework activities compared to only 22.0% of not-FTR cases that include recurrent activities. Due to the specific nature of the company's purchasing model however, these numbers could be misleading. As discussed in the briefing document, the data flows go beyond the four categories described, and involve multiple goods receipts, payment blocks and invoices per case. This adds complexity to the FTR analysis, as recurrent activities are actually permitted (e.g. multiple goods receipts and invoice receipts). Thus, in order to provide more meaningful insights, the analysis is categorized into the four types of flows discussed within the assignment brief, enabling more specific filters to be applied.

Fig. 11. First-Time-Right (FTR) trend analysis



First, for the process flow ***"3-way Matching, Invoice After Goods Receipt"*** to be conformant, the invoice should be received after goods receipt, and the value at PO creation should be equal to those at goods receipt and invoice receipt. When considering this process flow, 9,554 cases flow through the activity "Clear Invoice", of which 5,851 (61%) cases are classified as FTR and 9,022 (94%) cases are process flow conformant. When filtering on cases conforming to FTR, the data flow conformance drops below the average, to only 92%. Of the not-FTR cases, had a higher than average process flow conformance, with 3,643 conforming cases compared to 60 non-conforming (98.3%).

A total of 7,959 rework activities are present in these not-FTR cases, with "Change Price" as the most frequent with 1,106 activities (14%) across 972 cases (26% of all cases). Interestingly, when filtering for cases that include a price change, the conformance to the data flow jumps to 99.8%, indicating that this rework activity is performed to attain data flow conformance. This additional effort enables the business to be conformant in 970 of the 972 cases. One of the two cases, "4508071078_00010" is nonconformant as the invoice is cleared, however no goods are received, with an item value of 334 Euro. The other case, "4507015727_00010" is non-conformant as the "Goods Receipt" value (66,995 Euros) does not equal that of the PO Item or clear invoice value (33,497 Euros), and thus it should not have been cleared.

The second most frequent rework activity, "Change Quantity" is present in 689 cases, occurring a total of 790 times. Of the cases that have this activity, the data flow conformance ratio is 99.9%, with only one case non-conformant. The violation occurs in the same case as one of the price change violations, "4507015727_00010", with the same user (User_171) responsible. Interestingly, the value of the activity "Vendor creates invoice" is in fact the PO item value and the clear invoice value, which may indicate a user entry error, or a field automation error.

The third most frequent rework activity, "Cancel Invoice Receipt" occurs 649 times across 418 cases. The vendors whose invoice receipts are cancelled most often are: vendorID_0236 (78 cases & 80 occurrences), vendorID_0470 (43 cases & 43 occurrences), vendorID_0157 (26 cases & 41 occurrences), vendorID_0404 (23 cases & 27 occurrences) and vendorID_0183 (16 cases & 18 occurrences) - accounting for 45% of the cases. With this activity present, the data flow conformance rises to 98%, with 9 cases in violation. Case “4508046522_00001” appears to be the most serious violation, due to the cumulative net worth equaling over one million Euros at multiple stages in the process, although the PO item was valued at only 1,582 Euros at creation. Three invoices are recorded during this case, with only one of them being cancelled, and only one being cleared after the removal of a payment block. Case “4507007756_00001” is cleared even though no goods receipt was recorded, the goods receipt values for case “4507010781_00001” differed from the PO and invoice clearing value, case “4508046183_00001” begins with the invoice creation and looped multiple times between goods receipt and record service entry activities, and the goods receipt value differs from the clear invoice value.

Second, the process flow ***“3-way Matching, Invoice Before Goods Receipt”*** corresponds to the vast majority of all cases within the data set, with a total of 173,536 flowing through the clear invoice activity (95% of all cleared cases) and a FTR ratio of 87%. For this data flow to be conformant, the value at PO item creation must match that of the value at clear invoice, and a goods receipt has to occur prior to clearing. The conformance ratio to this data flow is 99.69% (172,008 conforming cases). From all the cleared cases, there is a total of 6,293 variants from the happy path, whereby the FTR cases have 95 distinct variants, and the not-FTR cases have 6,198 variants.

For FTR cases, the process flow conformance ratio jumps to 99.96%, with only 53 non-conforming cases out of 143,200 cleared cases. These non-conforming cases mostly come from vendorID_0660 whom is responsible for 37 of the violations, of which 32 cases are cleared without a goods receipt and 5 are cleared prior to goods receipts. Surprisingly all values at PO creation matched the cleared invoice values, which indicates that a conformance check was performed on the values, but not on the process itself. All other non-conforming cases result from the same issue, whereby the invoice is cleared prior without goods receipt.

For the not-FTR cases, the process flow conformance ratio drops slightly to 98.10%, whereby 474 out of 24,972 cases are classified as non-conforming. In 472 of these cases, the invoice is cleared even though the goods are never received, and in the other three cases, the goods receipt happens after invoice clearing. Like the FTR cases, the values at PO creation match those at invoice clearing in every single case. The most frequent manual rework activities are: “Cancel Invoice Receipt” - occurring 467 times across 454 cases, “Vendor Creates Debit Memo” - occurring 459 times across 456 cases and “Cancel Goods Receipt” - occurring 275 times across the same number of cases. These rework activities are all caused by the vendor’s debit memo, which may stem from the vendor incorrectly invoiced the company in the past or over-delivering goods. The rectification of these issues takes place within a procurement sourcing process, even though it is of a different nature to the second process flow. As this use case differs from that of the desired data flow, it is recommended to the customer that they change their process, or incorporate a new field, that allows for a rectification to be identified. Further, as these rectifications involve manual rework to the company, it is important to note that the vendor with the most cases for these issues is vendorID_0246, responsible for 232 cases (49%). The spend areas most influential in these non-conformances are: “Trading & End Products” - 263 cases (55%), “Sales” - 107 cases (23%) and “Packaging” - 68 cases (14%).

From the 24,972 not-FTR cases, there are a total of 199,247 activities, of which 40,937 (19%) are manual rework activities. The most frequently occurring manual rework activities, by case count, percentile of all cases, occurrence count, occurrence ratio, the number of non-conforming cases that include the activity and the process flow conformance ratio of all cases that include the activity, are shown in Table 5.

Table 5. Rework activity metrics

Activity name	Case count	% of cases (Not-FTR)	Rework activity count	% of all rework activity	Non-conforming cases	Process flow conformance ratio
Change Quantity	11846	47.4%	14536	35.5%	19	99.84%
Change Price	7985	32.0%	8909	21.8%	92	98.85%
Vendor Creates Debit Memo	4680	18.7%	4801	11.7%	456	92.26%
Cancel Invoice Receipt	4243	17.0%	4544	11.1%	454	89.30%
Change Approval for Purchase Order	1595	6.4%	2472	6.0%	0	100%

The activities “Change Quantity”, “Change Price”, and “Change Approval for Purchase Order” all have a positive impact on the process flow conformance, whereby the cases including these rework activities are conformant above the average not-FTR value of 98.1%. The other activities are by nature non-conforming, and result from issues on the vendors side, such as using the wrong rates when calculating line values or sending an oversupply of stock.

The median throughput time between “Create PO item” and “Record Goods Receipt” of not-FTR cases for this process flow type is 16 days. Interestingly, the median time between “Create PO item” and “Change Quantity” is 12 days, and when this activity is present the median time between PO creation and goods receipt increases to 21 days. So for cases without the activity, goods arrive sooner. The fact that quantity changing occurs 12 days after PO creation (median could be interpreted as late changes of POs. A root cause might be that POs are created too far in advance for accurate planning, and thus required changes (increases or decreases) closer to the time of goods receipt. In these cases, the throughput time increases, which is harmful for purchasing planning. There could be some reasoning to create a PO as a type of “blocker” in the vendor’s system, and then change it with improved accuracy when it is closer to the date when the goods are actually needed. However, depending on the vendor lead time, it may also make sense to simply order items closer to when they are needed, which reduces additional effort and enables a more streamlined purchasing process.

Price changes have less impact on throughput times than quantity changes, however they have a significant impact on manual rework, and thus create risk for overpayments. The vendor “VendorID_0197” is involved in the most violations, accounting for 911 price changes across 766 cases. The process flow conformance ratio is 100%, indicating the integrity of the review process implemented in the company. With more data it would be possible to understand whether the price changes come from the vendor or the business, enabling root causes to be identified: such as master data issues, typos, unit price scaling errors and more.

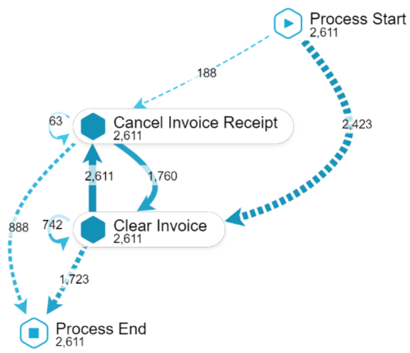
Third, the process flow ***"2-way matching (no goods receipt needed)"*** is considered conformant if the value of the invoice is equal to the value at creation. These types of process flows are small in numbers, only accounting for 297 cleared cases, of which all were considered conformant to the respective data flow. These cases contain the rework activity "Change Approval for Purchase Order". The spend area "Real Estate" is the most accountable, with 222 violations (75%) and 1,084 PO approval changes. The most expensive PO came from EBELN: 4507075967, with three-line items totaling to 451k euros. Due to the value of these cases, the additional rework effort is most certainly justified. Interestingly, the only two users that make the activity changes are User_602 and User_603.

Fourth, as the process flow ***"Consignment"*** does not consider invoices on an item level, the invoice activities (such as invoice receipt and clearings) are not present. So far, the FTR formula analyzes only cases where the clear invoice activity is present, which is not significant for this flow type. Thus, the end activity was adjusted to "Record Goods Receipt", meaning only cases with this activity are analyzed. With this logic applied, a total of 13,466 cases are closed, of which 12,028 (89%) are labelled as FTR. As this data flow does not have values assigned to PO items, further process flow conformance is not possible, similar as for the previous data flows.

5 Limitations

Limitations are in the essence of each project. Our analysis and findings are also limited to the provided data and information availability. We had complications that might refer to noisy data but also lack of information on the process steps that might lead to process misunderstanding. For instance, there are 2,611 cases that are canceling the invoice receipt after clearing the invoice and 1,760 cases that are clearing the invoice after canceling the invoice receipt (Figure 12).

Fig. 12. Process graph showing invoice canceling after clearing and clearing after canceling



Additionally, we can refer to data anonymization that limits comprehensive understanding of the data. However, we have tried to get the most knowledge possible out of the provided data and within the limited time. We hope that our findings would support the process owner in their process transformation initiatives.

Finally, process mining is a powerful tool to analyze and understand business processes. However, the findings generated should be interpreted cautiously (especially regarding direct actions taken on the basis of process mining insights), as correlation does not necessarily imply causation.

References

1. Van Der Aalst, W. (2011). Process mining: discovery, conformance and enhancement of business processes (Vol. 2). Heidelberg: Springer.
2. Pickell, D. (2018, September 4). Everything you need to know about big data analytics. Retrieved from https://learn.g2crowd.com/big-data-analytics
3. https://icpmconference.org/icpm-2019/contests-challenges/bpi-challenge-2019/
4. van Dongen, B.F., Dataset BPI Challenge 2019. 4TU.Centre for Research Data. https://doi.org/10.4121/uuid:d06aff4b-79f0-45e6-8ec8-e19730c248f1
5. Rozinat, A., & Van der Aalst, W. M. (2008). Conformance checking of processes based on monitoring real behavior. Information Systems, 33(1), 64-95.
6. Crobsy, P. B. (1984). Quality without tears: the art of hassle-free management. McGraw-Hill Book Company, New York, N.Y.
7. Veit, F. et al. (2017): The Proactive Insights Engine: Process Mining meets Machine Learning and Artificial Intelligence. In: Proceedings of the BPM Demo Track and BPM Dissertation Award co-located with 15th International Conference on Business Process Modeling, Barcelona, Spain.
8. Badakhshan, P. et al. (2019): The Action Engine - Turning Process Insights into Action. In: Proceedings of International Conference on Process Mining (ICPM) Demo Track, Aachen, Germany.

Compliance and Performance Analysis of Procurement Processes Using Process Mining
BPI Challenge 2019

Kiarash Diba, Simon Remy, and Luise Pufahl

Hasso Plattner Institute, University of Potsdam, Potsdam, Germany
{firstname.lastname}@hpi.de

Abstract. Procurement processes are essential part of the value chain of an organization to provide services and products. A regular analysis and optimization of these processes is beneficial for companies to increase quality, reduce costs, and identify process risks. Process mining techniques provide insights into the common flows of activities, detect deviations and compliance issues, and report process performance. In this paper, we present the analysis and evaluation of procurement processes of a company in the area of coatings and paints from the Netherlands. Applying filtering and discovering techniques, the paper presents six process models for handling different types of purchase orders. In addition, deviations of these flows and compliance violations, their frequencies, and root causes are described. Finally, key performance indicators regarding throughput time are defined and measured.

Keywords: BPI challenge · Procurement · Process Mining · Compliance.‘

1 Introduction

Business Process Intelligence (BPI) Challenges, organized annually, aim at advancing the field of process mining through providing real-life data sets and problems. By involving the process mining community from both academia and industry, the BPI challenges showcase the power and value of process mining, and trigger the development of new techniques and novel solutions. This year the data is provided by a multinational company located in the Netherlands in the area of coatings and paints. The data refers to the *Purchase-To-Pay* process as part of the procurement process of the company. Procurement processes are crucial to the value chain of organizations. These processes can be subject to business risks, such as the risk of long delivery times, decreasing efficiency of production or increasing costs, or more severely, risk of potential fraud. Analyzing these processes regularly provides insight on improvement potentials and prevention of potential risks.

Process Mining as the main component of process intelligence provides techniques for the analysis of processes based on recorded data, such as the provided event log for the BPI challenge. Real process flows can be automatically discovered, revealing the behavior of the recorded process execution. Conformance of the recorded behavior to the expected behavior can be checked, and deviating cases can be detected, exposing violations to compliance rules. Root causes of violating cases can be explored. Performance metrics can be measured, unusually long cases be detected, and bottlenecks identified. These can lead to identification of process diagnostics, insights for process improvement, and preventive action for potential risks and fraud.

This report outlines the details and results of a process analysis on the provided data for the *Purchase-To-Pay* process, focusing on the three questions posed by the process owners. These questions are related to the three aspects of process mining, namely process discovery, conformance (compliance checking), and performance analysis. In order to organize the analysis flow and take a step towards a standardized and repeatable process mining analysis, we have followed a methodology inspired by works in the literature [5, 4, 2]. In the context of this BPI challenge, our methodology consists of five phases of *data and process understanding*, *discovery and design*, *Conformance and compliance checking*, *Performance analysis*, and *report*.

Section 2 details the understanding phase and provides statistics and models necessary for understanding the process, in addition to identification of data quality issues existing in the data. Section 3 zooms into specific analysis focusing on process discovery and design, before section 4 outlines deviations and compliance issues, while Section 5 details performance analysis of the process. Finally, section 6 summarize the main results, and suggestions for process improvement, and concludes the report.

2 Data and Process Understanding

In this part, we report the result of exploration and inspection of the data and the underlying processes.

2.1 The Data

The data is recorded for the execution of *Purchase-To-Pay* processes (possibly of an ERP system) of purchase orders submitted in 2018. The log contains 1,595,923 events, belonging to 251,734 traces (cases). The case notion adopted for the log is individual purchase order item. Each purchase order item is part of a purchase order (PO) document. Each PO document can consist of multiple items. There are 76,349 PO documents in total.

The data contains several attributes. A number of attributes are on the event level including:

- `Case ID`: A combination of PO ID and item ID as the case identifier,
- `Activity`: The name of the activity that the events refer to,
- `Timestamp`: of activity completion, and

– `User`: the resources recording the activity.

Majority of attributes, however, are on the case level recorded for each item including:

– `Company`: The anonymized ID of the respective subsidiary that the case relates to,
– `Name`: Anonymized name of the vendor,
– `Spend area text`: The purchasing area,
– `Cumulative net worth`: The value of the item in Euro,
– `Document type`: The type of purchasing document,
– `Item category`: The invoicing procedure is determined based on this attribute, and
– `Item type`: the type of the item.

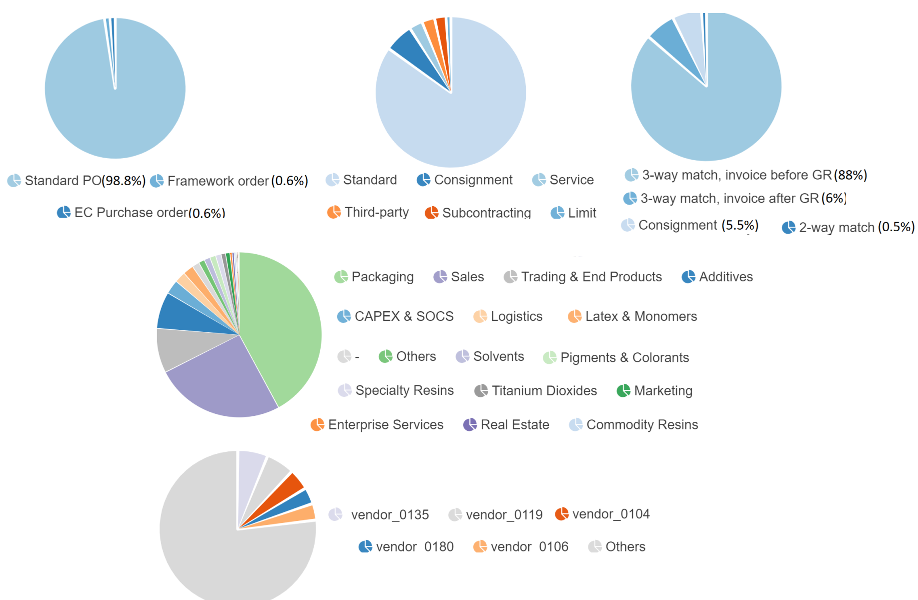


Fig. 1. Proportion of values of five case attributes in the log

Statistics on attributes: Figure 1 displays the proportion of the different values for a number of attributes. Majority of PO documents are of type "standard PO", while "framework", and "EC purchase orders" are together around 1% of the documents. The most frequent `item type` is "standard" followed by "consignment" and "Service". `Item category` is dominated by "3-way match, invoice before Goods Receipt (GR)" (88%), whereas the "2-way match" constitute a small proportion (0.5%). The biggest proportion of items are in the area of "packaging". The other frequent areas are "sales", "trading & end products", and "additives". There are around 1900 unique vendors recorded in the log and the last chart in figure 1 reveals the top 5 frequent vendors (vendor_0135, vendor_0119, vendor_0104, vendor_0180, vendor_01060).

There are 42 different `activities` in the log. The five most frequent activities are: "Record goods receipt", "Create purchase order item", "Record invoice receipt", "Vendor creates invoice", and "Clear invoice". For each case, there is exactly one instance of "Create purchase order item", while the other activities can happen an arbitrary number of times for each case (e.g. several goods receipt might exist for one item). In total there are 627 different `Users` recorded, divided into batch and human users. Data related to 4 of the subsidiaries are recorded in the log. However, two of which ("Company-0001", and "Company-0002") only contain 2 cases each which are eventually deleted, one paid before deletion.

	Standard PO			Framework order	EC purchase order	$\sum$	%
	Company_0000	Company_0001	Company_0002	Company_0003	Company_0000		
3-way match after GR	14,077	2	0	494	611	15,182	6%
3-way match before GR	221,008	0	2	0	829	221,010	88%
2-way matching	0	0	0	1,044	0	1,044	0.5%
Consignment	14,498	0	0	0	0	14,498	5.5%
$\sum$	249,583	2	2	1,538	2,484	251,734	251,734

Table 1. Item category distribution across all companies and document types (created in Celonis).

Table 1 provides an overview to the number of cases belonging to each item category, and document type in their respective subsidiaries. As shown in the table, "3-way match, invoice after goods receipt" items comprise around 6% of all cases originating from "Company-0000" and from all three document types but mostly "standard PO". "3-way match, invoice before goods receipt" compose the biggest part of the log at around 88% all from "Company-0000", mostly "standard PO" with a few EC purchase order and 1 exceptional, incomplete case recorded as framework order. "consignment" order items constitute around 5.5% of the total number of items and are all originated from "Company-0000" and are as document type of "standard PO". Around 0.5% of cases are of category "2-way match", all coming from "Company-0003" and from "framework order" purchase documents.

Relation between attributes: The classification trees in the figure 2 reveal relations between attributes `item category`, `document type`, `item type`, and`spend area text`. "EC purchase order" documents can contain items from categories, "3-way match, invoice before GR", and "3-way match, invoice after

GR". "Framework orders" are either "3-way match, invoice after GR" or "2-way match". "standard PO" contain, all three except "2-way match". On the other level, "3-way match, invoice before GR" items are of item types "standard", "subcontracting", or "third party", while "3-way match, invoice after GR" are mostly "services". "2-way match" are type "limit" , and "consignment" obviously "consignment" types. Furthermore, "Invoice before GR" are mostly (around 65 %) production (direct procurement) materials such as "packaging", and "trading and end products". "3-way match invoice after GR" on the other hand are mostly (88%) non production (indirect procurement) such as "logistics". "2-way match" items are almost always indirect procurement, mostly in the purchase area of "real state" and "energy", and "consignment" almost always(98.64%) direct procurement mostly "titanium dioxides", and "latex & monomers.

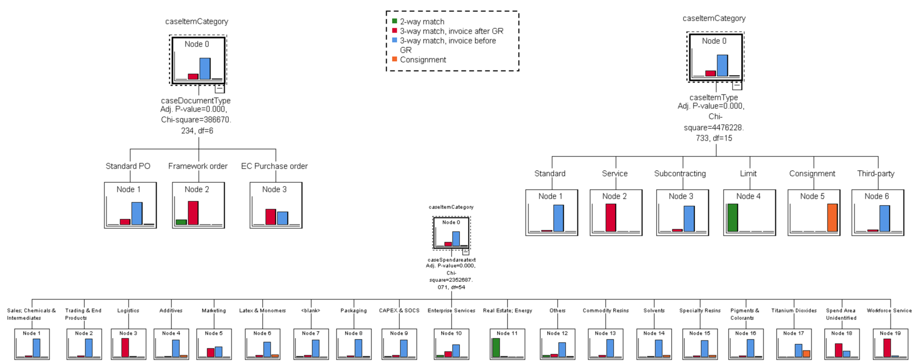


Fig. 2. relation between item category with item type, document type, and spend area

2.2 The process

The process is the *Purchase-To-Pay* process of a company with more than 60 subsidiaries. Since the case notion is purchase order items, the process describes the flow that items go through, since creation to payment. The purchase order items are divided into four categories (identified by `Item Category` attribute) based on the procedure for matching the invoice before payment. Therefore, items from each category go through a different flow. The categories are as follows:

- *3-way match, invoice after goods receipt* These items follow the classical invoicing approach when the value of an invoice is matched against the value of the goods receipt message and the value at its creation.
- *3-way match, invoice before goods receipt* The same matching procedure is applied with an only difference that invoices can be recorded before the goods are received but they must be blocked until the goods have been received. Only then the invoice can be paid after removing the payment block.

- *2-way matching (no goods receipt needed)* For these items, the goods receipt is not required. Before payment, the value of the invoice(s) is matched with regard to value at creation. These items are part of framework order documents (so called "blanket PO) which refer to items with recurring invoices over a period of time.
- *Consignment* for these items, the product is received and stored in the warehouse, but not paid until only after it has been used for production. Due to the fact that invoices are handled in a different process, the data do not contain information about their invoices.

Beside the essential activities contained in the normal flows, four other types of activity can be performed in *Purchase-To-Pay* processes which affect the performance of the process.

- Changes: change activities such as "change price" and "change quantity" can cause rework, decrease efficiency and therefore, can be candidates for process improvement.
- Cancellation and deletion: These activities often imply waste of time and resources.
- Release workflows: Activities such as "release PO" and "reactivate PO" can also lengthen the throughput times and decrease efficiency.
- Messages: Events such as "order confirmation" and overdue notices such as "vendor creates debit memo" can become potential bottlenecks of the process.

2.3 Data quality issues

There are a number of data quality issues in the event log which could influence the analysis result.

- *Incorrect timestamps*: there are a few cases with events dating back to outside of the scope of the event log (e.g 1948), or to a future date. As most of this cases show a regular behavior in terms of process flow and were even complete cases, we kept them for the discovery and conformance checking, while neglecting them in the performance analysis which results would be biased with the incorrect timestamps (e.g. unusually long duration would impact the average).
- *Incomplete cases*: Incomplete case, not ending with an expected end event (e.g. "clear invoice"), affect the result of analysis. The reasons for incompleteness might vary and is up to domain expert to identify them, however, for some cases is simply the snapshot effect, i.e. the case was still open at the time of data extraction and will possibly end in the future.
- *Multiplied events*: Another data quality issue is the fact that some events happen at PO document level but are copied in the traces of each item. Since this event is copied with the same timestamp into each trace, while items have different creation times and their events have different timestamps, these events are positioned wrongly in the traces. (For example if

PO document x is created at 2 pm with two items, y created at 2pm and z at 4pm (it is specially the case for framework orders). now imagine a change approval activity is performed for PO x at 3 pm. While the change approval event would appear after creation of item y, indicating a sequence from create purchase order item to change approval, the creation of item z would be positioned after the change approval, indicating the sequence from change approval to create purchase order item. This will impact the result of process discovery negatively.

- *Missing events*: In cases that goods are receipt before invoices, a payment block has to be set to prevent early payments without having the goods. However, these event has not been properly recorded and reflected in the log. Although the event "Remove payment block" exist 57,136 times, "Set payment block" is recorded 124 times.
- *Missing attributes*: Another rather minor problem is that the value for a number of attributes are missing for some cases. For example, in the "spend area text" or "Users" sometimes the value is not recorded.
- *Case data on event level*: Lastly, the fact that majority of attributes are at case level but copied for each event not only creates overload, but leads to incorrect statistics (e.g. frequencies of certain type of items). To avoid this, for collecting statistics on this attributes we grouped the events and considered on event per each case.

3 Process Discovery of the Common Flow

Process discovery is one of the three main types of process mining and aims at constructing process models based on event logs [1]. It can be seen in various perspectives like the control-flow, the organizational, and case perspective. In this section, we focus on discovery of control-flow. The discovered process models should capture the behavior recorded in the event log. However, there is a need for a trade-off between desired and undesired behavior. The latter, also called noise, may occur in several ways, such as missing data, perturbed order of events, or additional events. Especially for processes with high variance, as observed in this year's log, this trade-off between noise and desirable behavior can be challenging. In general, the discovered model should represent the recorded events, therefore it should be precise. At the same time it should fit the log.

In this challenge we roughly followed the steps below, to handle infrequent behavior but retain most of the observed events:

1. Filter out artificial start and end events,
2. filter out incomplete cases,
3. filter out infrequent cases[1],
4. use inductive miner[2] and heuristic miner[3] with appropriate thresholds,

[1] Eric Verbeek, Filter Out Low-Occurrence Traces (Single Log), ProM Plugin
[2] S. J. J. Leemans, Mine Petri net with Inductive Miner, ProM Plugin
[3] F. Mannhardt, Ineractive Data-aware Heuristic Miner, ProM Plugin

5. create and check activity to resource mapping,
6. Design a BPMN process diagram based on the discovered model.

To accomplish the steps above, we used different process mining tools like Disco, Celonis, ProcessGold, ProM, and Apromore. In this section, especially Celonis and ProcessGold were used to obtain more in-depth insights into the data and to understand root causes for certain control-flow deviations. In the following the result of process discovery for each `item category` is presented.

3.1 3-way Match, Invoice after Goods Receipt

This section describes the models which depict the process of items falling into the first category. For such items, the value of the goods receipt message should be matched against the value of a corresponding invoice received message and the value put during the creation of the item. However, the invoice should be received after the goods have been received. As stated in Table 1, 15,182 cases belong to this group. After exploring the event log, we decided to distinguish between EC purchase orders and all others. Since the former ones are partly processed by a supplier relationship management (SRM) system, we separate these process instances.

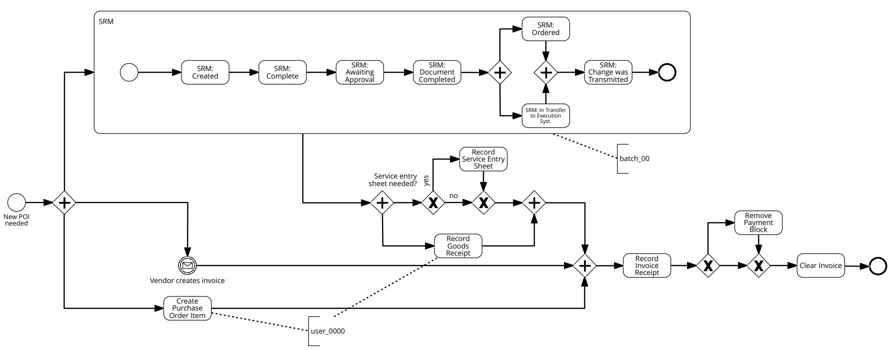


Fig. 3. Designed model for item category *3-way match, invoice after GR*, and document type EC purchase order.

EC purchase order Initially, this sublog contains 610 cases with 279 variants. After filtering out incomplete instances, we filtered out very infrequent variants. In the next step, we applied the inductive miner discovery algorithm to obtain a process model. We improved the resulting model by grouping activities into a subprocess and adding resource information. Figure 3 depicts the discovered process as BPMN process diagram. The discovered model has a trace fitness of 84% on the original log.

A batch resource executes all activities in the subprocess across all process instances. Once the subprocess ends, two activities are concurrently executed. On one side, the receipt of the goods gets recorded while on the other side a service entry sheet gets registered. The later one is optional and only observed for service-oriented purchase order items which holds for 47% of all variants. Since the model reflects the common flow, it does not contain infrequent loops. However, in 11% of all cases, a vendor creates invoice at least three times. In at least 24% of all cases in which this activity is executed at least three times, the vendor_0040 creates the invoice. In addition, 50% of these cases belong to the "Workforce Service". We observed similar findings in section 4 with violations of business rules.

While several different resources execute most of the activities which do not belong to the subprocess, only one user creates the order items and also registers the receipt of the goods. Further, many different resources execute most of the user tasks, therefore we did not model the resource allocation with pools and lanes but added comments to highlight activities which are connected to specific resources.

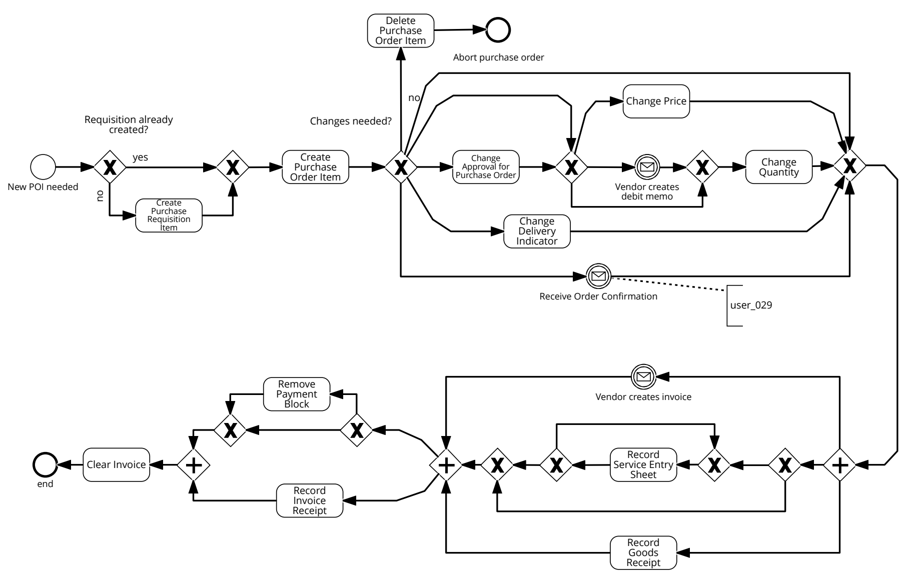


Fig. 4. Designed model for purchase orders of item category *3-way match, invoice after GR*, excluding EC purchase order documents.

Standard and Framework Orders While the previous process model describes only EC purchase orders, the model in Figure 4 depicts all other document types, as there are framework orders and standard purchase orders. Similar to the former process, we observe a high variance of cases. Nevertheless, the discovered model has a trace fitness of 83% on the original log.

While the activity "Create Purchase Requisition Item" is optional in the model, we observe a process drift, which is indicated by a more frequent execution of this activity since September 2018. Like the previous process model, this model only reflects the common behavior and only one loop. However, there are activities in the log which are executed multiple times, even if the model does not allow for this. One example is the "Change Price" activity. In 68 cases it happens at least two times and in 79% of cases in the Logistics area. For example, vendor_0388 participates most in cases where the price has changed multiple times. "Change Quantity" is another activity that occurs at least two times in 115 cases, even if the model does not allow for that. Other than in the previous example, the activity is repeated the most in "Packaging" (61%), especially by vendor_0264 (24%). Again, we observe similar behavior concerning violation of cardinality rules as described in section 4. Based on domain knowledge the process owner should decide whether this exceptional behavior should be reflected in the model in the future or not. Maybe it would also make sense to keep cases of certain spend areas in separate process models, like logistics.

Another activity which stands out is the "Remove Payment Block" activity. It is present in 15% of all cases. Following the description of the item category, this activity does not belong to this process. It mostly occurs in "Packaging" and "Sales".

3.2 3-way Match, Invoice before Goods Receipt

This category of items go through the same matching procedure as the previous category (the value of invoice, goods receipt and at PO creation time). However, the invoice can be received and recorded before the goods are received. In this case the invoice should be blocked for payment until the goods are received. This category of items are the most common category in the log constituting around 88% of cases and contain items from standard, and EC purchase documents. In the same manner as the previous category we have divided the process based on the PO type. In the following we detail the process for each part.

Standard PO These cases compose 87% of cases in the log. The model for this category is shown in figure 5. A purchase requisition item is created prior to the creation of PO item; or create purchase order item is performed first. Next an order confirmation can be received (based on the log this is optional) in case vendor has not sent the invoice yet this event might happen next, before goods and invoice are receipt (in any order). Finally, the payment block needs to be removed (if exist) before the invoice can be paid. The designed model based on the recorded behavior in the log, and adjusted by the logical (desired) process flow maintains the fitness of around 85 percent compared to the complete log and of 91 percent against the log with only completed cases. This implies that although there are various variants and deviations in the log (which will be investigated in detail in the next section), most cases can be explained by the model depicted in figure 5.

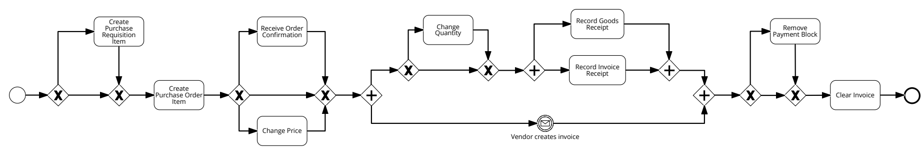


Fig. 5. BPMN process diagram for PO items following the *3-way match, Invoice before Good Receipt Standard Po*

EC purchase order There are 829 cases belonging in this category and the process model describing them is depicted in figure 6. The process start with the SRM system events (or by vendor sending a debit memo), followed by creation of purchase order item. Afterwards, receiving the goods can happen after the creation and recording the invoice. If the invoice is recorded before goods are received a payment block has to be set, which is not reflected in the log properly. After the goods are received and the invoice is recorded the payment block can be removed (if set previously), which, unlike setting the block, is reflected in the log (57,136 times) and then invoice can be cleared.

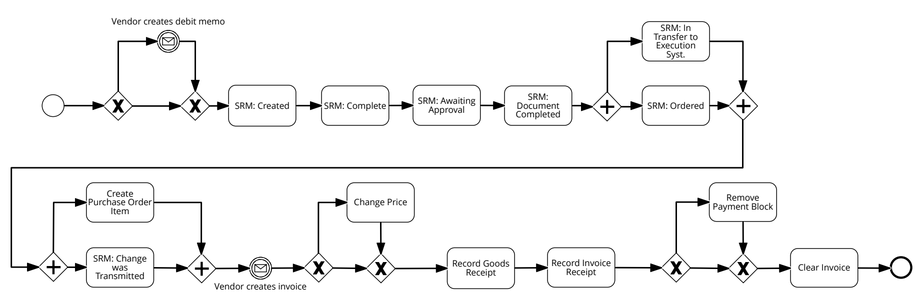


Fig. 6. BPMN process diagram for PO items following the *3-way match, Invoice before Good Receipt EC Po type*

3.3 2-way match

The 2-way matching process executed only by Company0003, in which no goods but an invoice is received, is shown in Fig. 7.

This part of the event log has a major data quality issue; the event "Change approval for Purchase Order" is occurring multiple times for each item, although it happens on a PO document level, it is copied for each item in the document. This means, although an item was created much later, the change approval activity which was executed previously for the PO, was added to its trace. The logging of every change approval for each purchase item leads to the pattern that items had multiple "Change approval for Purchase Order" events at the beginning or end of the trace. Thus, we filter this event type out from the traces prior to discovery (as it indicates wrong sequences), and added the activity in the designed BPMN model after "Create Purchase Order Item" where it occurs in 96% of the traces.

Additionally also infrequent behaviour was filtered out (traces less than 10%) resulting in 935 traces (originally 1,044 traces). Based on the filtered log, the inductive miner could produce a model with a trace fitness of 97.2% on the original log without the "Change approval for Purchase Order" events. The result was improved in a designed model in Fig. 7 by visualizing the "Vendor creates invoice" as a BPMN start event and by adding several end events. Besides, the invoice handling is often only captured in one of the PO items, although it might cover several items. This means that an invoice by a vendor can initiate the creation of several PO items, but only in one of them the invoice creation, its receipt and clearing is shown. Thus, we added a multiple-instance activity which captures this pattern that one or more purchase orders can be created and get approved. Additionally, we extended the diagram by adding common users executing an activity in its annotation.

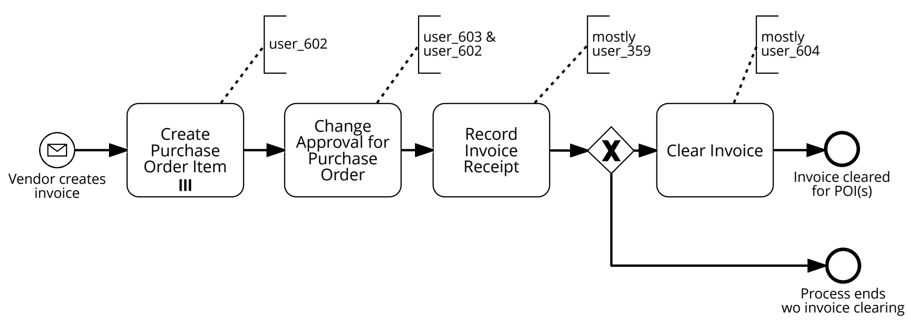


Fig. 7. BPMN process diagram for PO items following the *2-way matching*

If a vendor creates an invoice, then user_602 can create one or several PO items at the same time. Next, the user_603 changes the approval of the PO. Then, the invoice receipt is recorded, mostly by user_395 and also usually the invoice cleared, mostly by user_604.

The most common variant in this part of the log, with 22%, consists only of the two activities "Create Purchase Order Item" and "Change Approval for Purchase Order". Those are the cases which were created together with several others based on one incoming invoice where the invoice handling is not recorded.

The second most common variant 21% covers the exactly the described process, without having the activity "Clear invoice".

3.4 Consignment

The consignment process, where a good is received in the warehouse and is paid only after usage, is shown as BPMN process diagram in Fig. 8. The actual usage of the product is not anymore part of the process log, such that the net value of all purchase order items is zero.

Filtering out incomplete cases (not ending with "Record goods receipt", "Delete purchase order item", "Cancel goods receipt") and infrequent behavior (traces less than 10%), left 12,172 traces from the original 14,498 traces (cf. Table 1). On this filtered log, we ran the inductive miner which resulted with a model having a trace fitness of 97% by replaying the complete log for consignment. The resulting process model of the inductive miner was minimally corrected by showing different end events and a loop activity.

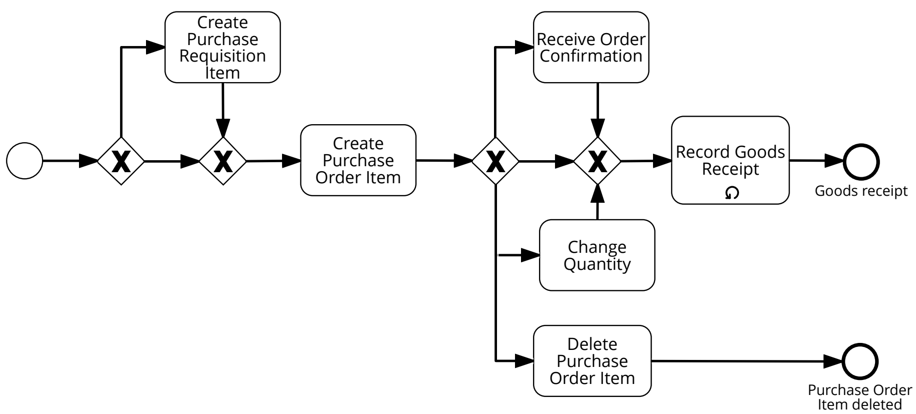


Fig. 8. BPMN process diagram for *Consignment*

Most commonly in consignment, first a purchase order item is created and then the goods are received covering 61% of the traces (8,776 traces). Goods can also received multiple times. This is visualized in the BPMN process diagram as a loop activity, which is finished as soon as all ordered goods were obtained. In 9% of the traces, it could be observed that the goods were received more than two times.

Since September 2018, we could observe a slight drift. Since then, first Create Purchase Requisition Item is executed before the purchase order item is created. This is the second most common variant with 12% of the traces.

After the purchase order was created also three other activities can occur: "Change Quantity" (3.7% of the traces) or "Receive Order Confirmation" (2.1%

of the traces), where the goods are still received, or "Delete Purchase Order Item" (1.2% of the traces) where the case is then terminated.

4 Detection of Compliance Issues with Rule-based Conformance Checking

Conformance checking aims to analyze the relationship between observed and desirable process behavior. The later one can be, for example, expressed as a normative business model or defined by rules. Rules can be derived from process models as a set of constraints given by the model's control flow. In a less technical perspective, such rules can also express business rules, like fulfilling the four-eyes principle, or more related to the given process, particular matching behavior. Compared to other conformance checking approaches, rule checking provides useful but straightforward insights into process conformance. In general, there exist five types of rules: (i) ordering rules, (ii) cardinality rules, (iii) exclusiveness rules, (iv) response rules, and (v) precedence rules [3].

Rules are always defined for a pair of activities. While the first three types are quite self-explanatory, response rules define whenever one activity occurs in a trace, there eventually has to be the other activity in the same trace. Similarly, precedence rules define activities, which have to occur in the same trace before the requesting activity takes place. By counting the violation and satisfaction of each rule, the conformance can be measured.

In the following, we analyze rules for each item type and point out violations which are strongly connected to compliance issues and therefore of particular interest. However, across all item types, response rules are frequently violated. One crucial factor for this could be a relatively large number of open cases. Therefore we do not consider them in our compliance analysis. Further, we will not report violations of cardinality rules in more detail, as we covered them in section 3 already. In general, these rules are strongly connected to rework and loops of activities, which are not modeled in our discovered process models since they only occur infrequently. In the context of the given process, such exceptional behavior may also not be considered as conformance checking problem since various goods and invoices could be received for single line items.

We implemented a Python library[4] to define and check rules for a given event log. We expand the given definition of precedence and response rules given in [3]. Since by default one occurrence of either the preceding or responding activity satisfies the respective rule, we allow enforcing the existence of a matching partner for all requesting activities, e.g. in one trace for each instance of "Record Goods Receipft" a "Record Invoice Receipt" activity has to exist. In the following, the default case will be marked with a "True" flag.

[4] https://github.com/bptlab/bpic19

Rule	Preceding activity	Requesting activity	# violations	% of traces	single occ.
	3-way match, Invoice after Goods Receipt with EC				
1	Record Goods Receipt	→ Clear Invoice	31	**5.80**	False
2	Record Invoice Receipt	→ Clear Invoice	3	0.67	False
3	Record Goods Receipt	→ Record Invoice Receipt	55	**6.88**	False
	3-way match, Invoice after GR without EC				
4	Record Goods Receipt	→ Clear Invoice	534	**4.58**	False
5	Record Invoice Receipt	→ Clear Invoice	91	0.87	False
6	Record Goods Receipt	→ Record Invoice Receipt	861	**5.02**	False
	3-way match, Invoice before GR with EC				
7	Record Invoice Receipt	→ Clear Invoice	3	0.28	False
8	Set Payment Block	→ Remove Payment Block	101	**100.00**	False
9	Record Goods Receipt	→ Clear Invoice	18	**2.36**	False
10	Record Goods Receipt	→ Clear Invoice	1	0.14	True
	3-way match, Invoice before GR without EC				
11	Record Invoice Receipt	→ Clear Invoice	1,670	0.91	False
12	Record Invoice Receipt	→ Clear Invoice	399	0.23	True
13	Record Goods Receipt	→ Clear Invoice	5,552	**2.83**	False
14	Record Goods Receipt	→ Clear Invoice	653	0.38	True
15	Set Payment Block	→ Remove Payment Block	54,532	**99.98**	False
	2-way matching				
16	Record Invoice Receipt	→ Clear Invoice	0	0	False

Table 2. Considered precedence rules for compliance analysis

4.1 3-way Match, Invoice after Goods Receipt

For this item category recording goods before recording the invoice is crucial and any violation a compliance issue.

EC purchase orders As shown in table 4, about 7% of all cases in this type, that contain "Record Invoice Receipt" at least once, violate this rule. Further in three cases, not all recorded invoices have been cleared and in about 6% of all cases invoices have been payed without recording the goods first. Paying an invoice without receiving goods may hint for fraud. The average amount of these items is 130,322.85€. The most common spend area (13 cases) is "Marketing", followed by "Workforce Services" (10 cases). The same observations can be made for the third rule.

Standard and Framework orders This orders share similar conformance and compliance issues as the previously described ones. In about 4.5% of all considered cases, the invoice has been cleared before recording any goods. Slightly more frequent (5%) an invoice has been recorded before recording goods. Regarding the first rule, two vendors, namely vendorID_0404 and vendorID_0236, stand out. Out of 423 violated cases, the first vendor participates in 89 cases, the second in 82 cases. Further, the majority of cases are of item type "Standard" (281 cases), followed by "Service" types (124 cases). Finally, about 46% of all cases belong to "Sales".

vendor ID	participated cases
vendorID_0118	385
vendorID_0246	234
vendorID_0136	166
vendorID_0108	156
vendorID_0236	151
vendorID_0110	134
vendorID_0404	126
vendorID_0122	114

Table 3. Most frequent vendors, participating in cases which violate rule 13 in table 4 (all vendors in average = 10 and median = 2).

For the second rule, the same vendors (vendorID_0404: 120 times, vendorID_0236: 88 times out of 533) participate the most in all affected cases. Again around 44% of all cases belong to "Sales", followed by "Packaging" in about 17%.

4.2 3-way Match, Invoice before Goods Receipt

PO of this item category can be handled more flexible. Since it is possible to record an invoice receipt before and after goods have been received. However, if an invoice has been received before the goods received a payment block has to be set.

EC purchase orders As stated in table 4, all cases with the activity "Remove Payment Block" lack of a preceding "Set Payment Block" activity. This activity is not contained in any case at all. While all purchase orders are "Standard" types, almost 50% of the cases belong to the "Enterprise Service" spend area. Since this activity is not recorded once, it is may hint for another data quality issue. In about 2.4% of all cases, invoices have been cleared without receiving goods before. However, there seems no direct root cause to exist.

Standard PO Most cases of the log fall into this item type. Like described before, the "Set Payment Block" does not occur in 99.98% of all cases containing "Remove Payment Block". Around 52% of these cases belong to "Packaging" and another 25% to the "Sales" area. In total 98% are standard orders, while the remaining 2% are split between "Subcontracting" and "Third-party" orders. Again, this might be hint for a general data collection issue.

The 13'th rule, "Record Goods Receipt" before "Clear Invoice", is violated in 2.83% of all cases. Eight vendors participate in these cases way above the average of 10 times, as listed in table 4.2. In the median, each vendor participates only twice. About one-third of these cases belong to the "Sales" area, 22% to "Packaging" and 21% fall into "Trading & End Products". Together, these spend areas make about 30% of the cumulated order value of the affected orders.

4.3 2-way Match and Consignment

Cases of purchase orders belonging to either 2-way match or consignment are the least violated ones. Since there are no invoices recorded for consignment purchase orders, there cannot be any business rules derived from the control flow, for compliance analysis. Also, only one rule could be derived for 2-way match items. According to our analysis, this is never violated in the observed cases (see table 4).

5 Process Performance

The event log of this year's challenge provides timestamps for the completion of each activity, such that throughput time can be further analyzed as performance dimension. For this, we first define the performance measures for the *Purchase-To-Pay* process whereby we consider two types of customers: (1) the internal departments and (2) the vendors. Then, the results are visualized and presented, based on which, question were raised. Finally, these are answered by a root cause analysis.

5.1 Definition of Performance Indicators

A *Purchase-To-Pay* process has actually two customers to-be served: (1) the internal production department or other departments (called here internal customer) requiring certain goods for producing the final products based on which the company generates revenue and (2) the suppliers which deliver the goods. A good supplier relation is important to receive high quality products, good prices, and a fast delivery. Based on these, we define performance measurements for the internal customers and the vendors.

Internal Customer. For internal customers, we assume that fast processing of the purchase requisition is important. Thus, the following aspects are measured:

– *From Requisition/SRM to PO Item* (Time between "Create Purchase Requisition Item" or "SRM:created' and "Create Purchase Order Item")
– *From PO Item to Goods* (Time between "Create Purchase Order Item" and "Record Good receipt")

Whereas the first indicator measures the time until a PO item is created, the latter measures the time between its creation and the good receipt.

Vendor Relationship. For vendors, we assume that they want to get quickly paid for the delivered goods.

– *From invoice to invoice receipt* (Time between "Vendor creates invoice" and "Record Invoice receipt")
– *From invoice receipt to clear invoice* (Time between "Record Invoice receipt" and "Clear Invoice")

– *From goods to clear invoice* (Time between "Record good receipt" and "Clear Invoice")

The first performance indicator computes how long it takes to actually record the invoice of the vendor. The second and third one measure the time between invoice receipt and also good receipt until the invoice is finally cleared which is important for the vendor.

5.2 Results

We measured the performance indicators with the help of Lana Labs Tool by using the dashboard functionality to analyze the throughput for a selected eventually follow relation. The tool provides the mean and the median for the throughput time between two selected events. We performed this evaluation individually for each item category and considered thereby only those performance indicators relevant for a certain category. Traces with incorrect timestamps as reported in Section 2.3 were filtered out beforehand. The results are visualized in Fig. 9 and Fig. 11 and presented in the following.

Internal Customer Fig. 9 shows the mean and median throughput times in days to create a PO item based on a requisition (or SRM:created for EC purchase orders) and to actually receive the goods for the different item categories. Thereby, it is distinguished between without and with EC purchase orders for the 3-way match procedure, because EC purchase orders have the extra SRM activities involved. 2-way match is not considered here because no tangible goods are received which can be checked, such as government payments and rents.

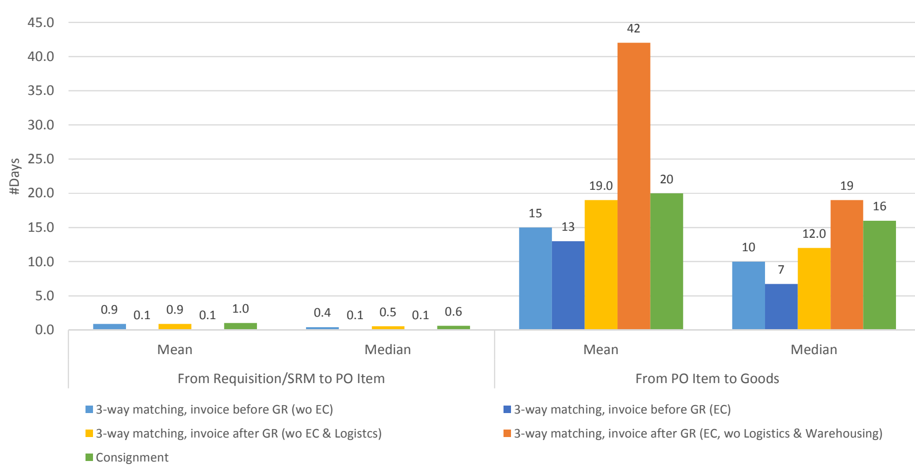


Fig. 9. Resulting performance measures for the internal customer (throughput time for creating a PO item and for receiving the goods)

From the recorded data, it seems that a PO item is usually generated quickly within a day. Especially for EC purchase orders, where a SRM process runs before the actual creation of the PO item, this is conducted in less than 2.5 hours in most cases.

After placing the PO item, goods are delivered in different time. For the items following the **3-way match, invoice before GR** it is in average 15 days (median value 10 days) and for EC purchase orders in average 13 days (median value 7 days). That means that there are some outliers. This item category has the most items processed and most of them are production-relevant (85%). Thus, it is positive that these items are delivered in a relatively short timeframe. Still, we want to place a question, because these PO items seems to be very relevant for the company: *"Why can we observe a difference between the mean and median value for the delivery time of items in the category: 3-way match, invoice before GR?"*.

Regarding **3-way match, invoice after GR**, we filtered out some cases to report meaningful results. In the case of non-EC purchase orders, logistics items (4,498 traces, 32% of these traces), such as road were filtered because they are usually received instantly leading to a median of 4 minutes and an average value of 7 days. Considering not the logistics items, the average receipt time is 19 days with a median value of 12 days. In comparison, EC purchase orders need in average 42 days and have mean of 19 days when not considering items of the spend area *Logistics* and *Workflow Services* (18% of these traces, 93 traces). These encompass Express, HR Services, and Third Party Labor and have quite long times until the good is recognized. Still, they have a comparable long delivery time because this type of purchase orders includes usually the payment of services, e.g., laboratory services, consultancy, design, market research etc. (cf. Fig. 10) We assume that "Record good receipt" is recorded as soon as the service is terminated, such that a longer throughput time is not harmful as long as it takes the time as specified in the contract of the vendor. This is not part of the data.

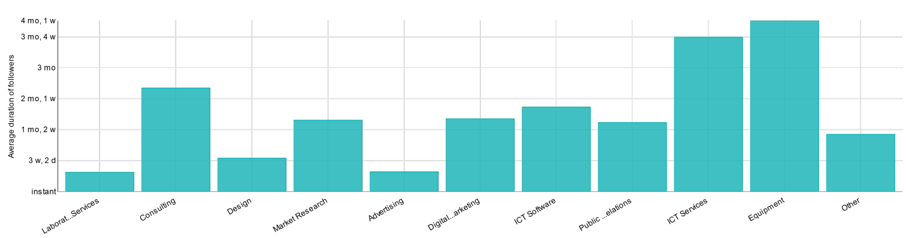


Fig. 10. Average throughput times from PO item creation until good receipt for EC POs in different sub-spend areas in 3-way match after GR

Consignment PO items are received in average in 20 days with a median value of 16 days. We assume that this timeframe is not harmful for the internal customers as it delivered in a warehouse where usually enough capacity should be available. Here it would be interesting to compare the delivery times with the warehouse data. It could be observed that Titanium Dioxides (ordered in 2,453 PO items) needs longer time until goods are receipt, up to 2 months (56 days) in average.

Vendor Relationship Fig. 11 shows the throughput time in days as mean and median from the creation of an invoice until recording it, from invoice receipt until clearing, and from goods receipt until clearing. Thereby, it is again distinguished between the different item categories. Consignment items are not considered because the invoice handling for them is not included in the event log.

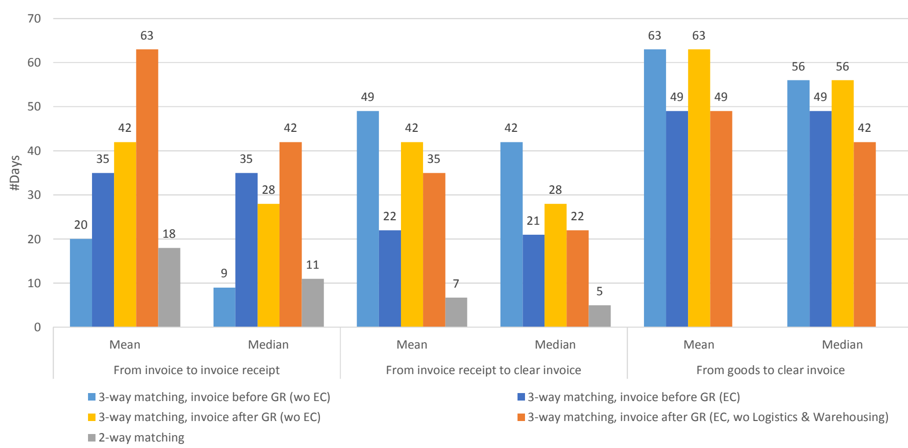


Fig. 11. Resulting performance measures for the vendor relationship (throughput times between goods, invoices, and payment)

For items of the category **3-way match, invoice before GR**, the invoice could be received before receiving the goods.

Invoice receipt usually happens in average 20 days (median: 9 days) after the invoice was placed. For EC purchase order items, it is done mostly after 35 days. After invoice receipt, it needs in average 49 days with a median of 42 days to clear the invoice. For EC purchase order, the clearing after invoice receipt is quite stably conducted after mostly 21 days, and also the clearing after goods receipt is executed mostly after 49 days. For all other purchase order items, the time between goods receipt and clearing the invoice is in average 63 days with a median of 56 days.

For items of the category **3-way match, invoice after GR**, first the goods have to be receipt until the invoice can be receipt and cleared. That means the vendor can create its invoice at any time but it is only receipt after goods receipt. Therefore it is not surprisingly that EC purchase order items in this category take more time to receive them because also the delivery takes longer, in average 63 days with a median of 42 days. The Standard and Framework PO items need in average 42 days with a median of 28 days what is in comparison to receipt of goods quite high.

The clearing of the invoice needs in average another 42 days (for EC: 35 days) with a median of 28 days (for EC: 28 days).

This leads to a total time between goods receipt and clearing the invoice of in average 63 days with a median of 56 days. For EC purchase orders is less, in average 49 days with a median of 42 days. This leads to the assumption that vendors are creating their invoices quite early. Between the two categories of 3-way match not so much differences can be observed, although the invoice in the first category could be received earlier and then payed as soon as the goods were received. Thus, our question is: *"Why is the time between goods receipt and clear invoice is comparatively long for items where the invoice can be already received before the goods receipt?"*

2-way matching only requests to check the invoice, such that we can observe a quick clearing of an invoice as soon as the invoice receipt was recorded, in average in 7 days with a median of 5 days. However, it takes usually in average 18 days (median of 11 days) to record the invoice receipt after the vendor has created the invoice. That might be due to that usually first the vendor sends the invoice and then the purchase order (item) is created. We observed some outliers in that category: the purchase order items for *Vendor_1914* in the sub-spend area *Business park* (12 cases) which were created more than 10 months later after the vendor has created the invoice.

5.3 Deep-dive Analysis

In the following we want to discuss the questions placed earlier in more depth:

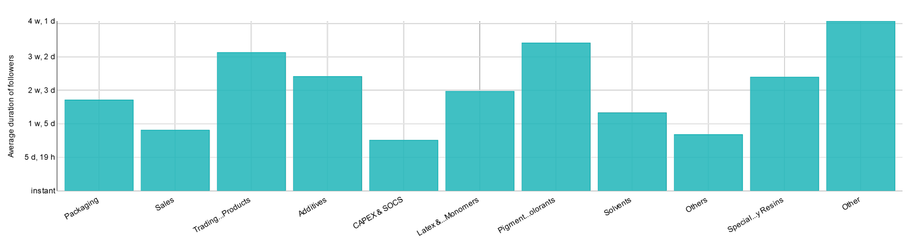


Fig. 12. Average throughput times from PO item creation until good receipt for the different spend areas in 3-way match before GR, without EC.

1. "Why can we observe a difference between the mean and median value for the delivery time of items in the category: 3-way match, invoice before GR?" Fig. 9 shows the throughput times for the different spend areas for this item category without EC purchase orders. *Trading&End Products* and *Pigments&Colorants* of the most common items have higher average delivery time than the others. When looking on the changes which can occur after the PO item was created, the "Change quantity" is the most relevant ones. This happens in 13140 cases (6.9% of these cases) and leads to an increase in the throughput time up to in average 28 days (median of 23 days), which is twice as much than in the most cases. Thus, this activity has a negative influence for a fast delivery of products.

2. "Why is the time between goods receipt and clear invoice is comparatively long for items in which the invoice can be already received before the goods receipt?" Interestingly, the invoice is received before the good receipt in only 8% (16,168 cases). Those have also a minor improved throughput between goods received to invoice clearing of two weeks less than the reported average of all those PO items. The Fig. 13 shows that items from different sub-spend areas are also treated differently. For example, the *Lables* have a comparable long payment time of in average three month whereas, *MRO (components)* only takes in average 3 month.

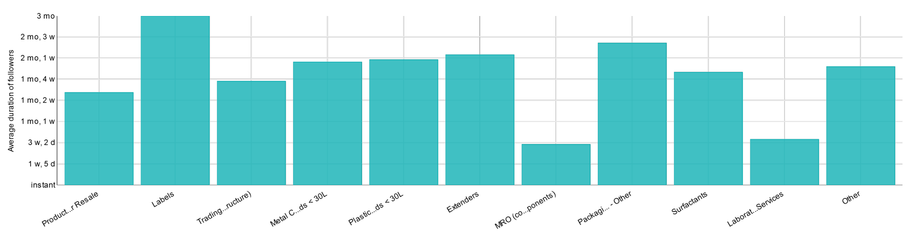


Fig. 13. Average throughput times from good receipt until clear invoice for the different sub-spend areas in 3-way match before GR, without EC.

In general, we assume that regularly a batch for the invoice is run. How this batch behavior influence the performance for invoice clearing needs to be further evaluated.

6 Conclusion

After presenting some statistics, insights, and data quality issues on the event log for a *Purchase-to-Pay* Process, we discovered in this report the most common flows of the different item categories and presented them as understandable BPMN process diagrams. In this step, the high variance and flexibility of the process were challenging.

Next, we performed conformance checking and discovered compliance issues based on control-flow rules. To this end, we developed a Python tool, made publicly available. Among other things, we found that 54,534 times, a payment block was not recorded before removing it. For 3-way match items, in 6,135 observations invoices have been cleared without recording goods before. Last but not least, 1,767 times invoices have been cleared without recording an invoice before.

Finally, we defined performance indicators for the two customers of this process: the internal departments of the company and the vendors. This revealed that the receipt of goods takes place within less than two weeks for the standard purchase orders which are relevant for production. However, clearing the invoice takes several weeks. The same holds for the *3-way match, invoice before goods receipt* items. Here we recommend checking whether these type of invoice could be already receipt earlier to decrease the time of payment.

We think that a fast payment of invoice can lead to a better vendor relationship. Therefore, in the future, we would like to investigate the root causes of the payment duration further. We assume that this is regularly done in a batch-like fashion for which we would like to identify the rules and pattern.

References

1. van der Aalst, W.M.P.: Process Mining: Discovery, Conformance and Enhancement of Business Processes, vol. 2. Springer (2011)
2. Bozkaya, M., Gabriels, J., van der Werf, J.M.: Process diagnostics: a method based on process mining. In: 2009 International Conference on Information, Process, and Knowledge Management. pp. 22–27. IEEE (2009)
3. Carmona, J., van Dongen, B., Solti, A., Weidlich, M.: Conformance Checking: Relating Processes and Models. Springer (2018)
4. van Eck, M.L., Lu, X., Leemans, S.J., van der Aalst, W.M.: PM^2: A process mining project methodology. In: International Conference on Advanced Information Systems Engineering. pp. 297–313. Springer (2015)
5. Wirth, R., Hipp, J.: Crisp-dm: Towards a standard process model for data mining. In: Proceedings of the 4th international conference on the practical applications of knowledge discovery and data mining. pp. 29–39. Citeseer (2000)

BPI Challenge 2019

Dominik Hüser, Philipp Heisenberger

RWTH Aachen University
dominik.hueser@rwth-aachen.de
philipp.heisenberger@rwth-aachen.de

Abstract. Recently, Process Mining became a major technique for auditing teams in the area of business process intelligence. Especially in purchase to pay processes (P2P process), process mining can uncover unrealized optimization potential, which again results in overall cost reduction in the company. In this year's BPI challenge we therefore deal with a P2P process log offered by a company which is operating in the context of coatings and paints.
The report deals with the detailed analysis of this log, starting with a general analysis of the unfiltered log, to gain information how to filter the log properly. The filtered log is then analyzed regarding the given case attributes. The main part will focus on the creation of a process model that describes the as-is process properly. Reasonable sublogs are defined to gain several simple and precise models rather than one large, complicated, imprecise one. We split the log according to the cases' category: SRM (supplier relationship management), 2-way matches, consignments, 3-way matches with invoice before goods receipt and 3-way matches with invoice after good receipt. The main KPIs this report focuses on are the throughput time, automation rate, rework rate as well as a social analysis. Coming from found bottlenecks and inefficiencies, we will identify deviating cases, vendors and resources to finally round up the whole report with a summary.
To mine proper models we will rely on the tools ProM, RapidProM and PM4Py. We use Celonis to gain further insights into the log, and to define proper KPIs via the Process Querying Language (PQL).

Keywords: Business Process Intelligence, Purchase to Pay, Process Mining

1 General Information

According to [2] starts a typical Purchase to Pay Process (P2P) with a purchase requisition order, which is a formal request for goods or services, followed by a vendor selection who then receives the purchase order from the company. After that, we receive the goods and receiving documents. Then the vendors sends the invoice which then is entered to the ERP. Then a three way match is performed automatically and line items don't follow that schema need to be investigated manually. After that the invoice can be cleared.

1.1 General Information about the unfiltered log

The log [3] which will be analysed in the following describes a purchase to pay process excluding the approval workflows for purchase orders and invoices. From fig. 1, obtained via Prom 6, we can find out key information about the log: We observe 251734 cases with in total 1595923 events of 42 different event classes. Every event got the same type. On average we got 6 events of 5 different event classes, variating between one and 990 events per case, while we observe between 1 and 20 different event classes per case. The log contains events which were executed between 26.01.1948 and 09.04.2020.

Fig. 1. General Information about the unfiltered log.

In fig. 2 we see the 42 different event classes and their occurrences. We observe that 32 of the the 42 have less than one percent relative occurrence. On the other hand we see that "Record Goods Receipt" (19,7%), "Create Purchase Order Item" (15,8%), "Record Invoice Receipt" (14,33%), "Vendor creates Invoice" (13,8%), "Clear Invoice" (12,2%) and "Record Service Entry Sheet" (10,3%) appear highly frequently. As we observe later those events often are executed several times per case. Additionally, we spot several event classes labeled with the prefix SRM, while activities without that prefix refer to standard purchase to pay actvities, SRM activities relates to the Supplier Relation Management.

All events

Total number of classes: **42**

Class	Occurrences (absolute)	Occurrences (relative)
Record Goods Receipt	314097	19,681%
Create Purchase Order Item	251734	15,774%
Record Invoice Receipt	228760	14,334%
Vendor creates invoice	219919	13,78%
Clear Invoice	194393	12,181%
Record Service Entry Sheet	164975	10,337%
Remove Payment Block	57136	3,58%
Create Purchase Requisition Item	46592	2,919%
Receive Order Confirmation	32065	2,009%
Change Quantity	21449	1,344%
Change Price	12423	0,778%
Delete Purchase Order Item	8875	0,556%
Change Approval for Purchase Order	7541	0,473%
Cancel Invoice Receipt	7096	0,445%
Vendor creates debit memo	6255	0,392%
Change Delivery Indicator	3289	0,206%
Cancel Goods Receipt	3096	0,194%
SRM: In Transfer to Execution Syst.	1765	0,111%
SRM: Ordered	1628	0,102%
SRM: Created	1628	0,102%
SRM: Awaiting Approval	1628	0,102%
SRM: Document Completed	1628	0,102%
SRM: Complete	1628	0,102%
Release Purchase Order	1610	0,101%
SRM: Change was Transmitted	1440	0,09%
Reactivate Purchase Order Item	543	0,034%
Block Purchase Order Item	514	0,032%
Cancel Subsequent Invoice	491	0,031%
Release Purchase Requisition	467	0,029%
Change Storage Location	415	0,026%
Update Order Confirmation	244	0,015%
SRM: Deleted	188	0,012%
Record Subsequent Invoice	167	0,01%
Set Payment Block	124	0,008%
SRM: Transfer Failed (E.Sys.)	45	0,003%
Change Currency	35	0,002%
Change Final Invoice Indicator	11	0,001%
SRM: Transaction Completed	8	0,001%
Change payment term	7	0,0%
SRM: Held	6	0,0%
SRM: Incomplete	6	0,0%
Change Rejection Indicator	2	0,0%

Fig. 2. General information about the unfiltered log.

Start events

Total number of classes: **8**

Class	Occurrences (absolute)	Occurrences (relative)
Create Purchase Order Item	199867	79,396%
Create Purchase Requisition Item	46526	18,482%
Vendor creates invoice	3457	1,373%
SRM: Created	1360	0,54%
Change Approval for Purchase Order	378	0,15%
Vendor creates debit memo	122	0,048%
Release Purchase Order	22	0,009%
Change Currency	2	0,001%

End events

Total number of classes: **32**

Class	Occurrences (absolute)	Occurrences (relative)
Clear Invoice	181328	72,032%
Record Invoice Receipt	23091	9,173%
Record Goods Receipt	22776	9,048%
Delete Purchase Order Item	8123	3,227%
Remove Payment Block	5397	2,144%
Create Purchase Order Item	4447	1,767%
Cancel Invoice Receipt	1348	0,535%
Change Approval for Purchase Order	1346	0,535%
Record Service Entry Sheet	921	0,366%
Change Delivery Indicator	718	0,285%
Change Price	529	0,21%
Change Quantity	525	0,209%
Cancel Goods Receipt	440	0,175%
Receive Order Confirmation	232	0,092%
Block Purchase Order Item	167	0,066%
Vendor creates invoice	68	0,027%
Set Payment Block	61	0,024%
SRM: Change was Transmitted	47	0,019%
SRM: Transfer Failed (E.Sys.)	44	0,017%
SRM: In Transfer to Execution Syst.	30	0,012%
Cancel Subsequent Invoice	24	0,01%
Change Storage Location	20	0,008%
Reactivate Purchase Order Item	14	0,006%
Record Subsequent Invoice	9	0,004%
SRM: Deleted	9	0,004%
SRM: Transaction Completed	6	0,002%
Vendor creates debit memo	5	0,002%
Update Order Confirmation	3	0,001%
Change Final Invoice Indicator	3	0,001%
Release Purchase Order	1	0,0%
Change payment term	1	0,0%
Change Currency	1	0,0%

Fig. 3. General information about the unfiltered log.

In fig. 3, we see that nearly 80% of all cases start with the creation of an Purchase Order Item, while 18% start with the creation of a Purchase Requisition Item. Much less frequently a process starts with a creation of the invoice by the vendor (1,4%). 72% of all cases then end with "Clear Invoice", 9,2% with "Record Invoice Receipt" and 9% with "Record Goods Receipt". Much less often a case ends with other events. We need to keep in mind, that this is a real log containing information about year 2018. Cases which were not completed in the logged timeframe are now falsely appearing as complete cases with an end activity from the middle of the actual incomplete case. How we deal with that issue will be described later on.

The process is highly complex with in total 11973 different case variants. The four most common variant can be seen in fig. 4.

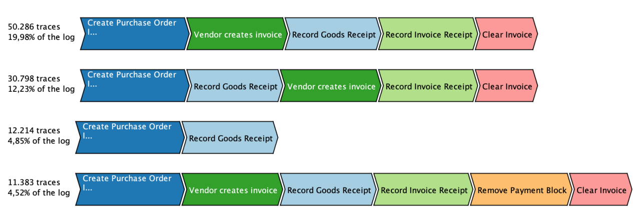


Fig. 4. Most common variants in the unfiltered log.

1.2 Filter the log

In the previous section we already talked about the time frame. We observe that there are cases reaching back to 1946 or already contain event executions from the future, namely from the year 2020. In a first step we filter the cases on the timestamps of the first event which must be executed on the 01.01.18 or later. In a second step we filter cases such that we only observe events where the last event is executed at latest in present, so that we won't regard cases containing future events. We filtered via a prom4py python script.

As already mentioned, do we also need to deal with filtering incomplete cases. This task is not as easy as it seems since incomplete cases can easily be confused with infrequent behavior of the process. We look into fig. 3 again to get information about the frequency of end activities. We immediately spot candidate endings which might have resulted from incompleteness (cf. "Release Purchase Order", "Change Payment Term", "Change Currency", each occurring only once in the end of a trace). We still decided to keep those traces inside the log and deal with them later, by leaving out a certain part of the less frequent cases when mining. The advantage of doing the filtering on the fly while mining is the possibility to play with the severeness of the filter operation and see the impact on the model much better, so that we hopefully find a much better fitting model describing the process more precisely. After filtering we lost 265 cases. That reduces the maximum number of events per case from 990 to 868 events per case. And the maximum amount of event classes per case from 20 to 19. Additionally, we eliminated 61 case variants.

1.3 Insight into the log

After filtering the log for outliers we can now make some basic observations about the given cases. We primarily use Celonis, Prom 6 and pm4py to perform that task.

Let us first regard the PO Document Types (cf. fig. 5): We see that by far most PO documents are standard PO documents (74598 docs), 978 are EC Purchase Orders and 704 are Framework Orders. If we look at the PO Item types (fig. 5), we see that 87% of all cases are Standard PO Items, while 6% of all cases are Consignment and 2% are Service PO Items.

Company When we regard the subsidiary from where this purchase originated, which is defined in the `Case-Company` attribute of an item, we spot the following distribution in fig. 6. Nearly all items are related to companyID_0000. Only 1027 (a share of 0,41%) are related to companyID_0003. CompanyID_0001 and companyID_0002 are only related to two items each. We want to dig deeper into those three less common cases, quickly: In fig. 7 we see dotted charts regarding the company and several other criteria. What we basically conclude is the following: For some reasons we don't have any information about the document type, GR Indicator, Item Category nor the spend area for comapnyID_0001 and

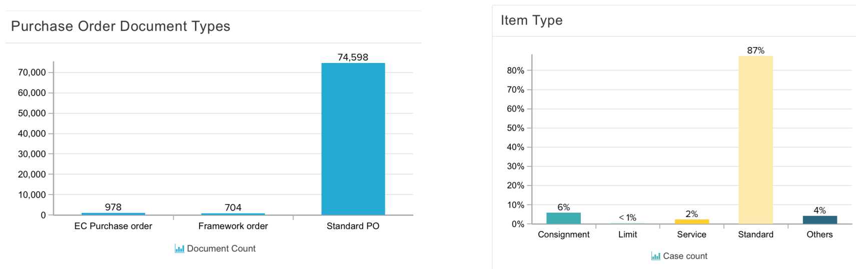


Fig. 5. Spend Area.

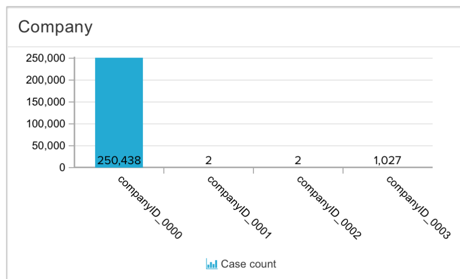


Fig. 6. Most common variants in the filtered log.

comapnyID_0002. The reason for that might be in the process: In all of those four cases the PO Item was deleted in the end of the process.

When we compare companyID_0000 (in the following called company 0) and companyID_0003 (in the following called company 3) then we observe that company 3 is related to Framework Orders only, while company 0 is related to all three types. Related to company 3 are only PO Items without GR. This also leads to the obvious observation that company 3 only deals with 2-way match items, as visible in fig. 7. This observation again fits to another observation we made earlier: Only Framework Orders are related to company 3.

When we compare the spend class of company 0 and 3 we see that company 3 only performs purchase orders of class NPR and OTHER while company 0 contains all kind of PO Item classes: NPR, PR, OTHER and NULL. When we dig deeper here we see that company 3 exclusively performs PO Items of spend area "Real Estate" and "Energy". Both companies perform items of spend area "Enterprise Services", "CAPEX & SOCS" , "Workforce Services" and "Others". Those are basic purchases which need to be made in every company. Excluding those areas, we can split company 0 and 3 in following two clusters:

- **Company 0:** Actual production (Titanium Dioxide, Latex and Monomers, Pigments and Colorants, Solvents, Specialty Resins, Commodity Resins, Additives) and shipping and related (Logistics, Packing, Sales, Marketing)
- **Company 3:** Real Estate, Energy

Let us in the following look closer into the spend areas of the company.

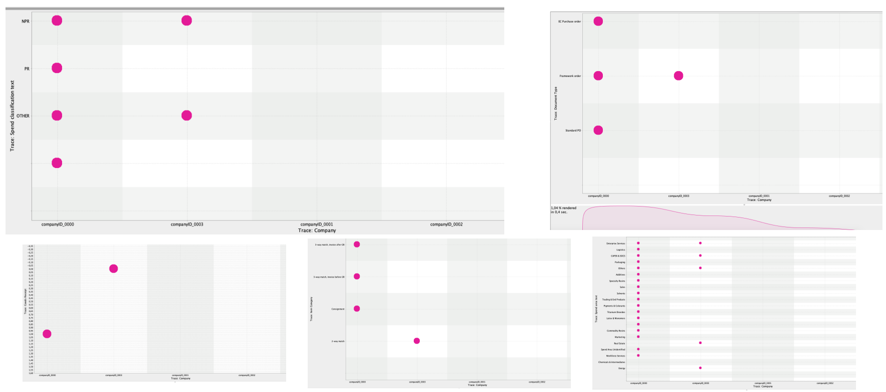


Fig. 7. Several comparisons betweent the company and a) the class, b) the doctype, c) the GR, d) the Item Spend category and e) the Spend Category Text

Spend Areas In fig. 8 we see that most cases are assigned to class "PR" (more than 160000), followed by "NPR" (more than 80000) and then followed by "OTHERS" and NULL valued. If we regard the Spend Area explicitly then we see than most cases deal with "Packing" (109181), Sales (65790), Trading and End Products (22204) followed by "Additives" and others.

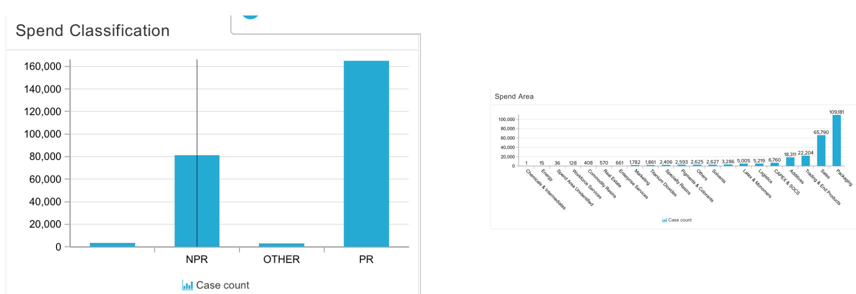


Fig. 8. Spend Classification and Area.

Vendors When we regard the vendors we spot in total 1965 vendors. In fig. 9 we see how many vendors each area of spend got in the log. We see that especially many different vendors are present in the area of "Sales", "Additives", "Packing" and "Marketing". Interesting is the amount of one time vendors: In total we saw that there are 432 vendors where we performed a purchase order only once in the given time frame.

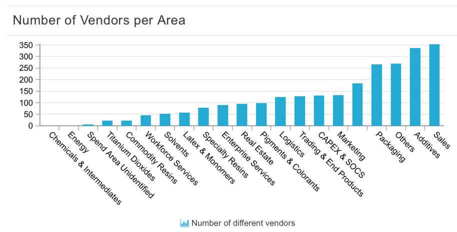


Fig. 9. Number of different vendors per spend area.

Matching When we look at the case count of each of the categories we see that by far (220811 times) most items are 3 way matches where the invoice was received before the good. Followed by 3-way matches where the good arrived before the invoice (15134 times). Tightly again followed by Consignments (14497 occurrences) and 2-way matches (1027 times). Interesting is the distribution of the actual total value of the purchase order items. We therefore summed up the net value of the creation of a purchase order item. Here we see that most value is actually in the category 3-way match with invoice after GR, followed by 3-way matches where the invoice arrived before the goods. In the previous analysis we saw that all 2-way matches were performed by company 3 and that those are all framework orders. That makes sense since it primarily deals with purchases of non-material goods, and therefore no goods are receipt. In the dotted chart below we additionally see that some framework orders are actually processed as 3 way match where the invoice arrived after the GR. Standard POs are either 3-matches or Consignments while EC Purchase Orders are always 3-way-matches.

1.4 Usability of the log

Due to the large variety of cases inside the log, we prefer to split the log into sublogs and then create one model for each sublog, to gain more detailed insight and simpler models. How we split further is described in the next section.

2 Model based Mining

In fig. 11 we see the distribution of throughput times. On average the throughput time is 70 days. We see additionally that there are cases, that take 383 days. To dig deeper now we derive sublogs and submodels from the overall log.

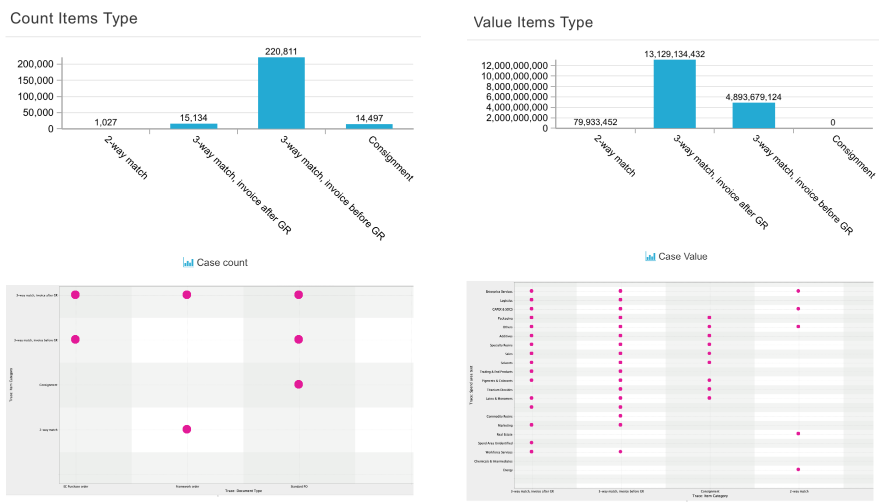


Fig. 10. Item Type Analysis.

2.1 Split data into sublogs and general mining procedure

The way we split the filtered log is visible in fig. 12. Since we first want to focus on the pure P2P process we filter out cases including SRM events. Those can then be analyzed separately, later on. The SRM filtered log is then split via the case attribute `Case-category`. The three categories are ”3-way match (invoice before GR)”, ”3-way-match (Invoice after GR)”, ”Consignment” and ”2-way-match”.

In the following we will mine one model for each of those subsets in Prom:

1. Mine a basic model with the inductive miner on ca. 90% of the most common traces.
2. Regard the traces and derivations and manually adapt the model if necessary.
3. Calculate the fitness and precision via the Multi Perspective Process Explorer / RapidProm and adapt the model manually if necessary.

That resulted in the following models:

2.2 Model A: 2-way match

Here we observe those traces which are marked as two way matches. In that case IV = false and GR = false. According to [1] we expect to receive only an invoice and no goods.

The Model In fig. 13 we see the mined model: Optionally the vendor starts with the creation of a debit memo, which happens rather infrequent as we see later. After that either 1) concurrently multiple changes to the approval for the

Fig. 11. General Throughput Time Distribution (left) and Bottlenecks by Celonis (right)

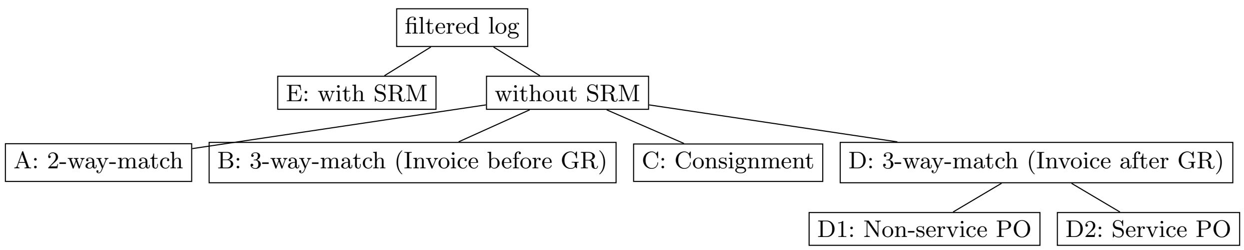


Fig. 12. Splitting the log into sublogs.

purchase order item and one creation of a purchase order item happens. Or 2) concurrently multiple changes to the approval for the purchase order item, a creation of the invoice by the vendor and the sequential flow "Create Purchase Order Item", "Record Invoice Receipt" and optionally a "Clear Invoice" happens. As expected we don't observe any Goods Receipt events as we deal with all 2-way matches, only.

Conformance On first sight we can easily see that this model is fairly simple and easy to understand. In fig. 13, we can additionally spot the fitness as well as the precision of the model on the sublog. We got a very good fitness of 98,9%. On the other hand we observe a precision of ca. 76,1%. In fig. 13 we can immediately spot the places where we lack in precision. Main reason for the lack of precision is the loop at "Change Approval for Purchase Order Item" since it allows a lot more behavior than actually visible in the log. Same holds for the place after the first concurrency.

Throughput and Flow In fig. 14 (up) we see the traces as well as the derivations from the previously mined model for two way matches.

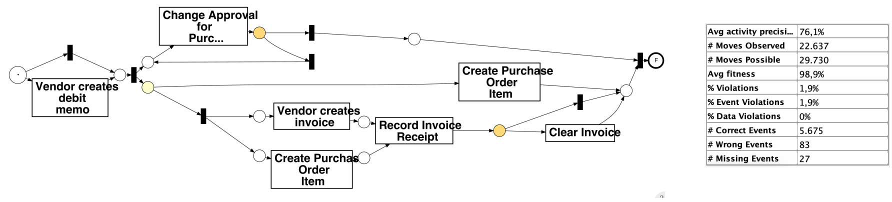


Fig. 13. Model for 2-way matches.

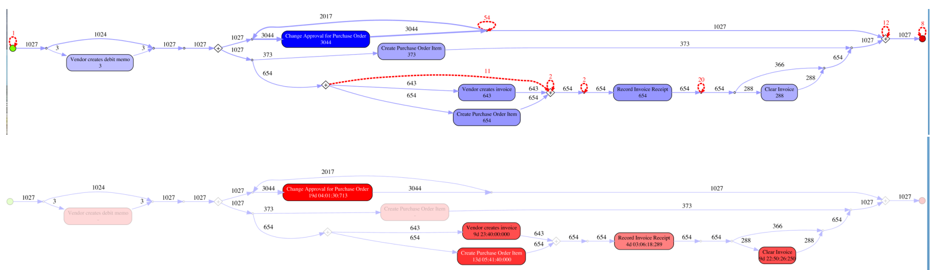


Fig. 14. Case Count and Throughput of Model A.

We see that only in 3 cases a debit memo was created by the vendor. 373 times a purchase order item was created while neither an invoice was created, received nor cleared. The other 654 cases invoices were handled. In 288 of those cases the invoice was finally cleared, while in the other 366 cases it was not. In total 3044 changes to the approval of the PO were performed, often multiple times per case. When we regard the deviations we spot no especially high number.

When we regard the throughput time of the other hand we directly see that the time between "Record Invoice Receipt" and "Clear Invoice" is 9 days and 22 hours on average.

When we filter on those cases, which got a significant higher throughput time here we spot the following: In fig. 15 we see the vendors and their average throughput time between recording of the invoice receipt and the payment. Focused on those vendors where payment took especially long time. Additionally we see those resources which are mostly clearing the invoices where the throughput is large.

We also see large waiting times for "Create Purchase Order Item", and "Change Approval for PO" of 13 and 19 days, which can't really be interpreted in this graphic though, due to the concurrent nature of the two activities.

2.3 Model B: 3-way match, invoice before goods receipt

Here we observe those traces which are marked as 3-way match, invoice before goods receipt. In that case IV = false and GR = true. From the BPI challenge project site [1] we take the following information: We expect to observe PO Items that require a GR while they don't require invoicing based on the GR. If

Cases per Vendor

Case-Vendor	Case count	Throughput time in days
vendorID_1797	1	87.00
vendorID_1690	2	84.00
vendorID_1689	4	76.00
vendorID_1687	4	61.00
vendorID_1855	1	54.00
vendorID_1859	1	49.00
vendorID_1895	2	38.00
vendorID_1853	2	35.00
vendorID_1700	1	34.00
vendorID_1717	1	31.00
vendorID_1808	2	27.00
vendorID_1844	14	27.00
vendorID_1878	2	27.00
vendorID_1950	2	22.00
vendorID_1942	1	22.00
vendorID_1868	1	22.00
vendorID_1935	2	21.00
vendorID_1867	1	21.00
vendorID_1945	1	21.00
vendorID_1947	1	21.00
vendorID_1946	1	20.00
vendorID_1688	1	20.00

Executors of Clear Invoice (>20 days)

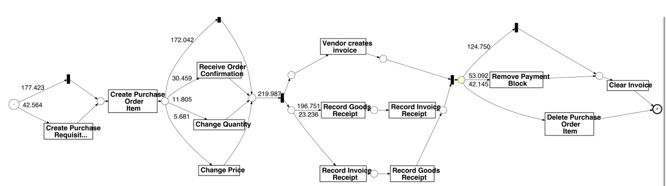
org:resource	Case count	Throughput time in days
user_294	3	52.00
user_197	12	49.00
user_205	1	48.00
user_604	27	35.00
user_359	5	27.00

Fig. 15. Throughput ($\geq$ 20 days) from "Record Invoice Receipt" to "Clear Invoice" in Model A

we record goods receipt then its ok that this happens after the invoice receipt. But the invoice should be blocked until the goods arrived (unblock). Invoices should only be cleared if goods are received and the value matches with the invoice and the value at creation of the item.

The Model In fig. 16 we see the mined model. The process starts with an optional creation of a purchase requisition item (19% of the cases). That is followed by the creation of the PO item and an xor split: Either we change the quantity or the price, receive order confirmation or do none of those. After that, in concurrency the vendor creates the invoice and we receive goods and the invoice. In 90% of the cases the goods receipt is reported before the invoice receipt. In 10% it is the other way around. In the end either the invoice is cleared direct or after a removal of a payment block or the PO item is deleted.

Conformance As already observed, is the model fairly simple. In fig. 16 we also see the fitness and precision of the model. With our model we gain a fitness of 92,3% and a precision of 99,9%.



Avg activity precisi...	99,9%
# Moves Observed	4.134.401
# Moves Possible	4.137.608
Avg fitness	92,3%
% Violations	13,6%
% Event Violations	13,6%
% Data Violations	0%
# Correct Events	1.143.381
# Wrong Events	79.977
# Missing Events	100.155

Fig. 16. Model for 3-way match, invoice before goods receipt.

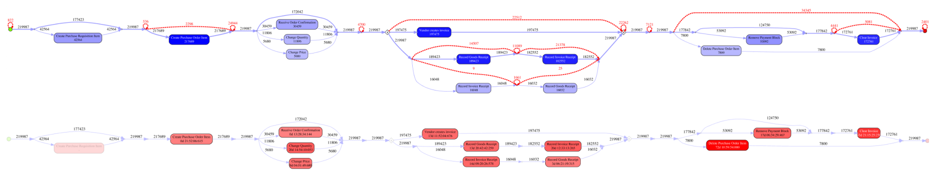


Fig. 17. Case Count and Throughput of Model B.

Frequencies We see that 80% of the cases do not contain a purchase order requisition item. We additionally see that not all cases in this log contains an order confirmation. We will look deeper into this insight later since it could result from the not perfectly fitting nature of the model. Most interesting is the later xor split: We see that 92,2% of the cases contain record goods receipt before record invoice receipt, while the other cases contain the two events ordered the other way around. In 7800 cases the purchase order item is deleted in the end. In 53092 cases a payment block needs to be removed before the invoice is cleaned finally.

Throughput Times When we look at the relevant throughput times now we see the following: When goods are receipt first, then it takes on average 20 days until the invoice is received. When the invoice is received first, then it only takes 3 days on average until the goods are received. When the purchase order item is deleted, it takes additional 72 days on average. This is definitely a bottleneck. In fig. 18 we see the throughput time from Record Goods/Invoice Receipt. In case when invoices are received before the goods, then on average 59 days pass until the invoice is cleared. In the other case that the goods are received before the invoice another 84 days pass. In general we can therefore says that the second case is significantly quicker in terms of throughput time. One reason for the high throughput times might be the mapping of items inside a purchase order. In this report, we calculate the throughput time with respect to the first occurrence of the recording and the last occurrence of the invoice clearing. This definitely needs to be kept in mind when regarding those times. We additionally want to check when invoices are cleared and when especially not, even though goods and invoices were received. This is done, now:

Fig. 18. Throughput of Model B from Good Receipt to Clear invoice when goods were received after the invoice (left) and from Invoice Receipt to Clear Invoice Receipt, when invoice was received after goods.

Derivations We additionally see the derivations: Those mostly result from traces where only a PO Item was created after an optional PO Requisition Item and then the case was over or goods were received but no invoice. The result are derivations at the first concurrent block where invoices and goods are received and in the end when usually a PO Item is deleted or payed. We therefore look again into the log, to spot cases where not both "Record Goods Receipt" and "Record Invoice Receipt" appeared. There are ca. 14000 cases from 747 vendors following that pattern. Those are especially interesting because the company thinks goods and invoices are both checked against each other in this sublog. Mostly Standard PO from company 0 with standard item type follow that pattern. In fig. 19 we additionally see that those cases usually are out of the PR and NPR class and additionally those vendors which are most commonly executing such cases. Interesting are those cases which additionally contain an invoice payment, even though not both record events occur in the log. We see that especially user 002 performs such payments the reason here is that user 002 performes most invoice clearances.

Fig. 19. Cases where not both, invoice and goods were received.

Order Confirmation Rate In fig. 21 we see the order confirmation rate of vendors (sorted by confirmation rate) of this sublog. The order confirmation rate is the fraction of cases containing an order confirmation. We defined a KPI in Celonis for the Order Confirmation Rate (cf. fig. 20) to generate that insight. Out of the 1374 vendors relevant for this sublog only 43 can hold an order confirmation rate above 70%. When we look at cases where the invoice came before the goods, then only vendor 599 (100%), vendor 120 (23%) vendor 136 (3%) perform order confirmations.

```
AVG(CASE WHEN MATCH_ACTIVITIES(NODE['Receive Order Confirmation'] ) > 0
THEN 1.0 ELSE 0.0 END)
```

Fig. 20. PQL for Order Confirmation Rate.

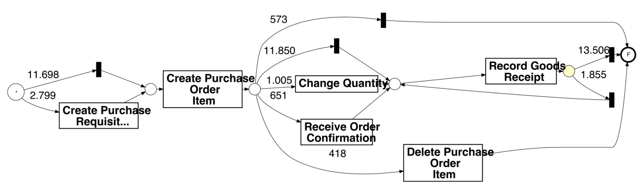
Order Confirmation Rate by Vendor

Case-Vendor	Order Confirmation Rate
vendorID_1125	100%
vendorID_1100	100%
vendorID_0549	100%
vendorID_0289	100%
vendorID_0892	100%
vendorID_1660	100%
vendorID_1300	100%
vendorID_1126	100%
vendorID_1353	100%
vendorID_0731	99%
vendorID_0159	98%
vendorID_0608	97%
vendorID_0680	96%
vendorID_0931	95%
vendorID_0290	95%
vendorID_0193	94%
vendorID_0280	93%
vendorID_0816	93%
vendorID_0293	93%
vendorID_1166	92%
vendorID_0681	92%
vendorID_0276	92%
vendorID_0103	92%
vendorID_0190	91%
vendorID_0104	90%
vendorID_0105	88%

Fig. 21. Order Confirmation rate in sublog B. (PQL in fig. 20).

2.4 Model C: Consignment

In the consignment subprocess we expect that there is no invoice handling present, since this is performed by another system [1]. GR is true and IV is false. This is indeed the case as we see in the following mined model:



Avg activity precisi...	99,9%
# Moves Observed	161.878
# Moves Possible	162.059
Avg fitness	98,1%
% Violations	4%
% Event Violations	4%
% Data Violations	0%
# Correct Events	34.678
# Wrong Events	1.402
# Missing Events	53

Fig. 22. Model for consignment.

The Model At the beginning of the process we again spot an optional "Purchase Order Requisition Item Creation". 19% of the cases contain that event. The other 81% start with "Create Purchase Order Item" directly. After that 4% of the cases stop immediately without any execution in the model. 2,88% then finish with "Delete Purchase Oder Item". For 4,5% of our cases we receive an order confirmation, for 6,9% a Quantity Change is necessary the other 81,7 % go directly to the next step. After that "Record Goods Receipt" is executed in 1855 cases this happens several times per case.

To be sure that there is no trace which contains invoice handling events, which is not conforming with the mined model, we plot in fig. 23 the occurring actvities and their frequency. As expected, we observe that especially Create PO

Item and Record Goods Receipt are executed in nearly every case. Additionally, we see that there is no execution of an invoice related activity.

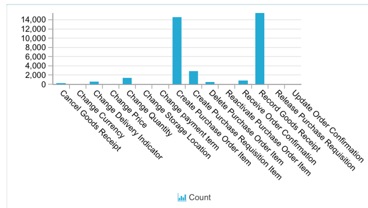


Fig. 23. Activity Count of Consignment Sublog

Conformance The mined model has fitness of 98,1% and precision of 99,9%. Therefore, represents the given subprocess quite accurately. The model is additionally simple which is a good feature for the further analysis. We see that 53 events are missing. Those can be tracked in fig. 23 again.

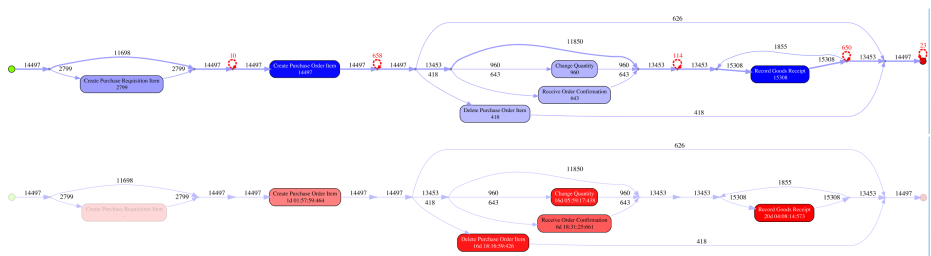


Fig. 24. Case Count and Throughput of Model C.

Flow and Derivations In fig. 24 we see the deviations and throughput time of our model w.r.t. the log. We see deviations after "Create Purchase Order Item" and "Record Goods Receipt". After Creation of PO Item we see 650 derivation, those are "Change Delivery Indicator" (229) and "Record Goods Receipt" (144) which are executed additionally to the models service. After "Received Goods" again around 650 derivations are visible. Mostly "Change Delivery Indicator" (276) and "Cancel Goods Receipt" (207) executions.

Throughput When we created a Purchase Requisition Item it takes on average one day until the corresponding Purchase Order Item was created. When the PO Item is then deleted, it takes more than 16 days on average to perform that task. Same holds for changes to the quantity: 16 days pass till this is executed. Order Confirmations are received after on average 7 days. Before "Record Goods Receipt" is executed, on average 20 days pass. We want to dig deeper into this in fig. 25, where we look into those case with especially large throughput time. When we look into the spend areas of extraordinary large throughput cases then especially "Packing" holds a large share here. Reason for that is the generally high occurrence of "Packing" cases in the log (cf. fig. 8). In the second largest spend area Titanium Dioxides, one third of those traces in the log end up as Consignment Cases with large throughput. Additionally, we spot those vendors which are most commonly occurring in high throughput time cases.

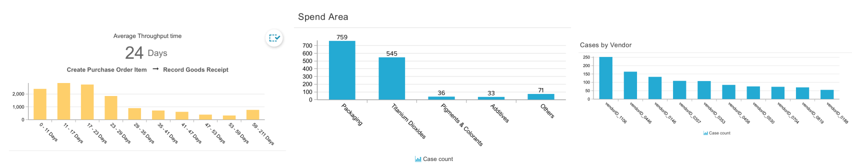


Fig. 25. Throughput between Create PO Item (first execution) and Record Goods Receipt (last execution), and the spend areas and vendors with especially high throughput (40 days or more)

Order Confirmation and other Insights Other interesting information can be found in fig. 26. In the first sub-figure we see those resources performing creations of PO items which later are deleted. In the second figure we see those vendors which most often perform order confirmation (in %). We see that there vendors which always confirm orders in the consignment sublog. Last but not least we want to look into those cases which stop after the execution of Create PO Item. Vendors which are most involved in those PO items can be found in the third graph of fig. 26.

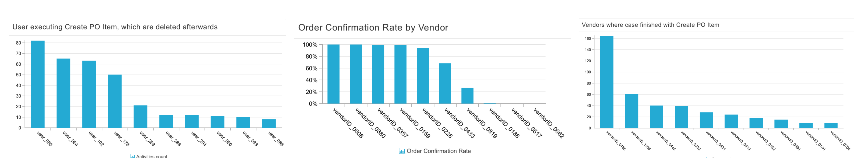
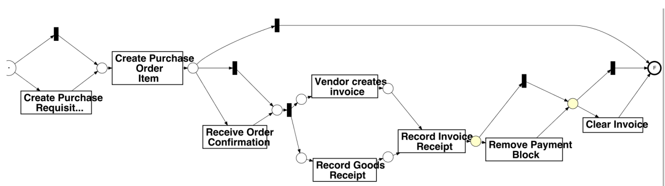


Fig. 26. a) user executing Create PO Items which are later deleted, b) Order Confirmation Rate by Vendor, c) vendors where a case finishes with PO Item Creation

2.5 Model D: 3-way matching, invoice after goods receipt

For those items we report goods receipt first and then an invoice receipt [1]. Invoices are only cleared if the values matched. IV = true and GR = true. Let us have a look at the mined model.

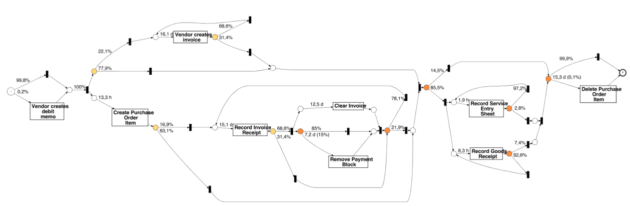
The Model Due to large variety, it is not easy to mine a proper model representing all traces with a high fitness. For simplicity reasons we filter the log for change actvities, since those are analysed individually in section 4. We additionally perform another log split into service orders and non service orders indicated by the occurance of the "record service entry sheet" event. The two respective models can be found in fig. 27 and fig. 28. The model for non-service goods represents the typical P2P process. Again we see the derivation that a case can end after the creation of an PO Item.



Avg activity precision	100%
# Moves Observed	129.117
# Moves Possible	129.144
Avg fitness	92,1%
% Violations	11,1%
% Event Violations	11,1%
% Data Violations	0%
# Correct Events	43.305
# Wrong Events	5.010

Fig. 27. Model D1: 3-way matching, invoice after goods receipt. Non-Service-PO.

In the second model (fig. 28) 0,2% of all cases start with a debit memo created by the vendor. The other 99,8% skip that step. Concurrently, the vendor can create several invoices while the company creates the purchase order item. After the item is created we can optionally record an invoice receipt which (with an optional payment block removal) is then cleared. After the concurrency optionally we can receive one or more Service Entry Sheets as well as goods. In the end 0,1 % of the PO items is deleted. Note, that the second model got a lower fitness than usual (88%). The precision is only 45,5 %, meaning that it allows for quit a lot more behavior than actually observed in the log. Reason for that is probably the large amount of loops in the model.



Avg activity precision	45,5%
# Moves Observed	613.415
# Moves Possible	1.348.865
Avg fitness	88,7%
% Violations	6,6%
% Event Violations	6,6%
% Data Violations	0%
# Correct Events	234.305
# Wrong Events	16.019

Fig. 28. Model D2: 3-way matching, invoice after goods receipt. Service PO.

Derivations For model D1 we spot quite a lot derivations after the creation of the purchase order. Those are cases where Goods Receipt were missing at this point in the alignment. For model D2 it was not possible to visualize the derivations in Prom 6.

Throughput Time The throughput time of model D1 is visible in fig. 29. On average takes 36 days until an invoice is received and another 31 days until the invoice is then cleared.

The throughput time of model D2 is visible in fig. 28. The waiting time for record invoice receipt is on average 15 days. After that it takes on average 13 days to clear that invoice. If a payment block needs to be removed this takes on average 7 days. The waiting time for record service entry sheet and record goods receipt is surprisingly small. If a PO item is deleted that takes on average 15 days.

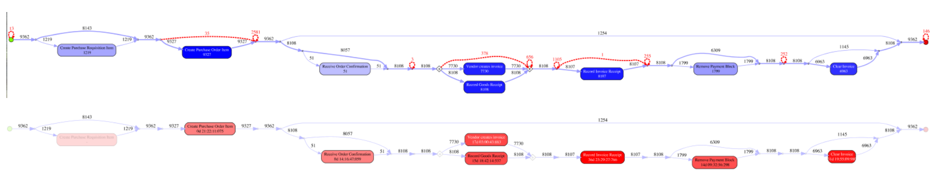


Fig. 29. Case Count and Throughput of Model D.

Order Confirmation Rate Interesting is the order confirmation rate though. Only vendor 433 confirms orders (ca 70% of them in this sublog). All other don't.

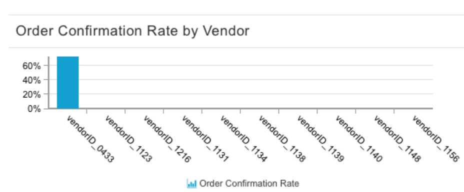


Fig. 30. Order Confirmation Rate of sublog D ordered descending per vendor. Only a subset of all vendors is visible.

2.6 Model E: Supplier Relation Management

We filtered the log so that it only contains traces with SRM Events, within the selected time frame. This log contains 21564 Events with 1425 Cases and 373

Variants of which 2/3 are unique traces. Taking a closer look at the log we see that the attribute `Document Type` has only the "EC Purchase order" value. Further intresting attributes are `Spend classification text` with values "NPR" , "OTHER" and marginally "PR", `Spend area text` with values "Enterprise Services", "CAPEX & SOCS", "Marketing", "Workforce Services", "Others", "Logistics", "Sales", "Real Estate" and "Trading & End Products" in descending case count, `Item Type` which has the values "Service" and "Standard". There is only one `Case-Company` involved, "company_0000". The `Item Category` attribute has "3 way match, invoice after GR" and "3 way match, invoice before GR" values.

We used the inductive miner plug-in from ProM to mine a model for the SRM containing traces which is shown in fig. 31. We used about 89.2% of paths and 100% of activities. This setting assured that we only get one possible loop for the "Record Service Entry Sheet" to maximize fitness and precision of the model. In fig. 31 we present the resulting values regarding fitness and precision of the model which were obtained using the "Multi-Porcess Explorer" ProM plug-in. We observe that there are some places which where marked by the Multi-Process Explorer such as the place after "SRM: Create" or "SRM: Document Complete". While the first one is rather light the latter one is clearly darker which indicates that the precision in this places is between 40% and 60% and therefore lower.

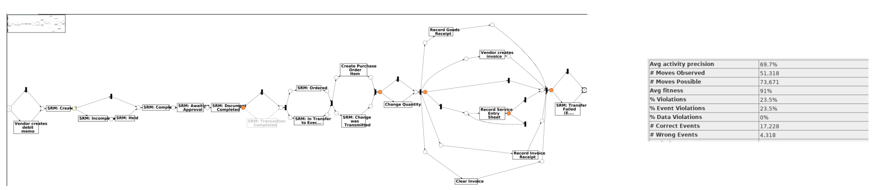


Fig. 31. Spend Classification and Area.

3 Automation Rate

Another KPI we will regard is the automation rate. We use Celonis and the integrated PQL interpreter to realize that KPI. The definition of the KPI can be found in fig. 32. We count the number of events having a user attribute including "batch", denoting an automatically executed event and then divide it by the overall number of events.

```
SUM(CASE WHEN "log.xes"."User" LIKE 'batch' THEN 1.0 ELSE 0.0 END) /
COUNT("log.xes"."Event-Id")
```

Fig. 32. PQL for Automation Rate.

In fig. 33 we can spot the automation rate per activity in the model. We spot at first sight that many activities are always executed manually: All change activities, Clearing Invoices, Block/Delete/Reactvitate PO Items, setting of payment blocks, record of a service entry sheet as well as the creation of debit memos and invoices by the vendors.

On the other hand we also spot a lot evens which are highly automated: Especially most SRM events, except from "SRM: transaction completed" and "SRM: Transfer Failed" which both are always executed manually, are highly automated. Other activities are medium to low automated. When we now also take the number of executions into account we realize that especially the already partly automated activities "Record Goods Receipt", "Create Purchase Order Item", and "Record Invoice Receipt" could be improved a lot. If possible, automating clearing invoices, recording service entry sheets and the creation of the invoice by the vendor are executed often, though never automatically. In fig. 33 we see the automation of the cases performed by the top 30 vendors in terms of number of items. We see that for the top vendor automation rate is 4% only, while there are also vendors which gain 20% and more. Especially, those vendors which we deal with often offer potential for automation by introducing electronic invoicing.

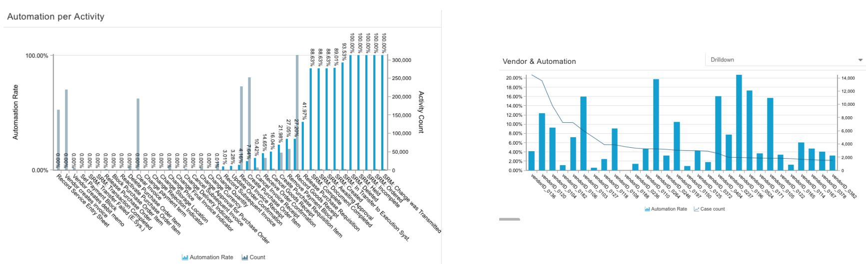


Fig. 33. Automation Rate by activity (left) and Automation Rate by vendor (top 30 vendors) (right)

In fig. 34 we see the automation rate per category. Lowest automation rate is visible in 2-way matches as well as in Consignment cases. three way matches where invoices are received after goods is much more automated than a three way match where the invoices arrives first.

4 Rework Rate

Now, we also want to check the rework rate of the cases. We can define rework in two different ways:

1. Rework as execution of a change activity, like "Change Price"
2. Rework as multiple executions of certain activties within a trace in a trace.

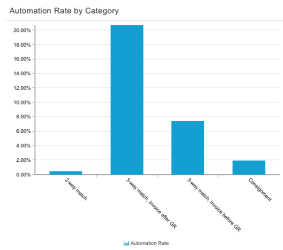


Fig. 34. Automation Rate per category

4.1 Change Activities

We will only regard the first approach in the following. Again we define a KPI for that rework via PQL in Celonis. The query can be found in fig. 35. We compute the relative proportion of all cases containing at least one of the defined rework activities.

```
AVG(CASE WHEN MATCH_ACTIVITIES(NODE_ANY[<%=rework_activities%>] ) = 1 THEN
1.0 ELSE 0.0 END)
```
with Rework Activities:
```
<%=rework_activities%> = 'Change Quantity', 'Change Price', 'Change
Approval for Purchase Order', 'Change Delivery Indicator', 'Change
Storage Location', 'Change Currency', 'Change payment term', 'Change
Rejection Indicator', 'Change Final Invoice Indicator'
```

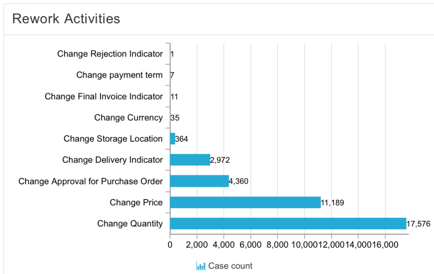
Fig. 35. PQL for Rework Rate.

In fig. 36 we see the number of times each rework event was executed in the filtered log. We observe that especially Change Quantity / Price Approval for Purchase Order / Change Delivery Indicator happen often.

In fig. 36 we also see in which spend area a lot of rework is performed and in which less. Especially "Energy", and "Real Estate" always contain at least one rework activity. These are by the way exactly those fields which exclusively are handled by CompanyID_003. When we dig deeper into that we see that CompanyID_003 in general got a rework rate of 100%. Change Activities are always either "Change Price" (10 times) or more often "Change Approval for PO" (1027). Visible is the low rework rate for large areas like Packing and Sales. Critical is the rework rate of 33% in "Trading and Endproducts" while the the number of cases is large (22204 cases).

Let us now look at the rework rate per vendor. In fig. 37 we see the top 25 vendors in terms of net value taken from the purchase order creation (left) and case count (right). Critical vendors are marked in orange and red.

In figure fig. 38 we now spot the rework rate per category.



Spend & Rework

Spend	net worth (EUR)	Rework Rate	Case Count
Energy	8,457,639	100.00%	15
Real Estate	36,548,592	100.00%	570
Workforce Services	718,095,070	58.82%	128
	133,586,716	39.50%	3,286
Trading & End Pr...	648,687,907	31.59%	22,204
Commodity Resins	40,228,311	29.77%	408
Spend Area Unid...	1,880,980	28.21%	36
Latex & Monomers	816,112,131	24.35%	5,005
Titanium Dioxides	503,060,138	23.38%	1,861
Pigments & Color...	150,738,929	22.34%	2,593
Specialty Resins	221,953,900	21.08%	2,406
Marketing	1,187,592,858	19.30%	1,782
Solvents	172,033,295	17.78%	2,627
Others	310,209,622	17.49%	2,625
Additives	836,984,963	17.32%	18,311
Enterprise Services	114,979,783	16.14%	661
Packaging	973,700,087	14.68%	109,181
Sales	528,810,347	10.13%	65,790
CAPEX & SOCS	142,733,331	8.78%	6,760
Logistics	12,050,993,715	4.85%	5,219
Chemicals & Inter...	2,545	0.00%	1

Fig. 36. Rework Activities (left) and Rework Rate by spend area (right)

Vendor & Rework

Vendor	net wort...	Rework Rate	Case Count
vendorID_0330	1,485,128,925	0.00%	215
vendorID_0396	1,361,026,196	42.64%	12
vendorID_0234	1,293,505,477	0.00%	214
vendorID_0472	916,422,763	0.00%	244
vendorID_0388	845,349,281	81.57%	29
vendorID_0020	568,095,267	82.91%	3
vendorID_0509	524,239,707	0.00%	2
vendorID_0395	392,828,940	0.00%	1
vendorID_0040	366,092,634	77.20%	19
vendorID_0877	354,796,759	91.10%	19
vendorID_0534	291,337,449	0.00%	50
vendorID_0397	288,386,436	0.00%	1
vendorID_0204	282,332,447	100.00%	1
vendorID_0213	270,726,869	100.00%	8
vendorID_0885	269,974,589	99.69%	28
vendorID_0392	236,944,440	0.00%	1
vendorID_0235	222,629,034	2.55%	255
vendorID_0104	215,215,080	5.38%	9,817
vendorID_0473	208,975,438	0.73%	220
vendorID_0538	207,522,312	0.00%	101
vendorID_0159	199,816,293	22.47%	968
vendorID_0910	192,904,542	12.50%	5
vendorID_0854	189,413,350	100.00%	19
vendorID_0017	185,042,166	52.55%	11
vendorID_0034	181,215,801	0.00%	1

Vendor & Rework

Vendor	net worth (E...	Rework Rate	Case Count
vendorID_0136	42,786,925	13.92%	14,470
vendorID_0120	39,633,153	8.47%	13,564
vendorID_0104	215,215,080	5.38%	9,817
vendorID_0182	20,664,506	11.70%	7,250
vendorID_0106	168,895,942	13.83%	7,231
vendorID_0103	14,127,205	10.12%	6,014
vendorID_0127	11,486,404	10.05%	4,980
vendorID_0118	2,972,618	7.17%	3,911
vendorID_0108	12,094,362	2.52%	3,891
vendorID_0188	552,062	13.40%	3,592
vendorID_0236	11,628,588	26.19%	3,374
vendorID_0110	11,035,874	3.07%	3,272
vendorID_0264	5,341,489	13.24%	3,264
vendorID_0246	8,550,720	6.26%	3,182
vendorID_0197	60,351,239	88.97%	3,052
vendorID_0150	3,854,366	3.49%	3,008
vendorID_0125	11,227,841	10.96%	2,989
vendorID_0372	4,940,689	4.30%	2,914
vendorID_0404	36,137,610	8.07%	2,550
vendorID_0237	8,826,199	6.00%	1,982
vendorID_0196	9,159,680	7.54%	1,958
vendorID_0524	20,145,407	10.18%	1,910
vendorID_0171	49,893,402	15.28%	1,868
vendorID_0105	8,347,992	2.30%	1,846
vendorID_0122	14,964,147	5.06%	1,842

Fig. 37. Rework top 25 vendors (value) (left) and Rework top 25 (case count) (right)

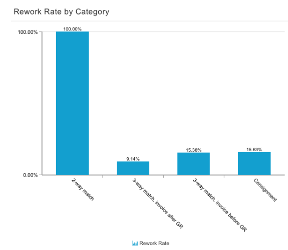


Fig. 38. Rework Rate per category

Notable is the 100% rework rate for 2 way matches. Reason for that is the obligatory execution of "Change Approval for Purchase Order Item" as visible in fig. 13. In three way matches where invoices arrived before goods we observe a slightly higher rework rate than in the case where goods arrived before the invoice.

5 Social Mining

First we will present general social facts about the log. The log contains two kinds of resources: users and batch. The difference between these types of resources is, as described in the challenge description, that "batch" resources are automatically executed while "user" resources are actually manual resources. As we can't really distinguish between the 20 batch users we decided that we aggregate all batch resources as one "batch" resource. After this manipulation we got 608 resources. Using the PM4PY modules we discovered that there are events where no resource was recorded, these are ['Vendor creates invoice', 'Record Service Entry Sheet', 'Vendor creates debit memo', 'Clear Invoice']. Batch resources execute the following tasks: ['SRM: Created', 'SRM: Complete', 'SRM: Awaiting Approval', 'SRM: Document Completed', 'SRM: Ordered', 'SRM: In Transfer to Execution Syst.', 'SRM: Change was Transmitted', 'Record Invoice Receipt', 'Create Purchase Order Item', 'Create Purchase Requisition Item', 'Remove Payment Block', 'Receive Order Confirmation', 'Update Order Confirmation', 'Record Goods Receipt', 'Record Subsequent Invoice', 'Cancel Invoice Receipt', 'SRM: Deleted', 'Cancel Goods Receipt', 'SRM: Incomplete', 'SRM: Held', 'Change Quantity', 'Release Purchase Requisition']. We have 288 users that only execute one task, with 123 users doing only 'Create Purchase Requisition Item'. In the next step we decided to filter out these 288 users to be able to progress with our Social Network Analysis, as the overall resource count was to high to be analysed properly. Additionally, we found that in 87355 traces the respective resource only executed the activities ['Record Goods Receipt', 'Cancel Goods Receipt'], not necessarily one time only but for sure only these two activities. Therefore, in order to be able to cluster resources by their activities, we renamed these resources to "rcgr_users". One thing that we noticed throughout the different social analysis is that 'user_297' always appears to be isolated. Looking closely this resource only performs two tasks 'Create Purchase Order Item' and 'Delete Purchase Order Item' on one specific day (31.01.2018). The 'Create Purchase Order Item' is performed four times consecutively with four different values and 'company_0000'. 2 minutes later 'Delete Purchase Order Item' is performed four times with the same values and company. 'user_013' performs the most activities as can be seen in the following figures.

6 Summary

In this report we analyzed the given log first in general and saw anomalies in company 1 and 2 which each only execute one PO Item in 2018 which was both deleted. We encountered a highly complex process with many variations. By splitting it according to the categories "2 way matches", "Consignments", "3way matches, invoice after goods" and "3 way matches, invoices before goods" we could get better insights. Then, models for each of those sublogs were created, each with sufficiently high fitness. Via a throughput analysis we found out that the time between reporting of the invoice receipt and the clearance is high, same holds for the time between record goods receipt and record invoice receipt. On some sublogs we computed the order confirmation rate. Other KPIs we observed were the automation rate where we spotted which activities are well automated and which offer potential for automation. Last but not least we looked into the rework rate and saw that especially often prices and quality changes were executed. For all KPIs we analyzed what vendor or resource was particularly often involved in not well working cases, so that improvements can be executed based on this analysis. For instance did we observe a high amount of one-time-vendors.

References

1. Bpi challenge 2019. https://icpmconference.org/icpm-2019/contests-challenges/bpi-challenge-2019/. Accessed: 2019-05-24.
2. An introduction to purchase to pay process flow. https://www.purchasecontrol.com/uk/blog/procure-to-pay-cycle/. Accessed: 2019-05-24.
3. van dongen, b.f., dataset bpi challenge 2019. 4tu.centre for research data. https://doi.org/10.4121/uuid:d06aff4b-79f0-45e6-8ec8-e19730c248f1.

Business Process Intelligence Challenge 2019: Process discovery and deviation analysis of purchase order handling process

Jongchan Kim, Jonghyeon Ko, and Suhwan Lee

Ulsan National Institute of Science and Technology, 50 UNIST-gil, Ulsan 44919, Republic of Korea
{jckim,whd1gus2,ghksdl6025}@unist.ac.kr

Abstract. Process mining is of paramount importance as the number of event logs cumulated in information systems is growing exponentially over time. Increased number of event logs implies that the amount of information that can be mined to be used to meet the stakeholders requirements through the generation of qualitative and quantitative outputs enhances. In compliance with the goal of the Business Process Intelligence Challenge to make the result of the analysis useful in the real world, we take a real-life event log of a company operating in the Netherlands in order to analyze and suggest improvements to the companys purchase order handling process for its subsidiaries. In this paper, we discover and refine business process models that encompass the four types of flows in the event log. Then, we also identify different types of deviation in the event log from the discovered processes and their extents based on alignment techniques and various statistical tests.

Keywords: Business Process Intelligence · Process mining · Business Process Model · Deviation analysis.

1 Introduction

Process mining has gained increasing traction in industry as a set of efficient techniques analyzing event logs to enhance business processes. Likewise, it is important to provide practical implications through the application of process mining techniques, which also aligns with the goal of Business Process Intelligence Challenge 2019 [1]. We were provided with the real-world problems that a company is facing with, and thereby, we strived to provide fruitful answers and explanations in detail for each question provided. Following the guidelines provided in the Process Mining Manifesto [2], our analysis concerns different phases of the BPM lifecycle [3].

1. Process identification : Before entering the deep-dive analysis step, the first and basic step is to acquire the knowledge drawn from the given event log, including where the event log is generated and clarifying what we want to discover with the event log. In this case, the event log is expected to be generated from the procurement system of a painting company.

2. Process discovery : The second step is to extract a business process model from an event log using process mining tools (e.g. *petri net*) based on identified process. It is essential to filter out deviations from the dataset and keep regular process flows unstained based on the obtained business process model reasonably fitted to the event log.

3. Process analysis : For the third step, discovered process is analyzed in different levels such as activity, timestamp, and other features. We, playing the role of consultants, diagnose the current situation of the system driven from the business process model and offer the reasons of deviations explaining where and when the discrepancy between deviants and regular model appeared in detail.

Following the three phases as in above, we mainly focus on 1) discovering the process model, and 2) analyzing the deviations in duration (between activities/instances), activities and other features as to answer for three questions provided in BPIC 2019.

The rest of the report is designed as follows. Section 2 provides comprehensive analysis of the event log. In section 3, 4, and 5 we focus on answering the three questions. Specifically in Section 3, a BPMN model that is constructed to encompass various different models in the process is described. In Section 4, throughput of invoicing processes are calculated, such as the duration between different activities or instances. Moreover, deviations in durations are identified and various factors affecting these deviations are also discussed. In Section 5, various types of deviations are analyzed using an alignment technique, and multiple statistical tests. Finally in Section 6, the overall summary of the analysis along with multiple minor and major findings are presented.

2 Overall Analysis

2.1 Understanding of the background of the company

Prior to analyzing an event log, it is extremely important to understand the general situation and the process of the given data by gathering information and integrating background knowledge. The given event log contains information about how the company and its subsidiaries handle purchase orders with vendors. General flows are best characterized by activities which handle purchase order, goods receipt and invoice, and can be described as follows. Firstly, one of the subsidiaries creates purchase order document and send it to vendors. After the ordered items are delivered, the company records goods receipt to confirm whether the amount and types of items are correctly delivered according to what had been ordered. For the last step, the company registers the invoice and clears it.

2.2 Data selection and preprocessing

We were provided with two files, where the first file is in xes format (*log_IEEE.xes*) and the other file is in csv format (*BPI_Challenge_2019.csv*). The xes file is an original event log and the csv file is a converted version of the xes file. Between these two files, however, it was revealed that there is a difference in timestamps. Although the number of cases, events, and activities are identical in both files, mean and median values of case duration are not identical as the starting and ending time of the event log are not equal. The starting and ending time of the xes file are 27.01.1948 07:59:00, 10.04.2020 06:59:00, respectively, whereas in the csv file, the starting and ending time are 26.01.1948 23:59:00 and 09.04.2020 23:59:00, respectively. Since not only the factors affecting the difference in the timestamp between two files cannot be identified but also the difference in timestamps was not immense, we decided to use the xes file.

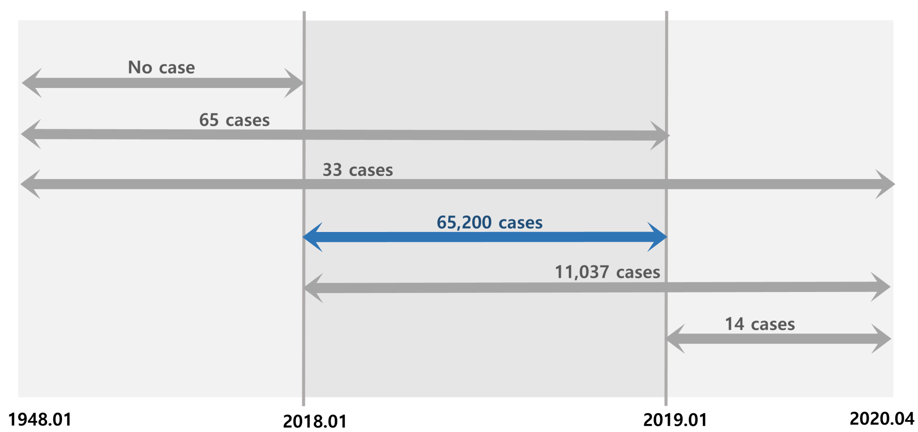


Fig. 1. Range of the timestamp in xes file

As mentioned earlier, the dataset has a wide range of timestamp, dating from Jan 27, 1948 to Apr 10, 2020, while the data should have been collected from events that happened in 2018 only. We figured out that the dataset can be classified into three types as in Fig. 1. The first type both starts and ends in 2018, the second type either starts or ends in 2018, and the third type neither starts nor ends in 2018. The distribution of timestamp shows that approximately 87% of all cases both started and ended in 2018, which belong to the first type. Hence, we argue that the second and the third type of the dataset are erroneous as these are out of the scope of the analysis. One of the reason the erroneous cases are introduced can be a system error as these cases are detected with an activity named, SRM: Transfer Failed (E. Sys.). Therefore, we regard these cases as outliers and we do not further consider these. To sum up, the filtered data contains 65,200 cases with 1,315,795 events and 41 activities.

2.3 Data exploration

Table 1 provides the description of variables of the dataset that are used in the analysis. As it describes, *case concept:name* corresponds to *Case ID*, *event concept:name* corresponds to *Activity*, *event time:timestamp* corresponds to *Timestamp*, and *event User* (or *event org:resource*) corresponds to *Resource*. Since the name of each variable is quite long, we changed the name of each variable by deleting the words, "case" and "event", located at the beginning of the name of each variable.

Table 1. Variable description

Variable	Definition	Data type
eventID	ID of an event	Categorical
case Spend area text	Upper-level classification of an item	Categorical
case Company	ID of a company	Categorical
case Document Type	The type of a flow of an item	Categorical
case Sub spend area text	Lower-level classification of an item	Categorical
case Item Type	The type of an item	Categorical
case Item Category	The category of an item	Categorical
case GR-Based Inv. Verif.	Binary indicator signifying whether GR-Based invoicing is necessary	Binary
case concept:name	Case ID	Categorical
case Goods Receipt	Binary indicator signifying whether 3-way match is necessary	Binary
event org:resource	Resource ID	Categorical
event concept:name	Activity	Categorical
event Cumulative net worth (EUR)	Cumulative net worth in Euro	Numeric
event time:timestamp	Timestamp	Date

Table 2. Number of cases by Company ID and Item Category

Item Category	CompanyID_0000	CompanyID_0001	CompanyID_0002	CompanyID_0003
3-way match, before GR	191,135	0	2	0
3-way match, after GR	13,205	2	0	0
2-way match	0	0	0	640
Consignment	13,837	0	0	0

The event log contains information about transactions between subsidiaries and vendors, and the number of items purchased by different subsidiaries are presented in Table 2. The Table 3 shows the overall data structure and the number of items that fall into each distinctive category defined by *Item Category*, *Document Type*, and *Item Type*. In addition to these three features, *Spend area text* and *Sub spend area text* are also used to further specify the distinctive category of an item. To summarize, each item can be defined by five features, *Item Category*, *Document Type*, *Item Type*, *Spend area text*, and *Sub spend area text*, and we will use this way of defining an item based on five features throughout the remaining part of the report (see Fig. 2).

Table 3. Data Structure

Item Category	Document Type	Item Type	Frequency
3-way match, after GR (mean duration= 75 days)	Standard PO	Service	4,547
		Standard	7,042
		Subcontracting	582
		Third-party	323
	EC purchase order	Service	341
		Standard	108
	Framework order	Service	264
3-way match, before GR (mean duration = 74.5 days)	Standard PO	Standard	182,160
		Subcontracting	3,664
		Third-party	4,598
	EC purchase order	Standard	714
	Framework order	Standard	1
2-way match (mean duration = 57.5 days)	Framework order	Limit	640
Consignment (mean duration = 24.1 days)	Standard PO	Consignment	13,837

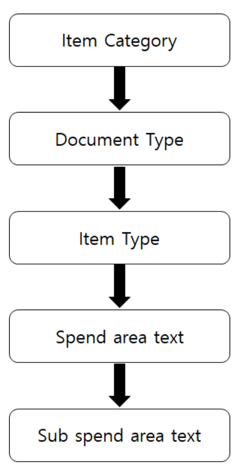


Fig. 2. Level of the analysis

2.4 Four types of flows

The event log contains four types of flows that can be defined by different combinations of three features, *case Goods Receipt*, *case GR-Based Inv. Verif*, and *Item Type*.

(1) 3-way matching, invoice after goods receipt: One of subsidiaries submits a purchase order for a specific item to a vendor. Then, the vendor will create an invoice and send the ordered item to the company. After receiving the item, the subsidiary will check the list of materials with the purchase order document and the invoice receipt. The company can revise the price or quantity of the item before receiving the ordered item. For the final step, the company will completely clear this order, if there is no problem on the invoicing process.
(2) 3-way matching, invoice before goods receipt: This flow is similar to *3-way matching, invoice after goods receipt*, except for the fact that invoices can be received before receiving goods receipts. As in *3-way matching, invoice after goods receipt*, the subsidiary will completely clear the order if there is no problem on this invoicing process.
(3) 2-way matching: In this process, goods receipt message is not required as all orders made in this flow are framework orders.
(4) Consignment: As the name of the flow implies, invoice messages are not required since it is handled in a separate process operated by the third-party company. Therefore, messages on goods receipt can only be observed in this flow.

As the name of each type of flow is long, we decided to change the name of each type of flow by deleting the word "invoice" and changing the word "matching" with "match" from each name of the type of flow for the remaining part of the report.

Table 4. Different *Goods Receipt & GR-based Inv. Verif* pair for four types of flows

Item Category	case Goods Receipt	case GR-Based Inv. Verif.	Item Type
3-way match, after GR	TRUE	TRUE	Various
3-way match, before GR	TRUE	FALSE	Various
2-way match	FALSE	FALSE	Various
Consignment	TRUE	FALSE	Consignment

3 Building a business process model

Question 1. Is there a collection of process models which together properly describe the process in this data. Based on the four categories above, at least 4 models are needed, but any collection of models that together explain the process well is appreciated. Preferably, the decision which model explains which purchase item best is based on properties of the item.

The BPMN model was derived from *petri net* constructed using ProM as in Fig. 3. Since the BPMN model is very complicated as it encompasses all four types of flows, subprocesses are introduced as in Fig. 4 to enhance the understandability by making the BPMN model simple and neat. Fig. 5 (a) is a subprocess called *SRM* that can be found in *3-way match, before GR*. Fig. 5 (b) is a subprocess called *SRM(Transfer followed by order)* that can be found in *3-way match, after GR*. Fig. 5 (c) is a subprocess called *Vendor Creates Invoice & Record Goods Receipt (Goods)* that can be found in more than two types of flow, and this subprocess deals with goods. Fig. 5 (d) is a subprocess called *Vendor Creates Invoice & Record Goods Receipt (Service)* that can be found in more than two types of flow, and this subprocess deals with services. Fig. 5 (e) is a subprocess called *Record Goods & Invoice receipt*that can be found in more than two types of flow. In Fig. 5 (c) to Fig. 5 (e), "&" means that any of two activities can have chance to happen earlier than the other activity.

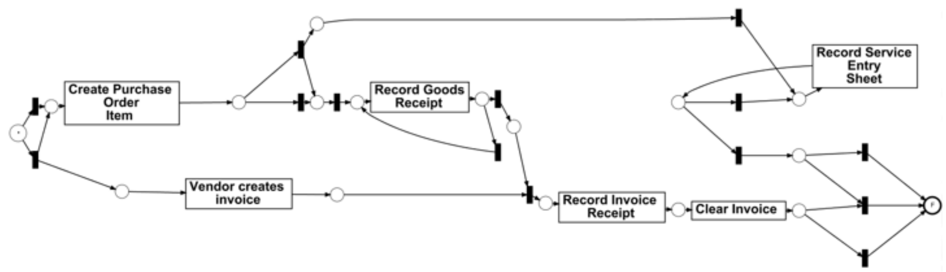


Fig. 3. An example of a petri net

4 Throughput of the invoicing process

Question 2. What is the throughput of the invoicing process, i.e. the time between goods receipt, invoice receipt and payment (clear invoice)? To answer this, a technique is sought to match these events within a line item, i.e. if there are multiple goods receipt messages and multiple invoices within a line item, how are they related and which belong together?

Each item contains one or more events with the information about its case duration. In a business process, case duration which means the order leading time is regarded as one of the most important factors from the perspective of a customer. Based on the business process model developed in Fig. 4, we investigate distributions of case duration for each type of 13 business process models.

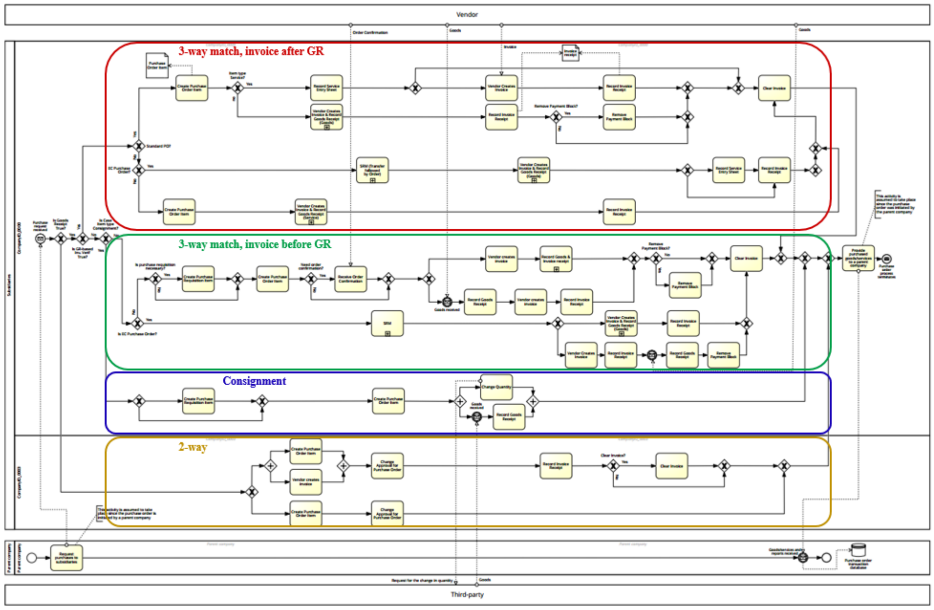


Fig. 4. A BPMN model

The analysis was conducted based on two approaches, *activity level approach* and *instance level approach*.

4.1 Activity level approach

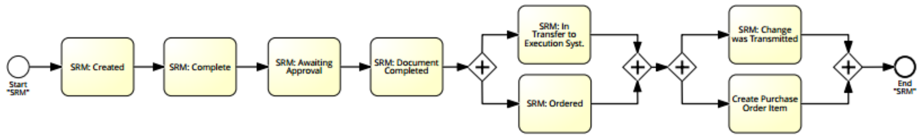
For each of 13 business process models which consists of different *Document Type* (D.T.) and *Item Type* (I.T.), we investigated durations for each activity. In order to exhibit the outcomes in a more efficient way, proportion of duration for each activity was calculated as in Fig. 6 and we focused on activities whose waiting time is longer than 30 percent of the total duration of an item.

(1) Duration in 3-way match, after GR (D.T. = EC purchase order, I.T. = service)
- from "Record Invoice Receipt" to "Clear Invoice" = 37%

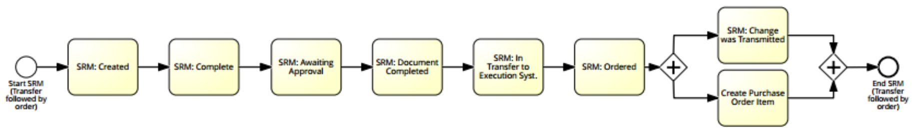
(2) Duration in 3-way match, after GR (D.T. = EC purchase order, I.T = standard)
- from "Record Goods Receipt" to "Record Invoice Receipt" = 43%
- from "Record Invoice Receipt" to "Clear Invoice" = 45%

(3) Duration in 3-way match, after GR (D.T. = Standard PO, I.T. = service)
- from "Vendor creates invoice" to "Record Invoice Receipt" = 33%
- from "Record Invoice Receipt" to "Clear Invoice" = 41%

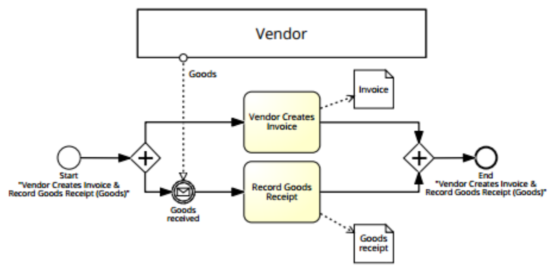
(4) Duration in 3-way match, after GR (D.T.= Standard PO, I.T. = Subcontracting)



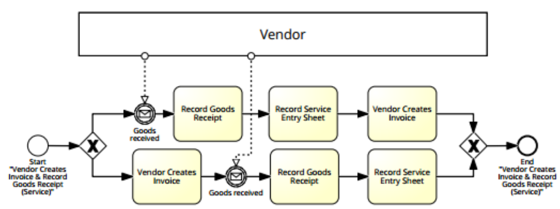
(a) SRM



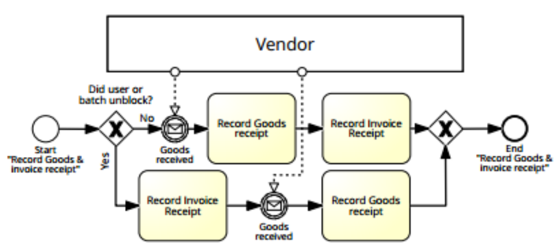
(b) SRM (Transfer followed by order)



(c) Vendor Creates Invoice & Record Goods Receipt (Goods)



(d) Vendor Creates Invoice & Record Goods Receipt (Service)



(e) Record Goods & Invoice receipt

Fig. 5. Subprocesses of BPMN model

- from "Record Goods Receipt" to "Record Invoice Receipt" = 100%

(5) Duration in 3-way match, after GR (D.T. = Standard PO, I.T. = Third-party)

- from "Vendor creates invoice" to "Record Invoice Receipt" = 31%
- from "Record Invoice Receipt" to "Clear Invoice" = 37%

(6) Duration in 3-way match, after GR (D.T. = Framework order, I.T. = Service)

- from "Create Purchase Order Item" to "Vendor creates invoice" = 37%

(7) Duration in 2-way match (D.T. = Framework Order, I.T. = Limit)

- from "Vendor creates invoice" to to "Create Purchase Order Item" = 65%

(8) Duration in Consignment (D.T. = Standard PO, I.T. = Consignment)

- from "Create Purchase Order Item" to "Record Goods Receipt" = 78%

4.2 Instance level approach

In this part, we assumed that some of distributions of case duration for each of 13 discovered process models is either similar or different according to its types such as *Document Type* (D.T.), *Item Type* (I.T.), and *Sub spend area text*, etc. In order to investigate this, some of the models were grouped.
Furthermore, we investigated some items with long order leading time, and tried to find factors in attribute level that affects the order leading time. For this, we implemented the regression analysis and extracted factors that are statistically significant. Our regression models were properly fitted under significant level=0.05 for all process models. Interestingly, we found 672 significant factor levels in 13 regression models and the summary is provided in Table 5-10. We showed maximum five factors for each attribute because of the limitation of the space of the table.

Note: For *Cumulative net worth*, we discovered that some of line items have multiple goods receipts. In addition, the total *Cumulative net worth* of one line item is calculated by summation of the *Cumulative net worth* in multiple good receipts.

(1) 3-way match ***after vs before*** goods receipt (D.T.= EC purchase order, I.T.= Standard)

Comparison: In terms of 3-way match with D.T.= EC purchase order and I.T.= Standard, *3-way match, after GR* has more cases with longer order leading time than *3-way match, before GR* as in Fig. 7. Additionally, *3-way match, before GR* has more transactions with vendors in different *sub spend areas* and it triggers deviation.

- 3-way match, after GR: Transaction with 5 vendors in 4 types of *Sub spend area*

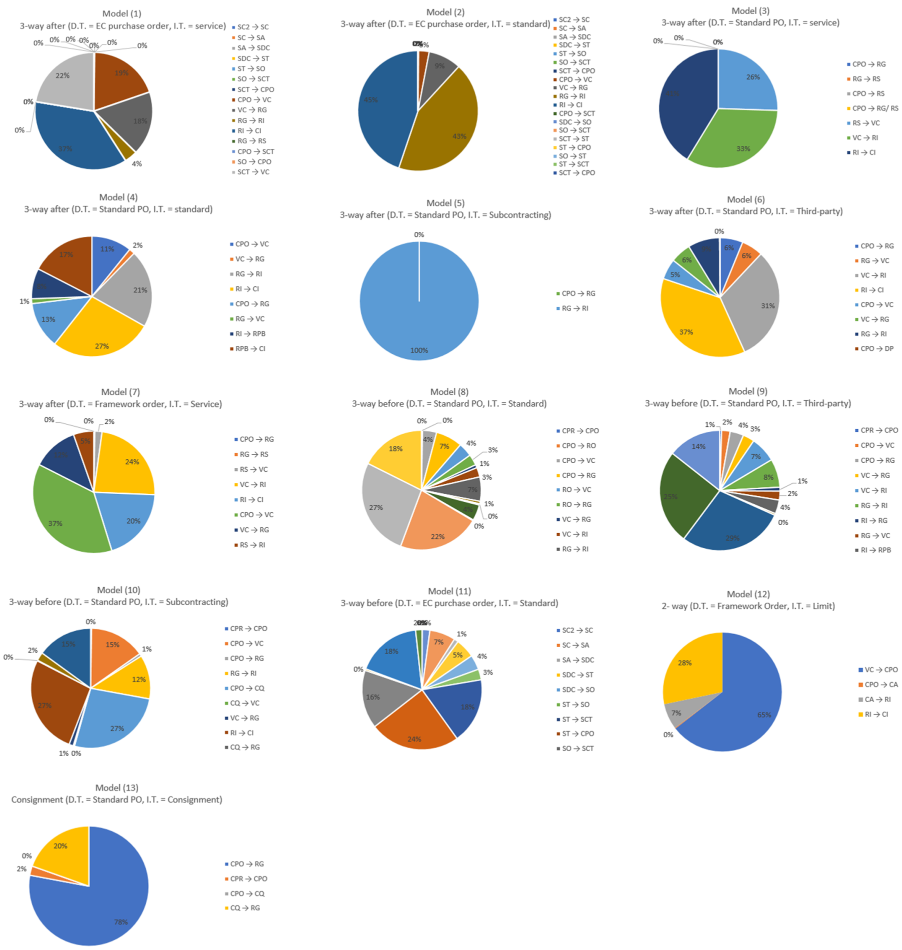


Fig. 6. Pie charts about proportion of duration for each activity

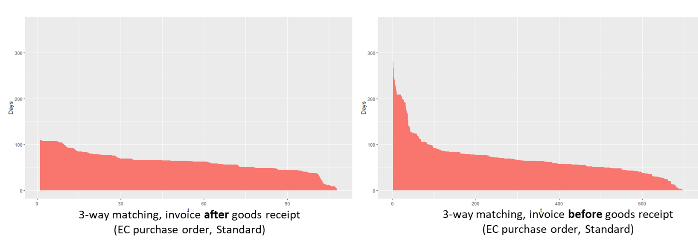


Fig. 7. Distribution of duration per each instance - part (1)

- 3-way match, before GR: Transaction with 70 vendors in 23 types of *Sub spend area*

Regression for each model: As a result of linear regression, Table 5 shows a list of significant factors that affect case duration. According to Table 5, the duration of the instance is decreased by 14.3288 days on average in case of "Laboratory Supplies Services" factor.

- 3-way match, after GR: For the decreasing effect on case duration, one factor named "Laboratory Supplies & Services" in *Sub spend area text* and one vendor named vendorID_0029 were found to be significant.
- 3-way match, before GR: For the increasing effect on case duration, 3 factors were observed in *Sub spend area text* and 5 factors were observed in *Vendor* including vendorID_0004.

Table 5. Regression analysis: the list of significant factors in part (1)

Attibutes	Factors	Estimate	P-value
3-way match, after GR			
-Sub spend area text	Laboratory Supplies & Services	-14.3288	0.0130
-Vendor	vendorID_0029	-17.3758	0.0000
3-way match, before GR			
-Sub spend area text	Facility Management	80.4243	0.0341
	Laboratory Supplies & Services	128.6787	0.0138
	Other Logistics Services	102.7780	0.0126
-Vendor	vendorID_0004	89.2219	0.0365
	vendorID_0024	-96.2948	0.0173
	vendorID_0027	-38.2701	0.0424
	vendorID_0042	-81.3129	0.0110
	vendorID_0045	-91.9423	0.0283

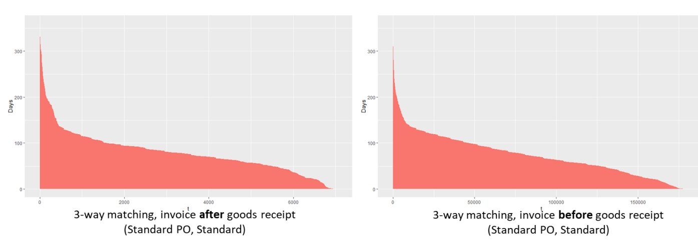


Fig. 8. Distribution of duration per each instance - part (2)

Table 6. Regression analysis: the list of significant factors in part (2)

Attibutes	Factors	Estimate	P-value
3-way match, after GR			
Sub spend area text	Business Gifts&Promotional Items	98.0376	0.0172
	Chloride	-59.9164	0.0090
	Commercial Printing	123.2229	0.0026
	Commodity Resins Precursors	-60.1448	0.0010
	HR Services	198.7098	0.0000
Vendor	vendorID_0106	-43.4573	0.0351
	vendorID_0120	-37.1949	0.0238
	vendorID_0123	155.1773	0.0036
	vendorID_0125	-60.5384	0.0004
	vendorID_0136	-42.6069	0.0104
Cumulative net worth	net worth (unit=100 EUR)	0.0097	0.0000
3-way match, before GR			
Sub spend area text	Additives - Other	-30.5778	0.0037
	Adhesion Promotors	-31.7756	0.0059
	Advertising	93.4852	0.0057
	Alcohol Solvents	-38.7260	0.0007
	Aliphatic Solvents	-30.3879	0.0039
Vendor	vendorID_0105	-55.4461	0.0449
	vendorID_0206	-64.6910	0.0267
	vendorID_0233	59.3381	0.0445
	vendorID_0275	-57.1588	0.0390
	vendorID_0318	-69.3700	0.0251
Cumulative net worth	net worth(unit=100 EUR)	0.0028	0.0000

(2) 3-way match ***after vs before*** goods receipt (D.T.= Standard PO, I.T.= Standard)

Comparison: In Fig. 8, the two plots have no clear difference in distribution of case duration even though *3-way match before GR* has much more transactions with different types of *Sub spend areas*.

- 3-way match, after GR: Transaction with 158 vendors in 48 types of *Sub spend area text*
- 3-way match, before GR: Transaction with 1263 vendors in 111 types of *Sub spend area text*

Regression for each model:
-3-way match, after GR: For the increasing effect on case duration, 9 factors were observed in *Sub spend area text* and 13 factors were observed in *Vendor*. On the other hand, transactions with 20 factors in *Sub spend area text* and 20 vendors have deceasing effects in case duration. About *Cumulative net worth*, the case duration is averagely increased with 9.7 days when the net worth is increased with the amount of 100,000 EUR.
-3-way match, before GR: Since the model has largest frequency among 13 business models and also has many levels of attributes, there have been observed 65 significant factors in *Sub spend area text* and 91 vendors. Among

those factors, transactions with 17 factors in *Sub spend area text* and 69 vendors were observed to have increasing effect on its case duration. For the effect of *Cumulative net worth*, the case duration was averagely increased by 2.8 days when the net worth increased by 10,000 EUR.

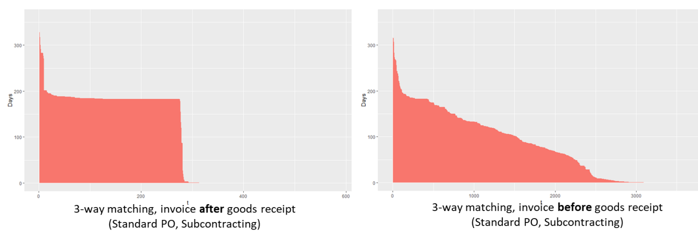


Fig. 9. Distribution of duration per each instance - part (3)

(3) 3-way match ***after vs before*** goods receipt (D.T.= Standard PO, I.T.= Subcontracting)

Comparison: Interestingly, in Fig. 9, distribution of case duration in *3-way match, after GR* looks quite constant while the other model is exponentially distributed. We can state that the prior process has been managed well with respect to order leading time.

- 3-way match, after GR: Transaction with 5 vendors in 3 types of *Sub spend area text*
- 3-way match, before GR: Transaction with 32 vendors in 14 types of *Sub spend area text*

Regression for each model:

- 3-way match, after GR: Because of the uniform distribution, there was no increasing or decreasing effect on attribute level.
- 3-way match, before GR: For the increasing effect, order leading time increased when a company had transactions with 4 vendors as in below. On the other hand, transactions with 2 factors in *Sub spend area text* and 3 vendors have increasing effects in case duration.

(4) 3-way match ***after vs before*** goods receipt (D.T.= Standard PO, I.T.= Third-party)

Comparison: Above two plots in Fig. 10 have no clear different distribution although the process model with *3-way match before GR* has much more transactions with different types of *Sub spend areas*.

Table 7. Regression analysis: the list of significant factors in part (3)

Attibutes	Factors	Estimate	p-value
3-way match, after GR			
No effect			
3-way match, before GR			
Sub spend area text	Other Logistics Services	-95.4080	0.0257
	Road Packed	-82.4135	0.0355
Vendor	vendorID_0176	30.2173	0.0003
	vendorID_0197	16.5945	0.0339
	vendorID_0260	-82.8960	0.0000
	vendorID_0548	-108.7705	0.0010
	vendorID_0581	23.7685	0.0309

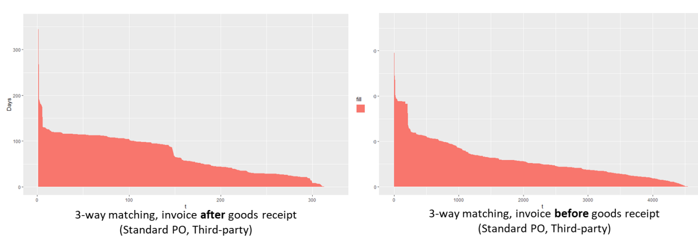


Fig. 10. Distribution of duration per each instance - part (4)

- 3-way match, after GR: Transaction with 8 vendors in 3 types of *Sub spend area text*
- 3-way match, before GR: Transaction with 40 vendors in 6 types of *Sub spend area text*

Regression for each model:

-3-way match, after GR: Among total 8 vendors, 3 vendors were observed to have significantly shorter case duration while one vendor was observed to have longer order leading time. In terms of *Cumulative net worth*, the case duration increased by 31.87 days on average, when the net worth increased by 10,000 EUR.

-3-way match, before GR: For the increasing effect in case duration, only one factor with Color Collateral showed significantly increasing effect. Regarding *Cumulative net worth*, when *Cumulative net worth* increased bof 10,000 EUR, the case duration is increased by 4.86 days on average.

Table 8. Regression analysis: the list of significant factors in part (4)

Attibutes	Factors	Estimate	p-value
3-way match, after GR			
Vendor	vendorID_0299	53.7619	0.0000
	vendorID_0374	-79.3028	0.0020
	vendorID_0381	-68.6248	0.0136
	vendorID_0476	-63.5625	0.0130
Cumulative net worth	net worth (unit=100 EUR)	0.3187	0.0182
3-way match, before GR			
Sub spend area text	Color Collateral	217.1101	0.0002
Cumulative net worth	net worth(unit=100 EUR)	0.0486	0.0003

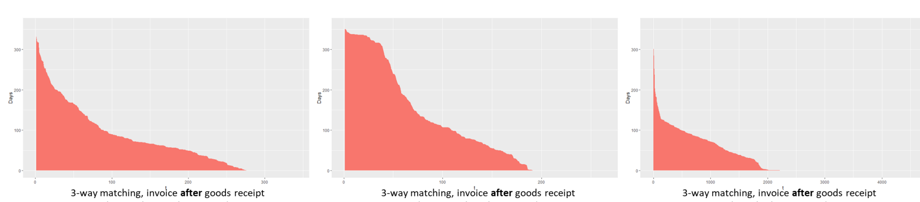


Fig. 11. Distribution of duration per each instance - part (5)

(5) 3-way match, after GR (D.T.= **EC purchase order vs Framework order vs Standard PO**, I.T.= Service)

Comparison: Fig. 11 shows that the ranges of case duration are similar as 3 processes have been handled with same *Item Type*. Due to different condition on *Document Type*, however, the amount of transactions is different for each model.

- EC purchase order: Transaction with 152 vendors in 32 types of *Sub spend area text*
- Framework order: Transaction with 88 vendors in 18 types of *Sub spend area text*
- Standard PO: Transaction with 64 vendors in 10 types of *Sub spend area text*

Regression for each model:
- EC purchase order: In *sub spend area*, 3 factors were observed to have strong increasing effect on case duration. In the meantime, 20 vendors increased the case duration while one vendor decreased it.
- Framework order: For the increasing effect, one factor about Customers in *sub spend area* and 23 vendors were observed to have significant probability. On the other hand, 6 factors in *sub spend area* and 3 vendors were observed to decrease the case duration.
- Standard PO: Two factors in *sub spend area* and 17 vendors had increasing effect on case duration while other 4 vendors had decreasing effect on it.

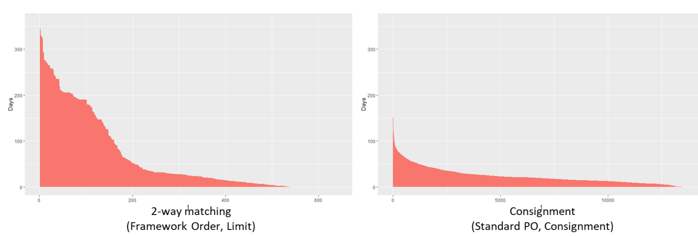


Fig. 12. Distribution of duration per each instance - part (6)

(6) The others

Comparison: In Fig. 12, *2-way match* process showed a huge deviation since it had transactions with large number of vendors in different *Sub spend areas*. For *Consignment* process, although some types of activities in ordering process

Table 9. Regression analysis: the list of significant factors in part (5)

Attibutes	Factors	Estimate	p-value
EC purchase order			
Sub spend area text	Information Services	208.2099	0.0032
	MRO (components)	223.8764	0.0020
	Process Automation & Instrumentation	202.1354	0.0042
Vendor	vendorID_0019	163.3033	0.0093
	vendorID_0022	334.6275	0.0000
	vendorID_0025	205.0393	0.0034
	vendorID_0038	318.3897	0.0000
	vendorID_0043	123.6410	0.0414
Framework order			
Sub spend area text	Customers	289.7913	0.0238
	Express	-220.9743	0.0372
	Other Logistics Services	-232.4003	0.0285
	Products for Resale	-272.1007	0.0105
	Road Packed	-272.4612	0.0102
Vendor	vendorID_0228	302.8015	0.0186
	vendorID_0231	253.7334	0.0004
	vendorID_0232	167.5334	0.0333
	vendorID_0335	-162.2973	0.0452
	vendorID_0338	174.3054	0.0269
Standard PO			
Sub spend area text	MRO (components)	100.5779	0.0092
	Other Logistics Services	60.7551	0.0096
Vendor	vendorID_0230	-29.5201	0.0000
	vendorID_0231	93.1153	0.0000
	vendorID_0232	56.7194	0.0000
	vendorID_0233	69.9751	0.0000
	vendorID_0263	125.8898	0.0048

may partly be skipped, we could extract meaningful features depending on the partial information obtained from regression analysis.

- 2-way match(D.T.= Framework order, I.T.= Limit): Transaction with 250 vendors in 17 types of *Sub spend area text*
- Consignment (D.T.= Standard PO, I.T.= Consignment): Transaction with 95 vendors in 43 types of *Sub spend area text*

Regression for each model:
- 2-way match: For increasing effect on case duration, one factor named "QHSE Services" in *Sub spend area text* made the order leading time longer while two factors in *Sub spend area text* had decreasing effect on it. With respect to *Vendor*, transactions with 10 vendors significantly increase the order leading time.
- Consignment: In spite of partial information, several significant effects were found in *Sub spend area text* and *Vendor*. Two factors in *Sub spend area text* were observed to increase the order leading time while five factors had decreasing effect. In addition, it was observed that transactions with 30 vendors increased the order leading time while transactions with 19 vendor decreased it.

Table 10. Regression analysis: the list of significant factors in part (6)

Attibutes	Factors	Estimate	p-value
2-way match(I.T=Limit)			
Sub spend area text	Government payments	-157.6528	0.0328
	QHSE Services	116.3163	0.0260
	Taxation	-199.9079	0.0031
Vendor	vendorID_1687	123.6710	0.0431
	vendorID_1688	214.3583	0.0007
	vendorID_1691	255.5862	0.0005
	vendorID_1695	197.0831	0.0023
	vendorID_1703	205.2750	0.0007
Consignment (I.T.=Consignment)			
Sub spend area text	Alcohol Solvents	-23.5552	0.0008
	Aliphatic Solvents	-24.3641	0.0003
	Aromatic Solvents	-24.2475	0.0009
	Chloride	10.9585	0.0391
	Color Collateral	-47.6652	0.0000
Vendor	vendorID_0062	-17.9757	0.0004
	vendorID_0146	15.4614	0.0002
	vendorID_0162	19.5504	0.0060
	vendorID_0175	35.8772	0.0000
	vendorID_0214	24.6655	0.0204

5 Deviation analysis based on business process model

Question 3. Finally, which Purchase Documents stand out from the event log? Where are deviations from the processes discovered in (1) and how severe are these deviations? Deviations may be according to the prescribed high-level process flow, but also with respect to the values of the invoices. Which customers produce a lot of rework as invoices are wrong, etc.?

5.1 Filtering deviant cases using Alignment technique in ProM

Our business process model is reconstructed based on *petri net* model which we assume as the baseline of model. In this problem, the *petri net* model is used for deviding the event log as 'general model' and 'deviant model' by applying *alignment* technique equipped in Prom tool.
The *alignment* is implemented for each of 13 business process models and we assumed that if a completed instance has fitness value less than 0.95, the instance is belong to 'deviant model'. The implemented result is summarized in table 11. As in Fig. 13, we observed that there are several types of deviant processes containing with missing or wrong(replicated) events in cases that fitness value is less 0.95. For the next part, we analyzed the two types of cases in 'general model' and 'deviant model' both interactively and independently.

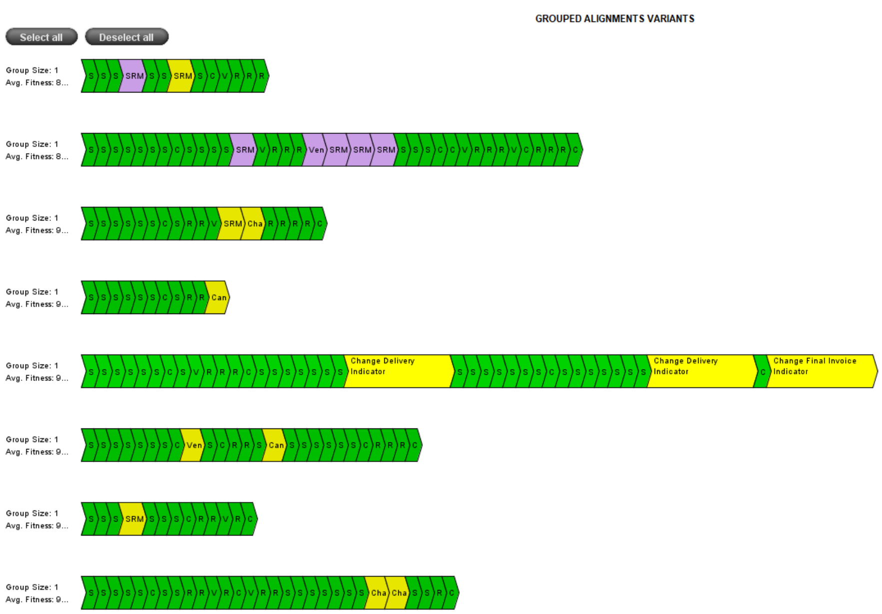


Fig. 13. An example of variants in alignment

Table 11. Alignment Regular and Deviation by Item Category

Item Category	Document Type	Item Type	# of Traces	Threshold of inductive miner	Avg. fitness	Min. fitness	General Cases (# of variants)	Deviant Cases (# of variants)
3-way match, after GR	EC Purchase order	Service	341	10%	99.6%	88.2%	309	32
		Standard	108	10%	99.3%	78.8%	98	10
	Framework order	Service	264	10%	98.3%	72.3%	221	43
	Standard PO	Service	4547	10%	97.3%	67.2%	3548	999
		Standard	7042	10%	93.4%	17.9%	5468	1574
		Subcontracting	582	10%	99%	71.4%	552	30
		Third-Party	323	10%	95.3%	57.1%	274	49
3-way match, before GR	EC Purchase order	Standard	714	10%	98.7%	55.1%	618	93
	Standard PO	Standard	182160	10%	91.7%	12.2%	125479	65293
		Subcontracting	3664	10%	99.7%	78.2%	2881	783
		Third-Party	4598	10%	97.9%	50%	3733	865
2-way match	Framework order	Limit	640	10%	99.6%	77.4%	614	26
Consignment	Standard PO	Consignment	13837	5%	96.3%	43.7%	13109	728

5.2 Comparison of distribution of features between regular and deviant cases

It is commonly expected that distributions of features between regular and deviant cases is not similar to each other. In order to test whether distributions of features of regular and deviant cases differ, various statistical tests were conducted.

Firstly, categorical features of regular and deviant cases are compared using *Chi-square test* and *Fisher's exact test*. It is widely known that *Chi-square test* can be used when the size of a contingency matrix is bigger than 22 and each cell of a matrix is bigger than or equal to 5, while Fishers exact test is acknowledged to be used when the size of a contingency matrix is 2x2 and at least one of a cell of a matrix is smaller than 5. In case where both the size of a contingency matrix is bigger than 2x2 and at least one of a cell of a matrix is smaller than 5, Monte-Carlo simulation is used to calculate the p-value. The Wikipedia says, "For hand calculations, the test is only feasible in the case of a 2 2 contingency table. However the principle of the test can be extended to the general case of an m n table, and some statistical packages provide a calculation (sometimes using a Monte Carlo method to obtain an approximation) for the more general case." There are few assumptions to be satisfied for Chi-square test and Fisher's exact test. These assumptions are that the value in a cell of a matrix should be the count of occurrences and the sum of all counts in a matrix should be equal to the size of a feature vector, which are all satisfied.

Secondly, a numeric feature of regular and deviant cases, *Cumulative net worth*, can be compared using *Kolmogorov-Smirnov test* and *Mann-Whitney-Wilcoxon test*. In order to conduct Kolmogorov-Smirnov test, normality assumption should be satisfied. To check normality, Shapiro-Wilk normality test was conducted as shown in the table below.

In our case, however, the normality assumption was not satisfied. Therefore, *Mann-Whitney-Wilcoxon test* was used to compare the distribution of a numeric feature between regular and deviation cases. The result verifies that the distribution of all four categorical features are different from regular cases to deviant cases.

Table 12. Shapiro-Wilk normality test

Item Category	Document Type	Item Type	Regular		Deviant	
			W	p-value	W	p-value
3-way match, before GR	Standard Purchase Order	Standard	0.30	<0.001	0.33	<0.001
		Third party	0.39	<0.001	0.51	<0.001
		Subcontracting	0.69	<0.001	0.71	<0.001
	EC Purchase Order	Standard	0.43	<0.001	0.19	<0.001
3-way match, after GR	Standard Purchase Order	Standard	0.38	<0.001	0.26	<0.001
		Third party	0.54	<0.001	0.64	<0.001
		Subcontracting	0.62	<0.001	0.81	<0.001
		Service	0.06	<0.001	0.08	<0.001
	EC Purchase Order	Standard	0.69	<0.001	0.68	<0.001
		Service	0.66	<0.001	0.55	<0.001
	Framework Order	Service	0.30	<0.001	0.30	<0.001
2-way match	Framework Order	Limit	0.59	<0.001	0.48	<0.001
Consignment			NA	NA	NA	NA

Table 13. Mann-Whitney-Wilcoxon test

Item Category	Document Type	Item Type	W	p-value
3-way match, before GR	Standard Purchase Order	Standard	1.02e+11	<0.001
		Third party	4.02e+7	<0.001
		Subcontracting	4.40e+7	<0.001
	EC Purchase Order	Standard	6.60e+6	<0.001
3-way match, after GR	Standard Purchase Order	Standard	2.14e+8	<0.001
		Third party	2.15e+5	0.001
		Subcontracting	1.36e+5	<0.001
		Service	1.45e+9	<0.001
	EC Purchase Order	Standard	8.61e+4	0.001
		Service	1.04e+6	<0.001
	Framework Order	Service	1.28e+6	<0.001
2-way match	Framework Order	Limit	5.67e+5	0.3358
Consignment			NA	NA

Table 14. Chi-square test / Fisher's exact test

Item Category	Document Type	Item Type	p-value			
			Resource	Spend area	Sub spend area	Vendor
3-way match, before GR	Standard Purchase Order	Standard	0.001	0.001	0.001	0.001
		Third party	0.001	0.001	0.001	0.001
		Subcontracting	0.001	0.001	0.001	0.001
	EC Purchase Order	Standard	0.001	0.001	0.001	0.001
3-way match, after GR	Standard Purchase Order	Standard	0.001	0.001	0.001	0.001
		Third party	0.001	0.001	0.001	0.001
		Subcontracting	0.001	0.001	0.001	0.001
		Service	0.001	0.001	0.001	0.001
	EC Purchase Order	Standard	0.001	0.001	0.001	0.001
		Service	0.001	0.001	0.001	0.001
	Framework Order	Service	0.001	0.001	0.001	0.001
2-way match	Framework Order	Limit	0.001	0.001	0.001	0.001
Consignment			0.007	0.001	0.001	0.001

5.3 Rework Attributes

With cases whose fitness was found to be low according to the *alignment* technique, we classified those cases into two groups based on whether lots of rework processes had occurred or not. The rework process can be discriminated based on two conditions. First condition is whether the case has at least one activity whose name has 'Change' in it (e.g. change quantity). The second condition is whether each activity occurred more than once. Using cases filtered with two conditions, we figured out the difference between rework and non-rework with respect to *Spend area text* and *Sub spend area text*. Table 15 is the list of values in (*Spend area text* and *Sub spend area text*) that only rework group has. Among 13 case groups, there were 3 groups that all the attributes in the rework also appeared at the non-rework group.

Table 15. Rework Attribute

Item Category	Document Type	Item Type	Spend area text	Sub spend area text
3-way match, before GR	EC Purchase Order	Standard	Order	MRO (components), Facility Management, Advertising, Packaging, Other Logistic Services
	Standard PO	Standarxd	Solvents	Styrene Acrylics, Paperboard
		Subcontracting	Titanium Dioxides, Pigments & Colorants, Marketing	Sulphate, MRO (components), Colorants, Point of Sales, Warehousing
		Thirdparty	-	-
3-way match, after GR	EC Purchase Order	Service	Enterprise Services	Consulting, HR Services, QHSE Services, Third Party Labor, Marketing Support Services
		Standard	-	-
	Framework Order	Service	CAPEX & SOCS, Workforce Services	Sea, Packaging, Transport & Hoisting Equipment, HR Services
	Standard PO	Service Standard Subcontracting Thirdparty	CAPEX & SOCS Specialty Resins, Commodity Resins Logistics -	MRO (components) Pure Acrylics, Vinyl Acrylics Road Packed -
2-way match	Framework Order	Limit	Real Estate, Energy	Real estate services, Real estate brokers or agents, Business park, Government payments, Electricity
Consignment	Standard PO	Consignment	Latex & Monomers	Opaque Polymers, Styrene Acrylics, Polyurethane Resins, Aliphatic Solvents, Alkyd Resins

6 Discussion

In the paper, we focused on discovering the business process model and analyzing the deviations in cases. Even though all four types of flows have disparate processes, we tried to find a comprehensive business process model that best describes the majority of the cases that belong to each type of flows. In deviation analysis, we tried to figure out whether there are deviations both in terms of the duration of time it takes from activity/instance level, and in terms of the order of activities based on business process model. In analyzing deviations in the duration of time, we investigated the deviations based on the distribution of duration for each type of processes using regression method. In analyzing deviations in activities, we applied Alignment plug-in made available in ProM to filter deviant cases, where the deviation is defined based on business process model discovered in Chapter 3. Furthermore, the extent of deviation is calculated using statistical tests, such as *Kolmogorov-Smirnov test*

and *Mann-Whitney-Wilcoxon test*.
In addition to the analysis we have conducted throughout this paper, we also came up with several findings as below.

6.1 Findings from process discovery

-According to the explanation of *3-way match, before GR* provided in BPIC 2019 guidelines, a specific user or a batch process enables invoices to be entered and receipt before the goods are receipt, while it is required to have invoices to be receipt after goods are receipt as in *3-way match, after GR*. Therefore, we could find some cases where invoices are received before goods receipt in *3-way match, before GR. In 3-way match, after GR*, however, we also could find few cases where invoices are receipt before goods receipt, which is not allowed in *3-way match, after GR*. Therefore, we assumed these cases to be outliers.
-In *2-way match*, "Create Purchase Order Item" does not always come first in *2-way match* (which should come first in usual cases). This is because most of the *spend areas* of items in *2-way match* are about real estate and the maintenance of a company, such as government payments, insurances and electricity. In case of the purchase of items related to the maintenance of a company, a company may have to pay without creating purchase order first, since the payment of these items are something compulsory, but not optional. However, the reason why "Create Purchase Order Item" comes first in case the *spend area* is real estate is not identified.
-"Change approval for purchase order" should come prior to "Delete purchase order item". This does not mean that "Delete purchase order item" should follow "Change approval for purchase order".
-In *3-way match, after GR*, there are a group of cases which has multiple "Record Goods Receipt" and "Record Service Entry Sheet" activities. If those cases are completed with the proper invoicing process activities, for instance, "Record invoice Receipt" and "Clear Invoice", there is a relationship between number of executed activities and *Cumulative net worth* (CNW). CNW in the "Record Invoice Receipt" is CNW of "Record Service Entry Sheet" multiplies the how many times the activity is executed. In case of the other invoicing activity, CNW of "Vendor creates invoice" and "Clear Invoice" are CNW of "Record Service Entry Sheet"
-When we take a look at the feature '(case) Company', "companyID_003" is in charge of "Real estate & maintenance system", which can only be found in *2-way match*. From the point of view of real estate related purchasing document, the values at the *Sub spend area text* are "Real estate brokers or agents"(27.4%) and "Real estate services"(25.58%). The values related with maintenance system are "Government payments"(23.87%), "Insurance"(3.43%), and "Escrow and title services"(3.06%), which can be considered as framework orders.

6.2 Findings from deviation analysis

-In answering the second question, we analyzed deviations of waiting time and found causes from the perspective of activity and instance level. Regarding the duration of each activity, waiting time of "Record Invoice Receipt" and "Clear Invoice" are observed to be longer than other activities for most of 13 process models. Unlike others, in case of *3-way match, after GR* with D.T.= framework order and I.T.= service, duration from "Create Purchase Oder Items" to "Vendor creates invoice" was the longest. Using these results, we recommend subsidiaries to keep track of these activities to decrease order leading time.
-In terms of case duration, *Sub spend area text*, *vendor*, and total *Cumulative net worth* factors were found to significantly affect the deviation in duration for each process model. As in Table 8 to Table 13, the list of factors shows which factors increase or decrease the case duration and how strongly or weakly they affect it. From the result, we were successful in finding the causes in attribute level and could make companies be aware of which factors in *Sub spend area text* or vendors make the ordering process longer. This finding will give an adequate solution for the problem of long order leading time in attribute levels.
-There exists a huge difference in distribution of features between regular and deviant cases. Almost all features are found to have heterogeneous distribution between regular and deviant cases.

7 Conclusion

The recommendation would be that the company must be able to clearly filter out deviant cases from event logs as not only deviant cases are found to have significantly different attributes compared to regular cases, but the difference in attributes may cause tremendous negative outcome to the company. Since the event log has handled different transactions between subsidiaries and 431 different vendors, it was challenging to make generic process model and give helpful results from the analysis for all subsidiaries. However, in this paper, we managed to provide generic results covering all participating subsidiaries and give solutions to process managers.

References

1. van Dongen, B.F. *Dataset BPI Challenge 2019.* Centre for Research Data. https://doi.org/10.4121/uuid:d06aff4b-79f0-45e6-8ec8-e19730c248f1.
2. VAN DER AALST, Wil, et al. *Process mining manifesto.* In: International Conference on Business Process Management. Springer, Berlin, Heidelberg, 2011. p. 169-194.
3. DUMAS, Marlon, et al. *Fundamentals of business process management.* Heidelberg: Springer, 2013.

Efficient and Compliant Purchase Order Handling

A Contribution to BPI Challenge 2019

Oliver Gutermuth, Johannes Lahann, Jana-Rebecca Rehse, Martin Scheid, Steffen Schuhmann, Sebastian Stephan, Peter Fettke

Institute for Information Systems (IWi) at the German Research Center for Artificial Intelligence (DFKI GmbH) and Saarland University, Campus D3.2, Saarbrücken, Germany
firstname.lastname@dfki.de

Abstract. The 2019 Business Process Intelligence challenge focuses on the purchase order handling process from a large multinational company in the area of paints and coatings, operating from the Netherlands. In this report, we describe how we analyzed the provided process data in order to answer the process owner's questions on process visualization, throughput efficiency, and compliance. To answer those questions, we used a combination of manual data analysis, established process mining techniques, and innovative machine learning approaches. After presenting our data understanding and tool chain, we first report on the results obtained by manual filtering, before addressing each challenge individually. We also discuss limitations and further recommendations, wherever applicable.

Keywords: Process Mining, Process Discovery, Process Compliance, Anomaly Detection, BPI Challenge

1 Introduction

With the help of information systems, business processes are increasingly digitized. Through the (semi-)automatic execution of process activities, the performance and the quality of the process can be improved. However, the proceeding digitization of business processes can also lead to an increasing complexity of organizations. This is particularly challenging in the context of ensuring process compliance, i.e., conformance with internal and external regulations. In this context, the recorded process data can be used to gain insights into the process, as well as its accompanying data, organizational, and social structures, with the help of process mining.

The yearly BPI challenge (BPIC) gives process mining researchers and professionals to test their tools, techniques, and methods on a real-life log. For the BPI Challenge 2019, the provided data comes from a large multinational company in the area of coatings and paints operating from The Netherlands. In particular, it is focused on the purchase order handling process for some of its 60 subsidiaries. This process is concerned with administrating and documenting purchases, including goods receipt and invoices. Each case corresponds to one purchase order item (or line item). Purchase orders (or purchase documents) may contain one or more purchase order items. For each line item, there are four different ways to handle it. If an item requires both an invoice and a goods receipt, the invoice may either be issued independent from the goods receipt (3-way-matching, invoice before goods receipt) or only after the goods receipt is entered (3-way-matching, invoice after goods receipt). Other items require an invoice without a goods receipt (2-way-matching) or vice versa (consignment).

From an analytical perspective, the process is challenging, because both the data and the underlying process are quite complex. This complexity not only makes it difficult the optimize the overall process flow, it also complicates to ensure its compliance regarding internal and external regulations. Process mining, artificial intelligence, and data analytics may help to reduce this complexity. According to the problem statement on the BPIC 2019 website, the providing company is interested in answering the following concrete questions:

- Is there a collection of process models which together properly describe the process in this data? Based on the four categories above, at least four models are needed, but any collection of models that together explain the process well is appreciated. Preferably, the decision which model explains which purchase item best is based on properties of the item.
- What is the throughput of the invoicing process, i.e. the time between goods receipt, invoice receipt and payment (clear invoice)? To answer this, a technique is sought to match these events within a line item, i.e. if there are multiple goods receipt messages and multiple invoices within a line item, how are they related and which belong together?
- Finally, which Purchase Documents stand out from the log? Where are deviations from the processes discovered in (1) and how severe are these deviations? Deviations may be according to the prescribed high-level process flow, but also with respect to the values of the invoices. Which customers produce a lot of rework as invoices are wrong, etc.?

To answer these questions, this report is organized as follows. In Sect. 2, we perform a first data analysis, describing which tools we have used, our data understanding and the results of a descriptive analysis. Manual data filtering is done in Sect. 3. Sect. 4 is centered on the first challenge, describing how the process log could be categorized further in order to simplify its analysis. In Sect. 5, we address the second challenge, using a combination of manual analytic tasks and automated data filtering. Section 6 is dedicated to the last challenge, identifying abnormal cases based on the results from the previous challenges, before the paper is concluded in Sect. 7.

2 Preliminary Analysis

2.1 Software Tools

A variety of tools was used to analyze the data from different perspectives and providing answers. While established tools were mainly used for descriptive analytics and standard process mining tasks, individualized solutions were implemented for the more challenging and data-specific questions. The following table provides a summary of the software frameworks and tools that have been used for generating the results presented in this report.

Table 1: Software tools and frameworks used for data analysis

Framework / Tool	**Purpose**
Fluxicon Disco	Process mining, process discovery, descriptive analysis
ProM 6.8	Process mining, process discovery, visualization
RefMod-Miner	Log manipulation, log conversion, trace similarity
PM4Py	Token Replay
Python 3.7 Including pandas and numpy	Log manipulation, descriptive analysis, clustering, anomaly detection
Pytorch	Language model creation
FastAI	Language model creation
Matplotlib	Visualizing
MS Excel	Analyzing a selection of individual logs using Power Pivot

2.2 Process Understanding

The main goal of the BPIC is to inspect the purchase order handling process regarding compliance issues. For this purpose, the purchase orders including the individual purchase order items should be analyzed. According to the BPIC 2019 website, purchase order items themselves can be processed using the following four different procedures, as summarized in Table 2. Here, a Product Order (PO) refers to each individual purchase order item, an invoice (IV) denotes the vendor invoice (not the invoicing) and goods receipt (GR) stands for the goods receipt document, e.g., a delivery note.

Table 2: Overview of the item categories

	3-way matching, invoice after GR	**3-way matching, invoice before GR**	**2-way matching**	**Consignment**
GR-flag	TRUE	TRUE	FALSE	TRUE
GR-based-IV-flag	TRUE	FALSE	FALSE	FALSE
Verification	PO + IV + GR	PO + IV + GR	PO + IV	PO + GR
Special feature	Invoice may only be entered after GR	Invoice can be entered before GR; however, settlement only takes place after GR	Settlement is independent of GR	Billing takes place in a separate process

As a further note, it was specified that several goods receipt declarations and invoices can be recorded for individual items (e.g., for installment payments or orders for time-delayed batches). This makes it difficult to compare the amounts, whose sum must match in the end. The values of each event were anonymized linearly such that their relation is preserved. In addition, the approval procedures (for purchase orders and settlement of invoices) were removed.

We also distinguish between two types of users: automated processes (“batch users”) and human actors (“normal users”). However, not all events are associated with a specified user. Finally, it should be mentioned that company, supplier, system, document name, and IDs have been anonymized, but product details are not.

2.3 Data Description

Before we start to discover and analyze process models, it is important to establish a general data understanding. This helps us to identify outliers and later filter the event log. For this BPIC, we will focus on the most common KPIs provided by DISCO. Using the project file provided by Fluxicon, we filtered the event log regarding the case Item category and compare the results with the whole event log.

Table 3: Data description along the four item categories

KPI	3-way matching, invoice after GR	3-way matching, invoice before GR	2-way matching	Consignment	Overall
#Events	319,233	1,234,708	5,898	36,084	1,595,923
#Start-Events	5	8	4	2	8
#End-Events	24	28	8	11	32
#Activities	38	39	11	15	42
#Cases	15,182	221,010	1,044	14,498	251,734
#Variants	4,228	7,832	148	281	11.973
Earliest Event	23.01.2001 23:59:00	26.01.1948 23:59:00	25.01.2017 23:59:00	31.12.2017 00:00:00	26.01.1948 23:59:00
Latest Event	09.04.2020 23:59:00	05.12.2019 23:59:00	01.02.2019 23:59:00	17.01.2019 22:26:00	09.04.2020 23:59:00
Median Case Duration	63.4 days	66.3 days	23.7 days	19.9 days	64 days
Average Case Duration	75 days	74.5 days	57.5 days	24.1 days	71.5 days
Max. Duration	17 years, 362 days	70 years, 120 days	1 year, 216 days	229 days, 20 hours	70 years, 120 days
Min. Duration	0 millis	0 millis	2 mins	0 millis	0 millis

2.4 Descriptive process analysis

As a next step, we employ process discovery to gain insight into the as-is process flow. Most commercial process mining tools (e.g. Fluxicon, Celonis) rely on heuristics miner and fuzzy miner. While heuristic process discovery algorithms are widely used and well-established. , we use the interactive Data-aware Heuristics Miner (iDHM) [1] for descriptive analysis in the following. The interactive exploration of the parameter space and the quality of the mined process model will provide a broad process understanding as a prerequisite for advanced analyses.

The iDHM provides several mining algorithms. Having a first look at the process model, we use the Flexible Heuristics Miner (FHM) [2]. Furthermore, we will use the all-task-connected heuristic. As stated in [1, 2], the advantage of using this heuristic is that many dependency relations are tracked without any influence of the used parameter setting.

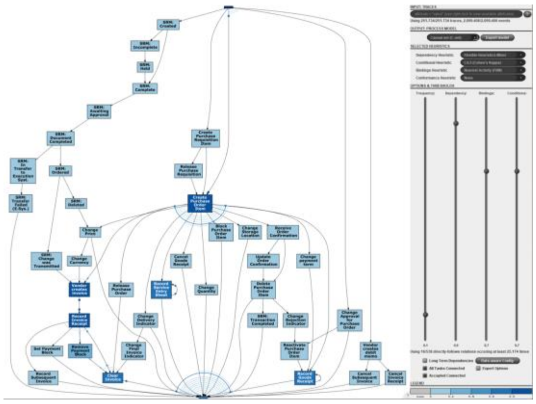


Figure 1: Mined process model with FHM (event log)

The heuristics calculates dependencies between events in a so-called dependency graph. Because the metric indicates how there is truly a dependency relation between two activities [3, 4], we only look at traces with a dependency >= 0.9. As depicted in Figure 1, several dependencies can be derived: (i) In general, the process starts with the creation of a purchase requisition item or the creation of a purchase order item. (ii) The activity *Record Service Entry Sheet* is only executed together with *Record Goods Receipt*. (iii) The activities *Vendor create invoice* and *Record Goods Receipt* are executed together. However, in some cases no vendor invoice is created.

In a next step, we compare the discovered process models regarding the four flow types. Therefore, we filtered the given event log regarding the Case Item Category to get four event logs. Doing this, all four flows look different. The bindings next to the artificial start event represent a parallel split, whereas disconnected dots represent exclusive splits. Looking at the process models, we can observe that for the 2-way matching, no goods receipt is needed (**Figure 2**). Analyzing the 2-way matching model in more detail, we see that after the invoice receipt is recorded, there are three possible ways to end the process. Conspicuously, there are cases for which no activity *clear invoice* can be observed. In addition, canceling the invoice receipt is contradictory to paying the invoice afterwards.

For the consignment flow, several activities which lead to *Record Goods Receipt* can be observed. As it is depicted in **Figure 3**, different actions can happen before the activity *Record Goods Receipt* is executed. It is interesting that there is a parallel split before the activities *Cancel Goods Receipt* and *Receive Order Confirmation*. It also looks like that in order to delete a purchase order item, an order confirmation has to be received. Here, a point of interest could be to analyze when a purchase order item is reactivated, and when not. In general, we see that all updates and changes regarding goods receipt happens before they are recorded. For the consignment process, it looks like that there are no cases where a recorded goods receipt is updated afterwards.

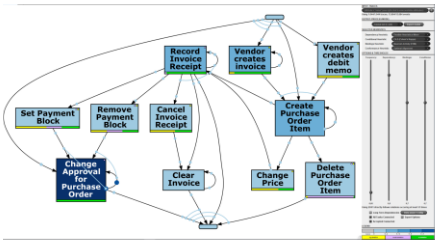


Figure 2: Discovered 2-way matching process (FHM)

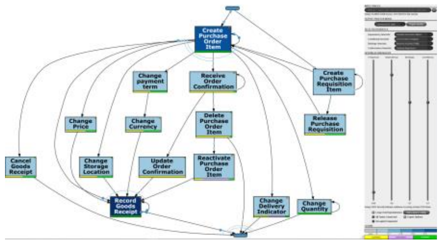


Figure 3: Discovered consignment process (FHM)

Looking at Figure 4, the flow for 3-way matching (invoice after GR) is much more complex. If we analyze the start point, we see that only 1,221 traces start with *Release Purchase Order* (1) or *Create Purchase Requisition Item* (1.220). According to the bindings, all other traces starts with more than one activity (parallel split). For example, we see that 7,440 traces start with the activity *Record Goods Receipt* AND *Create Purchase Order Item*. Interestingly, there are traces which consist of the activities *Vendor creates debit memo* and *Cancel Invoice Receipt*. These traces are conspicuous because no activity *Record Invoice Receipt* is executed.

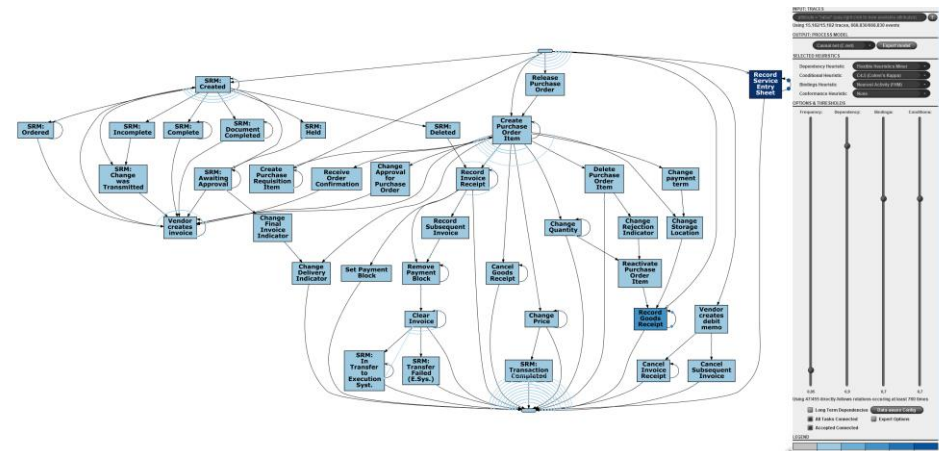


Figure 4: Discovered 3-way matching (invoice after GR) process (FHM)

If we look at Figure 5 (3-way matching, invoice before GR), we can observe that there are more or less the same SRM activities as in the process model shown before. In addition, looking at the structure, it looks like the SRM-activities form a separate and internally cohesive sub-process. As this cannot be observed in the previous two models, these activities are particular for 3-way matching flows.

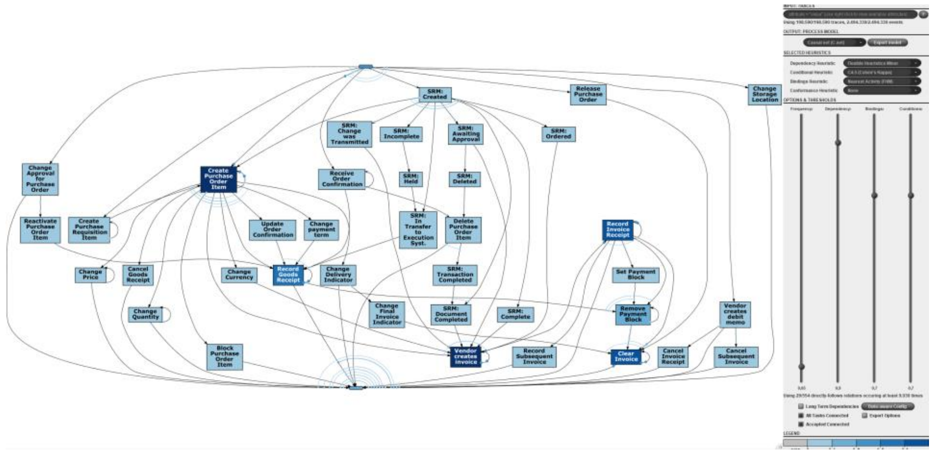


Figure 5: Discovered 3-way matching (invoice before GR) process (FHM)

3 Data Filtering

3.1 Approach

In a first step, the log was divided by process type (consignment, 2-way-matching, 3-way-matching with invoice before and after GR). This results in four sub-process logs. These sub-logs were then imported into Disco and analyzed. First, we focused on start events, beginning by inspecting the start events and decided whether or not to filter these cases according to the given timestamps. We also looked at some additional attributes of the log to find cases with potential start events before or after 2018. Additionally, we looked at other attributes, such as the order of events and their frequency to decide whether or not those cases will be included. Next, we followed an analogous procedure for the end events of the log. We looked at all potential end events and decided whether or not the corresponding cases will be included. In addition, if we were unclear about the sequence or logic behind the events, we looked at SAP's help (*https://help.sap.com* or *https://answers.sap.com*) to get a deeper understanding of the process and activity logic. This procedure was carried out on all four sub logs of the BPI Challenge dataset. Manually filtered results are also described as potential process anomalies in Sect. 6.2. The detailed filtering tables are available in Appendix A. The models resulting from this process can be considered as a proposal for the definition of a manually created reference process.

3.2 Findings of the manual filtering process of the consignment sub log

First, we filtered for the consignment process type which comprises 14,498 out of 251,734 cases. We looked for possible start events, i.e., events which occur as the first in time within a case. We identified six potential start activities (see Appendix A **Table 11**). Looking at the occurrence of the events inside the consignment sub log shows that three of the activities are start events which resulted from logs which started before 2018 and therefore resulted from cases which started before the time period we are looking at. These are therefore probably incomplete cases, where the real beginning was cut off. So, we did not include those cases inside our model. The activities *Delete Purchase Order Item* only functions as start event in cases which consists of itself and *Create Purchase Order Item*, so these are also filtered. This results in two start events: *Create Purchase Requisition Item* and *Create Purchase Order Item.*

Next, we inspected potential end events (see Appendix A **Table 12**). When considering these activities and the instantiated events, it became clear that eight of them were end events of pending process instances and were therefore no longer taken into account as end events. *Create Purchase Order Item* is only instantiated as end event if it is also start event in December 2018. We also found that *Delete Purchase Order Item* is only an end event if some process attribute was changed and update, cancel, change or reactivate events took place before. This resulted in the end events *Delete Purchase Order Item* and *Records Goods Receipt.* This filtering procedure subsequently led to a log with 13,534 cases and was subsequently discovered using DISCO (see Appendix B **Figure 21**). We also followed those steps in the following chapters.

3.3 Findings of the manual filtering process of the 2-way-matching sub log

The 2-way-matching sub log consists of only 1,044 cases and contains five potential start events (see Appendix A **Table 13**). *The filtering then resulted in the start events Create purchase order item* and *Vendor creates invoice.* Next, we identified potential end events based on the already found start events, finding seven potential end events (see Appendix A **Table 14**). This leads to *Clear Invoice* and *Records Invoice Receipt* as end events. Using those endpoint filters in addition with the filtered timeframe led to the discovered model in Appendix B Figure 22, based on 1,044 cases.

3.4 Findings of the manual filtering process of the 3-way-matching after GR sub log

In this subprocess, we first extract the SRM subprocess. The SRM activities appear very closely connected and are probably carried out conjointly in Supplier Relationship Management, a separate SAP module. Therefore, we filter this process out of the log and analyze it separately (538 out of 251,734). This led to the model visualized in Appendix B Figure 24. In the following, the activities considered here were filtered out in order to get a more exact impression of the remaining 3-way-matching log. The first filter results of the potential start events are listed in Appendix A Table 15. This results in a large number of potential end events. After the detailed analysis of these events we found eight potential end events, listed in Appendix A Table 16. For these cases, however, no exact decision could be made, so they were not filtered. The other activities resulted from pending and timeframe overlapping process instances. After filtering for those start and end points, we discovered the model visualized in Figure 6. However, it must be considered that due to the reduction of the displayed paths, some start and end events are not directly recognizable as such, since they are only present as such in a few cases.

3.5 Findings of the manual filtering process of the 3-way-matching before GR sub log

After filtering for the 3-way-match before GR sub log (194,288 cases), we identified four potential start events. Here, only *Vendor creates debit memo* can be excluded as a start event because of hanging or time-overlapping process instances. This leads to a large number of potential events as process start points (see Appendix A **Table 17**). After further filtering and detailed analysis, we identified e.g. *Cancel Goods Receipt, Cancel Invoice Receipt and Cancel Subsequent Invoice* as end events (see Appendix A **Table 18**). This results in a filtered log which consists of 194,228 cases and is visualized in Figure 7.

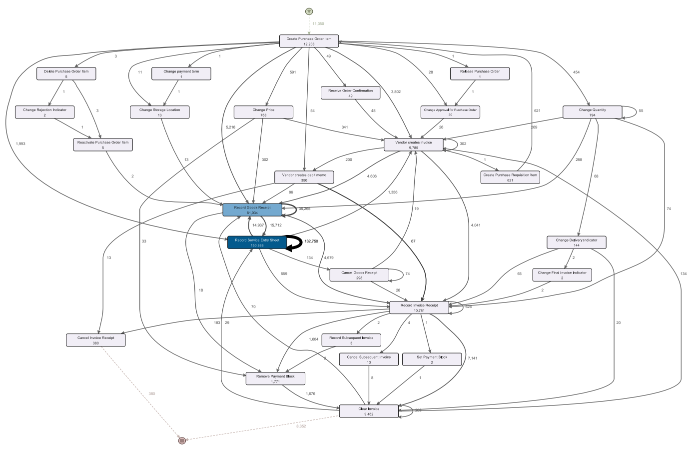


Figure 6: 3-way-matching after GR (100% of the activities and 40% of the paths)

4 Challenge 1: Process Model Collection Mining

4.1 Approach

The process owners' first question was to find a collection of process models, which together properly describe the process in this data. Since there are four fundamental ways to handle a line item, at least four models are needed, but the process owners were open to more models as long as they explain the process well. They preferred a collection of models, where the assignment of line items to models is based on properties of the item.

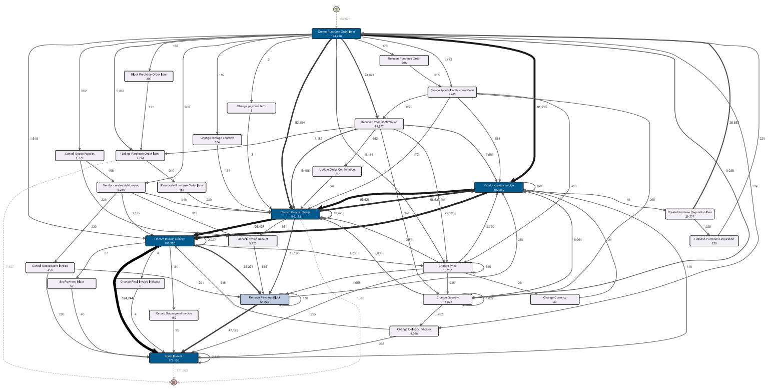


Figure 7: 3-way-matching before GR (100% of the activities and 20% of the paths)

The premise of this challenge was that we should inspect each of the four process types individually, deciding which one could benefit from being split into multiple processes. So, we first inspected the four sub-logs, as described in the previous section 3, but unfiltered to avoid potential distortions. Both the consignment and 2-way matching process that were mined for the respective logs were rather small and structured, giving stakeholders a good understanding of the process flow. For those two subprocesses, we saw no need for further separation. The same, however, could not be said about the two 3-way matching sub-logs. 3-way matching, invoice before GR is the largest log by a considerable margin; its many variants make it quite difficult to identify a clear structure. The same can be said about 3-way matching, invoice after GR, which is remarkable, because this log is much smaller. Nevertheless, manual filtering revealed that it has a very high number of distinct process variants. So, in this challenge, we focused on the two 3-way matching sub-logs, with the goal to separate them into more detailed process models. As explained above, we also removed the SRM process from both logs, because we considered it as a separate and independent subprocess of both.

For both datasets, we first approached this challenge in a data-centric way. The basic idea is to use clustering techniques in order to see similarities and differences between groups of process instances within one process type and use those results for ensuing manual analysis. The idea behind clustering techniques is to group a large dataset, such that the datapoints in one group (or cluster) are more similar to each other than to those objects in other groups. The similarity measure between datapoints can be individually defined. For our dataset, we executed the following steps in order to find a viable set of models to describe the purchase order handling process:

1. Separate logs: We separated the overall log into four (complete and unfiltered) sub-logs, each corresponding to one process type. The following steps were executed separately for each log.
2. Compute trace similarity: Trace clustering requires a similarity measure between traces, which can be defined in many different ways [5]. The result is a matrix with

pairwise similarity values for all traces, which was used as the input data for the clustering algorithm.

3. Dimensionality reduction: In order to group similar traces based on the generated similarity matrix, T-SNE was used to reduce the similarity to a two-dimensional space.
4. Separate cluster logs: After dimensionality reduction, our log was separated into multiple groups of similar traces, using the K-means algorithm to identify near located cluster in the projected space, which were then exported as individual logs for further handling.
5. Manual analysis and evaluation: Using Disco and the Inductive Miner [6], the separated logs and mined models then served as the basis for further manual data analysis, which was necessary to find a reasonable number of process models, with a clear connection to line item properties.

4.2 Identifying Process Models for 3-way-matching, invoice before GR

For analyzing the 3-way-matching, invoice before GR sub log, we used a trace clustering based on activity feature vectors. The log is quite large, so we required a computationally efficient way to determine the similarity between two traces. In this approach, the log is represented as a matrix, with a row for each activity and a column for each trace. Each value denotes how often the respective activity appears in the respective matrix. The dimensionality of this feature matrix is then reduced using the T-SNE algorithm, preserving the distance between the datapoints in the projected two-dimensional space. In order to optimize the hyperparameter, we have implemented an iterative approach to use different parameters for perplexity as well as iterations. The results were then compared to find the best hyperparameter combination based on the shape of the resulting clusters. For the further analyses we have chosen a value of 10 for perplexity with 500 iterations.

The results for the 3-way-before log are shown in Figure 8. The most discernible feature is the large cluster centered around the origin. It is surrounded by a ring of more scattered objects. The question remains, how the trace similarity related to features in the log and whether these results can be transformed into a viable set of process models.

After starting to inspect the center cluster, we realized that it corresponded to one specific process variant covering 79,487 cases, more than a third of the log. It consists of five activities (*create purchase order item*, *vendor creates invoice*, *record goods receipt*, *record invoice receipt*, *clear invoice*) and describes the handling of a line item (one case) with exactly one invoice, one goods receipt, and no other features. In the next step, we inspected further process variants, for which the same properties apply (activities *create purchase order item*, *vendor creates invoice*, *record goods receipt*, and *record invoice receipt* appear exactly once). We found that these amount to more than 90% of the 3-way-before log, while still producing a fairly structured process model. Most of the traces were rather small and they all contained each activity exactly once, whereas all other traces repeated activities for multiple invoices or goods receipt. The discovered model is shown in Appendix B Figure 26. The "happy path", i.e., the most frequently executed process flow is clearly visible, but it also includes the multiple variants to the process, such as attribute changes, the approval subprocess, or purchase cancellation.

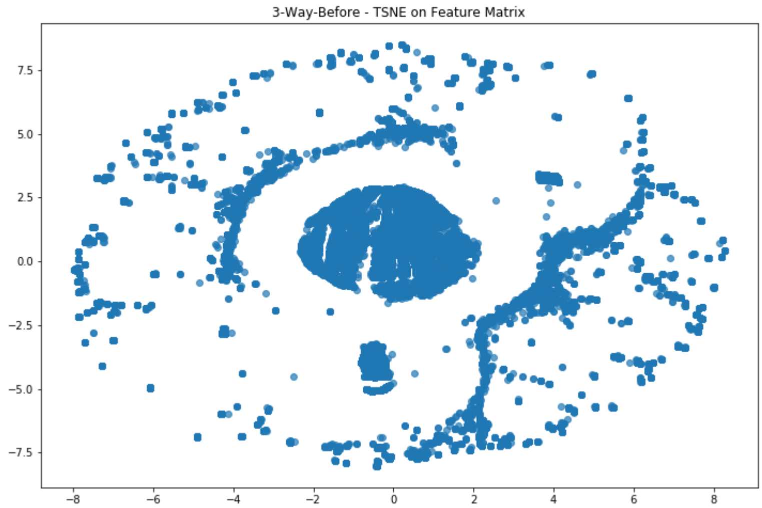


Figure 8: Clustering results for 3-way-before

After separating the line items with one invoice and one good receipt from the log, only about 10% traces remained, covering line items with multiple invoices and goods receipt. The model discovered for these cases is shown in Appendix B Figure 25. While not as structured as the model above, it also shows the generic process structure, including the prevailing repetition of activities.

4.3 Identifying Process Models for 3-way-matching, invoice after GR

Next, we focused on the sub-log for 3-way-matching, invoice after goods receipt. We essentially employed the same procedure as above, but with a different similarity measure. As the log is fairly small (14.571 cases), with 25 activities, but more than 4,700 process variants, our goal was to determine trace similarity with a focus on process structure instead of solely activities. Therefore, we used the Levenshtein trace similarity, defined as the normalized Levenshtein distance between two traces, i.e., the number of insert, delete, and replace operations required to transform one trace into the other [7]. This produced a similarity feature vector for each trace, which was again projected onto a two-dimensional space using the t-SNE algorithm. The clustering results are shown in Figure 9. Again, we see a large cluster centered around the origin and accompanied by a slightly smaller, but also quite compact cluster on its bottom left. Those two are surrounded by a ring of smaller clusters and individual objects.

These results again were the starting point for our manual analysis. Inspecting the individual clusters, we quickly saw that one of the central activities was *record service* *entry sheet*, indicating that the corresponding line item was some kind of service. Depending on the cluster, this activity was repeated frequently, typically alternating with *record goods receipt* and leading to many slightly different process variants, depending on the number of times a service was delivered. As the presence of this activity introduced a high variability into the process, we decided to split the log into those traces that contained the activity (presumably service line items) and those that did not (presumably non-service line items).

The process model for the non-service line items was discovered from 9,377 traces and is shown in Appendix B Figure 27. Its structure is similar to the process shown in Appendix B Figure 26, with a clearly discernible "happy path" and several optional or less frequent activities, mostly attribute changes and cancellations. The payment subprocess is isolated and can be found towards the end. Appendix B Figure 23 shows the process model for the service line items, discovered from 5,194 cases. It is less structured than the other processes, but also contains fewer activities (15), which still makes it fairly easy to understand. Even with 0% paths, one can see the frequent loop of activities create purchase order item, record service entry sheet, and record goods receipt, which only becomes more pronounced when displaying more paths.

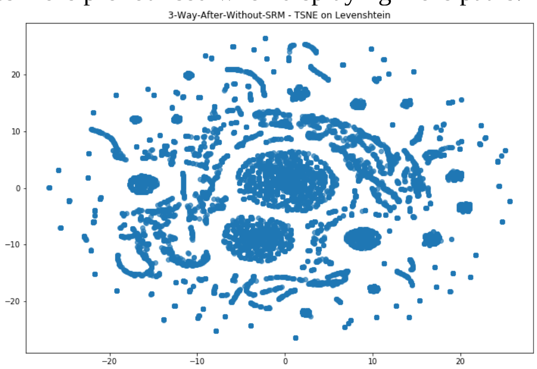


Figure 9: Clustering results for 3-way-after

4.4 Conclusion

To conclude, we suggest describing the process at hand with seven process models:

1. Consignment as shown in Appendix B Figure 21
2. 2-way-match as shown in Appendix B Figure 22
3. 3-way-matching, invoice before goods receipt: items with one invoice and one goods receipt (without SRM subprocess) as shown in Appendix B Figure 26
4. 3-way-matching, invoice before goods receipt: items with multiple invoices and goods receipts (without SRM subprocess) as shown in Appendix B Figure 25
5. 3-way-matching, invoice after goods receipt: service items (without SRM subprocess) as shown in Appendix B Figure 23
6. 3-way-matching, invoice after goods receipt: non-service items (without SRM subprocess) as shown in Appendix B Figure 27

7. Supplier Relationship Management (subprocess) as shown in Appendix B Figure 24

In our opinion, this collection of process models balances out the competing requirements of a having low number of models on the one hand and a having well structured, easy to understand models on the other hand. One could criticize that our separate sublogs differ considerably in size, but they still produce process models of comparable size and structure. This collection also fulfills the requirement that the assignment of line items to models is based on properties of the item. Regarding compliance (as addressed in the next section), it has another advantage: If the matching of line items with their respective invoices and goods receipt is determined by the assigned process model, it becomes a lot easier to check the compliance of individual line items and find potential violations.

5 Challenge 2: Position Matching

5.1 Approach

The second task focusses on the performance of the invoicing process. This process and its variants are important as they are subject to compliance guidelines and related to the four designated flow types. Invoices must be processed efficiently and with a low chance of errors. In addition to compliance requirements, the process steps and their timing should also be inspected, in order to obtain indications of possible disruptions. Analyzing the invoicing process is complicated by the differing complexity of cases within a purchase order document, which may have overlaps or dependencies as well as faulty or incomplete logs. Basically, cases can be defined at the level of a line item or a purchase order. As the selection of compliance procedure uses the properties of an individual item, a categorization on this hierarchical level seems appropriate. The results from Sect. 4 show significant differences between the process flows, depending the frequency of events within a case. Based on this insight, forming two disjunctive sets of cases seemed reasonable.

- Set 1 contains cases having a cardinality of 1:1 between case purchasing document and case concept name
- Set 2 contains cases having a cardinality of 1:n between case purchasing document and case concept: name.

These sets are used for an individual approach to assess the throughput of the invoicing process. To calculate the throughput, events must be selected to define relevant intervals. As there are 42 event types, for which their relevance must be decided, there are two major aspects to consider. The event type label indicating the process flow was used in conjunction with the absolute frequency of its occurrence for a selection. This way, the three most important event types related to the invoicing process were identified (events *Record Goods Receipt*, *Record Invoice receipt* and *Clear Invoice*) and used for further execution. The main challenge is the systematic examination of all cases containing relevant events of the invoicing process. Accordingly, the occurrence of the three event types per case is determined and the duration between them recorded. Cases with less than two of them are not considered for the measurement of time intervals. Figure 10 shows a scenario, where every relevant event type has a unique appearance:

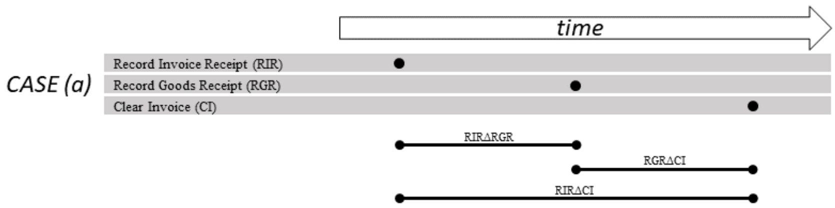


Figure 10: invoicing example: events, sequence, duration

As the order of the appearing events may differ, there are six possible durations to measure, but not all cases can be handled the same way. Basically, the net worth of invoices, the goods receipts, and the clearings should match for a line-item. But as there could be several recurring events for these types within a specific order, the selected method must fit the case characteristics. Therefore, two different methods must be used and chosen considering the cases' complexity. Appropriate approaches must take the frequency of occurring instances of an event type within a single case into account. This ensures that different categories of event quantities within a single event type can be handled separate. Figure 11 shows the criteria determining the procedure.

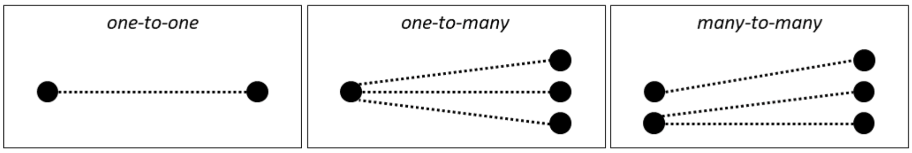


Figure 11: Event quantities and relations

There are three major aspects to consider for measuring the process performance. First, the two sets of cases were built as a significant difference can be expected. Second, the complexity of the relations towards relevant events of the invoicing process leads to the need for two different methods. For one-to-one and one-to-many-relations the explicitness of the sequence allows an accurate measurement of durations as temporal ties are explicit. But the many-to many-relations present the problem of assigning subsequences. A reference was needed. The examination of the XES-file did not provide an appropriate solution as it is limited to values which cannot assign the actual sequence of the process flow when many-to-many-relations appear. The approach to match information from the attribute *event Cumulative net worth (EUR)* failed as too many incongruities were identified. Even the inclusion of the events *Change Quantity, Change Price, Delete Purchase Order Item, Cancel invoice Receipt, Cancel Goods Receipt* which are expected to determine the cumulative net worth could not contribute to the clarification. Thus, the cases containing many-to-many relations were analyzed by using an approximation method considering all events without guessing the actual flow. In this case, the weighted mean of timestamps was calculated for every occurring event type. The number of relations was derived from the frequency of events within a many-to-many relation. It was assumed that sequences between them occurred just as often as the event with the most frequent expressions per type. For the calculation of average through times the following steps have been processed.

1. Two sets of cases are formed (single line item orders and multiple line item orders).
2. Within each group, all cases containing at least two of the relevant event types with only one type with more than one instance are located, thus analyzing events defined by one-to-one-relations and one-to-many-relations).
3. Depending on the order of occurring events, the average duration between the event types is measured using the accurate method.
4. The cases with many-to-many relations are measured using the approximate method.
5. As the throughput could also address additional useful metrics, the frequencies of event types and the cumulated net worth passing the flow are analyzed.
6. The results are presented and discussed.

5.2 Results

In order to the approach's description the tables within this chapter are created respecting the following scheme.

Table 4: Method dataset combination

	accurate method	approximate method
set-1 (single-case-orders)	**Table 5**	**Table 7**
set-2 (multiple-case-orders	**Table 6**	**Table 8**

To simplify the table's labeling the duration for a measured interval is defined by the timestamp of the limiting events using *RGR* for *record goods receipt*, *RIR* for *record invoice receipt* and *CI* for *clear invoice*.

Table 5: Applying the accurate method on case-set-1 (single-case-orders)

Category	Δ(RGR,RIR)	Δ(RIR,RGR)	Δ(RIR,CI)	Δ(CI,RIR)	Δ(RGR,CI)	Δ(CI,RGR)
2-way match	-	-	7d, 10h	-	-	-
3-way match, invoice after GR	40d, 13h	5d, 23h	33d, 22h	49d, 9h	78d, 16h	-
3-way match, invoice before GR	25d, 0h	4d, 8h	52d, 14h	47d, 23h	67d, 18h	42d, 13h

Table 6: Applying the accurate method on case-set-2 (multiple-case-orders)

Category	Δ(RGR,RIR)	Δ(RIR,RGR)	Δ(RIR,CI)	Δ(CI,RIR)	Δ(RGR,CI)	Δ(CI,RGR)
2-way match	-	-	10d, 1h	-	-	-
3-way match, invoice after GR	38d, 17h	16d, 10h	33d, 11h	41d, 3h	65d, 23h	-
3-way match, invoice before GR	21d, 3h	3d, 6h	48d, 18h	54d, 23h	65d, 17h	15d, 5h

Table 7: Applying the approximate method on case-set-1 (single-case-orders)

Category	Δ(RGR,RIR)	Δ(RIR,RGR)	Δ(RIR,CI)	Δ(CI,RIR)	Δ(RGR,CI)	Δ(CI,RGR)
2-way match	-	-	0d, 18h	-	-	-
3-way match, invoice after GR	28d, 2h	7d, 17h	16d, 2h	9d, 13h	36d, 2h	11d, 12h
3-way match, invoice before GR	27d, 18h	8d, 12h	38d, 2h	47d, 13h	66d, 12h	2d, 4h

Table 8: Applying the approximate method on case-set-2 (multiple-case-orders)

Category	Δ(RGR,RIR)	Δ(RIR,RGR)	Δ(RIR,CI)	Δ(CI,RIR)	Δ(RGR,CI)	Δ(CI,RGR)
2-way match	-	-	9d, 16h	-	-	-
3-way match, invoice after GR	37d, 4h	8d, 5h	27d, 4h	14d, 2h	69d, 4h	10d, 22h
3-way match, invoice before GR	21d, 21h	3d, 11h	48d, 14h	49d, 15h	66d, 0h	15d, 14h

Table 9: Overall average durations

Category	Δ(RGR,RIR)	Δ(RIR,RGR)	Δ(RIR,CI)	Δ(CI,RIR)	Δ(RGR,CI)	Δ(CI,RGR)
2-way match	-	-	311	-	-	-
3-way match, invoice after GR	49026	774	15637	1882	44398	930
3-way match, invoice before GR	203399	16803	182474	1175	191521	85

Table 10: Interval frequencies

Category	Δ(RGR,RIR)	Δ(RIR,RGR)	Δ(RIR,CI)	Δ(CI,RIR)	Δ(RGR,CI)	Δ(CI,RGR)
2-way match	-	-	10d, 18h	-	-	-
3-way match, invoice after GR	24d, 23h	10d, 8h	20d, 9h	16d, 2h	41d, 10h	10d, 11h
3-way match, invoice before GR	23d, 12h	3d, 10h	36d, 11h	36d, 17h	64d, 6h	-

5.3 Conclusion

For evaluating the process performance of invoicing, 708,415 intervals between relevant event types were used. Especially determining their durations and comparing them with respect to the flow variants revealed several findings. On the one hand, for the major part of 653,538 relations accurate key metrics could be calculated. On the other hand, it became obvious how versatile the logs for this process can show off and what kind of challenges their analysis entails. Regardless of the problems occurred by trying to measure complex process flows with many-to-many relations it was possible to also ascertain results for these 54,877 relations.

Not all cases in the data set have been useful for investigating the invoice process. Several of them did not contain any relevant events or they only revealed instances of one single event type, so an interval within the invoicing process could not be identified. This applies to the consignment process, because of the lack of invoices on purchase order level. A closer look at the remaining flow variants offered certain insights. As indicated by the tables, 2-way match processes do only show off one interval type as GR-events are obsolete and not recognized. Their duration met the expectation that they were rather short compared to other flow variants. However, the duration of the interval between recording invoices and recording goods receipts was recorded with an average of 3 days. A possible reason could be the fact that both events are triggered together as a batch process. The data within the 3-way-procedure confirms the assumption of longer processing times for invoice-after-goods-receipt-flow. The existence of sequences where recording and clearing invoices occur earlier than the record-goods-receipt-events for 3-way-matching with invoice-after-goods-receipt-requirement are remarkable as they show a possible compliance issue within 1704 sequences.

Overall, the results provide an overview of the invoicing process and especially mark the different duration as well as its relation to the respective flow variant. The average durations should be quite accurate as 92.3% of considered intervals have been processed by the accurate method. However, 7.7 % remain for an approximation. These performance indicators could be used for benchmarking and to analyze the company's evolution over time. Furthermore, some insights regarding compliance violations could encourage additional research towards the course of events within specific cases.

6 Challenge 3: Anomaly Detection

6.1 Approach

The process owner's third question deals with the detection of conspicuous and unusual process cases. On the one hand, process cases are considered anomalous if they do not fit any of the process models derived in challenge 1. On the other hand, a process case is regarded as potentially abnormal if it has either an unusual sequence of activities or extreme attributes, such as an unusually long duration, or extremely high costs. To answer the question, we applied multiple, partly interrelated approaches that address different aspects of the problem:

1. Descriptive attribute-based analysis: Based on a specific attribute, we searched the process log for conspicuous values or value combinations, i.e., process cases with unusual start and end times, or surprising long process executions.
2. Token-based replay: We used token-based replay to match the event log and the mined process models from challenge 1. All event logs that did not fit any of our mined models were considered anomalous.
3. Language model encoding: We trained a neural network to learn a language model from the case activity sequences. Next, we trained a second neural network to classify the event logs to one of the mined process models. We used the language model to encode the process cases before we passed them to the classifier. Subsequently, we compared the results of the classification task with those of token-based replay.
4. Root cause analysis: After identifying a series of abnormal process cases through 1-3, we compare the distributions of the potential anomalies and the normal process cases according to specific attributes, i.e., do the abnormal process cases often involve certain users or vendors?

6.2 Descriptive attribute-based analysis

During the detailed analysis and the filtering of the process we found some potential anomalies in the BPI log. The first revelation was that some process executions consists only of 2 events, which occurred frequently and independently of each other in terms of time (*Create Purchase Order Item* followed by *Delete Purchase Order Item*). This is particularly often done by certain users (e.g., in case 4507000392_00010, 4507000998_00040, 45070001217_00010 and several others by user 085). As these cases often take less than a minute, this may indicate an automated process or a batch processing. Another noteworthy feature is the *Create Purchase Oder Item* activity followed by *Change Price*. Those cases (e.g. 4507033253_00010 and 4507033261_ 00010) occur also in the consignment sub log and are carried out by the same users (388, 234) several times, with the start and end of the cases separated by one minute. However, the proximity of case ids could indicate a system-related error.

In the 2-way matching sub-log, we found that the most common variant of cases is the creating of a purchase order item followed by the activity *Change Approval for Purchase Order*. Additionally, those steps are always performed by user 602 and 603 and are also conspicuously close to each other or slightly offset in time (e.g. in case 4508076194_00010 or 4507075965_00040). In addition, it is noticeable that users 602, 601 and 60-65 are very often involved in conspicuously uniform processes. Here, the probability that these are not real system users is very high. When looking at the process frequencies, it was also found that there is a significant increase of the activity *Record Invoice Receipt* in the middle of the log period (e.g. case 4507075981_00010 and case 4507076007_00010). Looking at the log structures and sequences, it became clear that there are some conspicuous sequences within the log, including the continuous sequence of identical steps beyond those already described above, e.g., the activity *clear record invoice receipt* is followed by *clear invoice* and *vendor creates invoice.*

6.3 Token-based replay

By comparing the event log with the traces in the process model, token-based replay allows us to identify early deviations and have a first look at the replay ability of traces. As starting point for anomaly detection, we will use this technique to identify all the cases with potential deviations. Doing this, we focus on the measured fitness value. A fitness value of 1 means that a trace (or all traces of the event log) can be replayed without missing tokens, whereas a fitness value of 0 means that there are missing and remaining tokens. Thus, a valid execution sequence of the process model is not given.

To identify the abnormal case, we mined different process models based on six identified process models from Challenge 1 (excluding SRM). For token-based replay, we used the python package pm4py (http://pm4py.org/). The models were mined using the ProM plugin *Mine Petri net with Inductive Miner* [6], applying different noise threshold (5, 10, 20, 30). Applying token-based replay on the different petri-net models in PM4Py, we get the results shown in Figure 12.

Next, we selected only the models with the highest average trace fitness. These will be used to check conformance of each model based on the original event log. Our idea is that we want to know which models' conformance is given and can be replayed by the event log. For those which cannot replayed, a potential anomaly could result. Figure 13 shows the results.

log	model	average_trace_fitness	log_fitness	perc_fit_traces
consignment/consignment.xes	consignment/consignment-imi-05.pnml	80.700786	0.948929	0.929467
consignment/consignment.xes	consignment/consignment-imi-10.pnml	75.982894	0.917785	0.896832
consignment/consignment.xes	consignment/consignment-imi-20.pnml	75.962202	0.910183	0.883898
consignment/consignment.xes	consignment/consignment-imi-30.pnml	88.432887	0.952468	0.935357
2-way-match/2-way match.xes	2-way-match/2-way-match-imi-05.pnml	99.137931	0.998453	0.997536
2-way-match/2-way match.xes	2-way-match/2-way-match-imi-10.pnml	99.137931	0.998453	0.997536
2-way-match/2-way match.xes	2-way-match/2-way-match-imi-20.pnml	59.961686	0.926794	0.910910
2-way-match/2-way match.xes	2-way-match/2-way-match-imi-30.pnml	50.191571	0.781656	0.658020
3-way-before/3-way-before-one-line-item.xes	3-way-before/3-way-before-one-line-imi-05.pnml	88.429760	0.947728	0.959083
3-way-before/3-way-before-one-line-item.xes	3-way-before/3-way-before-one-line-imi-10.pnml	75.041886	0.923406	0.937011
3-way-before/3-way-before-one-line-item.xes	3-way-before/3-way-before-one-line-imi-20.pnml	87.071681	0.960404	0.967427
3-way-before/3-way-before-one-line-item.xes	3-way-before/3-way-before-one-line-imi-30.pnml	99.178068	0.945271	0.960953
3-way-before/3-way-before_mult_line_item_witho...	3-way-before/3-way-before-mult-line-imi-05.pnml	75.891678	0.963476	0.965985
3-way-before/3-way-before_mult_line_item_witho...	3-way-before/3-way-before-mult-line-imi-10.pnml	48.637715	0.881452	0.885334
3-way-before/3-way-before_mult_line_item_witho...	3-way-before/3-way-before-mult-line-imi-20.pnml	0.000000	0.671858	0.645219
3-way-before/3-way-before_mult_line_item_witho...	3-way-before/3-way-before-mult-line-imi-30.pnml	0.000000	0.741095	0.713111
3-way-after/3-way-after_without_srm_service.xes	3-way-after/3-way-after_without_srm_service-im...	92.164035	0.993245	0.994919
3-way-after/3-way-after_without_srm_service.xes	3-way-after/3-way-after_without_srm_service-im...	12.244898	0.854073	0.954206
3-way-after/3-way-after_without_srm_service.xes	3-way-after/3-way-after_without_srm_service-im...	12.244898	0.852917	0.953672
3-way-after/3-way-after_without_srm_service.xes	3-way-after/3-way-after_without_srm_service-im...	2.387370	0.722460	0.812670
3-way-after/3-way-after_without_srm_without_se...	3-way-after/3-way-after_without_srm_without_se...	92.417618	0.972833	0.964196
3-way-after/3-way-after_without_srm_without_se...	3-way-after/3-way-after_without_srm_without_se...	72.176602	0.912530	0.899601
3-way-after/3-way-after_without_srm_without_se...	3-way-after/3-way-after_without_srm_without_se...	68.859977	0.895967	0.893476
3-way-after/3-way-after_without_srm_without_se...	3-way-after/3-way-after_without_srm_without_se...	68.859977	0.888001	0.890800

Figure 12: Results of Token Replay using PM4Py

	consignment-imi-30.pnml	2-way-match-imi-05.pnml	3-way-before-one-line-imi-30.pnml	3-way-before-mult-line-imi-05.pnml	3-way-after_without_srm_service-imi-5.pnml	3-way-after_without_srm_without_service-imi-5.pnml	label
0	True	True	True	True	True	True	True
1	True	True	True	True	True	True	True
2	False	True	True	False	True	True	True
3	True	True	True	True	True	True	True
4	True	True	True	True	True	True	True
5	True	True	True	True	True	True	True
6	True	True	True	True	True	True	True
7	True	False	False	True	False	False	True
8	True	True	True	True	True	True	True
9	False	True	False	False	True	False	True

Figure 13: Results of token-based replay of the fittest models and the original event log

A TRUE-flag means that conformance is given. If a case has only TRUE-flags, this means that no anomaly could be identified. While conformance is given for each model, the case consists of standardized activities. For cases which have a TRUE-flag as well as a FALSE-flag, deviations are identified. For us, only the cases with only FALSE-flags are interesting. These are abnormal cases because they cannot related to even one of the mined models. We could identify 250.328 compliant cases, whereas 1.406 are abnormal. These will be analyzed more in detail in the following.

6.4 Language model encoding and anomaly classification

In this section, the objective is to develop a classifier that can map the process cases to the mined process models from challenge 1 and highlights cases that are particularly difficult to assign to the process models. We argue that in some ways event logs have similar characteristics than natural language, i.e., both types have long term dependencies and hierarchical relations and we, therefore, can apply state of the art text classification techniques to business process intelligence problems. In NLP, language modeling is the task of assigning a probability to sentences in a language. Besides appointing a probability to each sequence of words, the language models also assign a probability for the likelihood of a given word (or a sequence of words) to follow a sequence of words. A language model can be developed and used standalone, such as to generate new sequences of text that appear to have come from the corpus. However, language modeling is a root problem for a large range of natural language processing tasks. More practically, language models are used on the front-end or back-end of more sophisticated models for tasks that require language understanding. The idea is, that if we interpret the activity sequences of process cases as sentences, we can apply language models on the event logs to obtain process understanding of the underlying processes. The proposed approach is highly inspired by [8] and also has some similarities with [9]. The approach consists of two steps, an unsupervised language model pretraining and a target task classifier fine-tuning. In our experiments, we used the state-of-the-art language model AWD-LSTM [10], a regular LSTM (with no attention, short-cut connections, or other sophisticated additions) with various tuned dropout hyperparameters.

For fine-tuning the classifier, we augment the pretrained language model with two additional linear blocks. Following standard practice for computer vision classifiers, each block uses batch normalization [11] and dropout, with ReLU activations for the intermediate layer and a SoftMax activation that outputs a probability distribution over target classes at the last layer. For a more detailed description of the neural network architecture, we refer the reader to [8] as the applied architecture has been reimplemented with a few minor adaptations.

We represented each process case by the sequence of its contained activities. Afterward, we sorted all cases by date and concatenated the events of all claims in a full event stream. Additionally, a unique character was inserted at the beginning of a case to represent the beginning of a process instance. The event stream was split into training and validation data by ratio 0.8. Training the language model for 5 epochs on the training data resulted in an accuracy of 0.91 on the validation data. Figure 14 visualizes the training process.

epoch	train_loss	valid_loss	accuracy	time
0	0.274873	0.264056	0.907840	02:06
1	0.248734	0.247786	0.913724	02:08
2	0.249719	0.245029	0.913685	02:09
3	0.252426	0.242458	0.913646	02:10
4	0.249909	0.243616	0.913592	02:09
5	0.270247	0.248494	0.912880	02:09

Figure 14: Training phase of the language model along with accuracy on the validation set.

By comparison, current language models usually achieve significantly lower accuracy values in the range around 0.3. This can be partly explained by a significantly lower complexity of the underlying processes in comparison to natural language. However, it also clearly shows that the language model was able to learn a deep process understanding to be able to predict the next process steps. Figure 15 show a randomly selected sample from the validation set. In the next step, we trained a classifier to map from the event stream representation of a process case to the anomaly score. In this process, we used the pretrained language model as an encoder for the classifier. To define the labels for the cases, we selected the anomalies that we identified with the descriptive analysis (6.1.) and the token-based replay (6.2). In total, we flagged 250,328 cases as normal and 1,406 cases as anomalous. Since we had significantly fewer anomalies than normal cases, we had to manually ensure that a representative number of anomalies were both available in the training set and the validation set.

First, we trained our classifier for 3 epochs while freezing the layers of the language model. Afterward, we unfroze the layers of the language model piece by piece and trained them further with a lower learning rate. This procedure allows to benefit from the general process understanding of the language model and to fine-tune the classifier subsequently to the problem of anomaly detection.

target	pred
Create Purchase Order Item Record Goods Receipt xxbos Create Purchase Order Item Vendor creates invoice Record Invoice Receipt Remove Payment	Create Purchase Order Item Vendor Goods Receipt Vendor Create Purchase Order Item Vendor creates invoice Record Goods Receipt Record Payment
Clear Invoice xxbos Create Purchase Order Item Record Goods Receipt Vendor creates invoice Record Invoice Receipt Clear Invoice xxbos Create	Clear Invoice xxbos Create Purchase Order Item Vendor Goods Receipt Vendor creates invoice Record Invoice Receipt Clear Invoice xxbos Create
Receipt Record Invoice Receipt Clear Invoice xxbos Create Purchase Order Item Vendor creates invoice Record Goods Receipt Record Invoice Receipt	Receipt Record Invoice Receipt Clear Invoice xxbos Create Purchase Order Item Vendor creates invoice Record Goods Receipt Record Invoice Receipt
Receipt Clear Invoice xxbos Create Purchase Order Item Vendor creates invoice Record Invoice Receipt Record Goods Receipt Remove Payment Block	Receipt Clear Invoice xxbos Create Purchase Order Item Vendor creates invoice Record Goods Receipt Record Goods Receipt Remove Payment Block

Figure 15: Language model based next step prediction.

epoch	train_loss	valid_loss	accuracy	time
0	0.034394	0.035732	0.994465	01:51

epoch	train_loss	valid_loss	accuracy	time
0	0.034717	0.034180	0.994478	02:08

epoch	train_loss	valid_loss	accuracy	time
0	0.029757	0.034254	0.994478	03:10

Figure 16: Training phase of the classification of anomalous process cases along with accuracy on the validation set.

After training, we reached a classification accuracy for detecting the anomalous cases of 0.995. Figure 16 prints the training process and the validation accuracy of the classification. The results show that the presented approach can distinguish between normal and anomalous cases and confirms the anomalies from sections 6.1 and 6.2.

6.5 Root cause analysis

Next to the question of identifying anomalous process cases, it is also interesting to find the causes which led to the anomalies in the first place. To reveal those root causes we performed an attribute-based comparison between all normal process cases and all potential anomalous process cases. For example, we compared the vendors that were involved in normal process cases versus the vendors that were mainly involved in potential anomalous cases. If a vendor is only involved in conspicuous process cases, this is a strong indication that he has a direct impact on the process flow. Other attributes we have examined are the case item category, the activities and the involved users. We choose those attributes, since from a business perspective, we expect them to have the highest impact.

Looking at the attribute *case Vendor*, some conspicuous vendors to which the purchase document was sent can be identified. Figure 17 shows an excerpt of the anomaly detection. Here we see that for some vendors the deviation between abnormal and normal classified cases is quite high. Having a look at the users (Figure 18), we can see that there is one big deviation and thus a hint for a real anomaly. For this user, more measures should be taken to ensure process compliance.

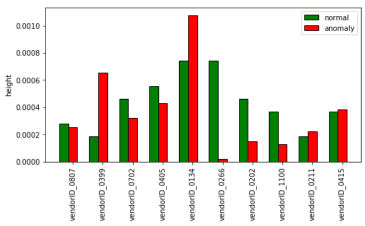


Figure 17: Excerpt results of root cause analysis for attribute "case Vendor"

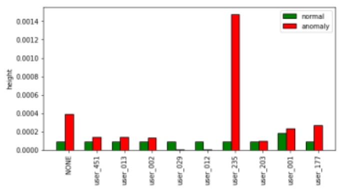


Figure 18: Excerpt results of root cause analysis for attribute "event User"

For the four different item categories, we see that no high deviations can be observed for 2-way matching and consignment. Especially for 3-way matching (invoice before GR) deviations exist, as shown in Figure 19. Doing the same for activities (Figure 20), anomalies for the two activities *Record Subsequent Invoice* and *Set Payment Block* can be observed. Compared to the other activities, the deviations are high. Thus, these activities must be reviewed.

6.6 Conclusion of the Anomaly Detection Process

In this chapter we have utilized established as well as innovative methods to identify a number of potentially anomalous process cases. For our investigations we elaborated on different methods considering case level attributes such as case item category and vendor as well as event level attributes such as time, amount and event type. In total, we were able to select 1406 potential anomalies. Subsequently, we used the selection of potential anomalies to locate different attribute deviations in the anomalies in order to uncover possible causes.

7 Conclusion

In this report, we describe our findings from the analysis of the purchase order handling process from a large coats and paints company operating from the Netherlands. We were tasked with analyzing the data in order to find conclusive answers to three leading questions regarding a collection of process models to properly represent the process, a technique to match line items, invoices and goods receipt in order to identify the throughput of a single instance, and the need for finding anomalous process behavior, i.e., instances that deviate from the prescribed process behavior. In the previous sections, we identified a collection of seven process models to describe the process. We developed a method that utilizes approximation techniques to measure the throughput time between specific activities. We combined different anomaly detection techniques to locate a selection of potential anomalous process traces and revealed different causes that led to the occurrence of the anomalies.

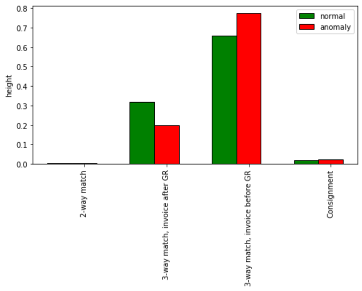


Figure 19: Excerpt results of root cause analysis for attribute "case Item Category"

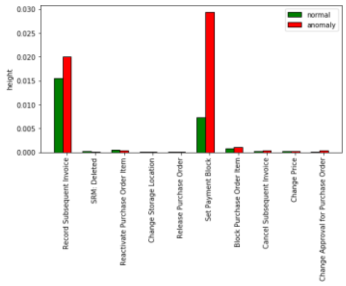


Figure 20: Excerpt results of root cause analysis for attribute "event concept:name"

Although we were able to explain some of the process behavior and context from freely-available SAP documentation, at some points we had to make assumptions. If we were to continue this analysis on a deeper level, we would prefer to talk directly to the process experts to gain an even deeper understanding of what this process entails. This would be particularly important for the second challenge, where we encountered peculiar behavior, but also for other features of the log.

Overall, we are convinced that the results from this challenge will help the case company to better understand their process, identify potential shortcomings, and optimize its purchase order handling in the future. We would like to express our gratitude to the case company for providing the data and to the BPIC committee for organizing this challenge.

References

Bose, R. P. J. C., & van der Aalst, W. (2009). Context Aware Trace Clustering: Towards Improving Process Mining Results. In *Proceedings of the 2009 SIAM International Conference on Data Mining* (pp. 401–412).

Carmona, J., Dongen, B. van, Solti, A., & Weidlich, M. (2018). *Conformance Checking*

- *Relating Processes and Models*. Springer International Publishing. https://doi.org/10.1007/978-3-319-99414-7

Howard, J., & Ruder, S. (2018). Universal language model fine-tuning for text classification. *ArXiv Preprint ArXiv:1801.06146*.

Leemans, S., Fahland, D., & van der Aalst, W. (2014). Discovering block-structured process models from event logs containing infrequent behaviour. In N. Lohmann, M. Song, & P. Wohed (Eds.), *Business Process Management Workshops* (pp. 66–78).

Mannhardt, F., De Leoni, M., & Reijers, H. A. (2017). Heuristic mining revamped: An interactive, data-Aware, and conformance-Aware miner. *CEUR Workshop Proceedings*, *1920*.

Nolle, T., Seeliger, A., & Mühlhäuser, M. (2018). BINet: Multivariate Business Process Anomaly Detection Using Deep Learning. In *International Conference on Business Process Management* (pp. 271–287).

Thaler, T., Ternis, S., Fettke, P., & Loos, P. (2015). A Comparative Analysis of Process Instance Cluster Techniques. *Proceedings Der 12. Internationalen Tagung Wirtschaftsinformatik (WI 2015)*, 423–437. Retrieved from http://www.tom-thaler.de/publications/Thaler2015_ProcessInstanceClustering.pdf

van der Aalst, W., Adriansyah, A., & van Dongen, B. (2011). Causal nets: A modeling language tailored towards process discovery. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *6901 LNCS*(1), 28–42. https://doi.org/10.1007/978-3-642-23217-6_3

van der Aalst, Wil. (2016). *Process Mining - Data Science in Action* (2nd ed.). Berlin, heidelberg: Springer Verlag. https://doi.org/10.1007/978-3-662-49851-4

Weijters, A. J. M. M., & Ribeiro, J. T. S. (2011). *Flexible heuristics miner (FHM)*. *IEEE SSCI 2011: Symposium Series on Computational Intelligence - CIDM 2011: 2011 IEEE Symposium on Computational Intelligence and Data Mining*. https://doi.org/10.1109/CIDM.2011.5949453

Weijters, A., van der Aalst, W., & de Medeiros, A. (2006). Process Mining with the HeuristicsMiner Algorithm. *Cirp Annals-Manufacturing Technology*, *166*(May), 1–34. https://doi.org/10.1.1.118.8288

Appendix A: Detailed table of the filtering process for the consignment sub log

Table 11: Filtering for start events of the consignment sub process

Name	Remarks
~~Change Quantity~~	Pending process with start before 2018
Create Purchase Requisition Item	
Create Purchase Order Item	

~~Record Goods Receipt~~	pending Process with start before 2018
~~Receive Order Confirmation~~	Pending Process with start before 2018, e.g. first event of 4508047120_00001 is on the 03.08.2018
~~Delete Purchase Order Item~~	Only Delete Purchase Order Item -> Create Purchase Order Item

Table 12: Filtering for end events of the consignment sub process

Name	Remarks
~~Create Purchase Order Item~~	Simultaneous start event when start in 12/2018
~~Change price~~	12 identical cases with 1 min delayed start each
~~Change Delivery Indicator~~	Many pending processes or processes at the end of 2018
~~Cancel Goods Receipt~~	Many pending processes or processes at the end of 2018
~~Change Quantity~~	Many pending processes or processes at the end of 2018
Delete Purchase Order Item	Possible end of process if something has been changed in the process (Update, Cancel, Change, Reactivate)
~~Reactivate Purchase Order Item~~	pending processes
~~Update Order Confirmation~~	pending processes
~~Receive Order Confirmation~~	pending processes, very often user 64, 30 29, 63, 65
Records Goods Receipt	
~~Change Storage Location~~	pending processes
Name	Remarks
Change Approval for Purchase Order	User_602, User_603 always executes identical steps
~~Create Purchase Order Item~~	Process at the end of 2018, end is not visible
Vendor creates invoice	
~~Vendor creates debit memo~~	Possible intermediate event with long lead time

Table 13: Filtering for start events of the 2-way-matching sub process

Name	Remarks
Change Approval for Purchase Order	User_602, User_603 always executes identical steps
~~Create Purchase Order Item~~	Pending processes at the end of 2018
Vendor creates invoice	
~~Vendor creates debit memo~~	Possible intermediate event with long lead time

Table 14: Filtering for end events of the 2-way-matching sub process

Name	Remarks
~~Change Approval for Purchase Order~~	Process at the end of 2018, end is not visible
Clear Invoice	
~~Create Purchase Order Item~~	Process at the end of 2018, end is not visible
~~Delete Purchase Order Item~~	Pending processes, often including user 602
Records Invoice Receipt	Clear accumulation in the middle of the log period
~~Set Payment Block~~	Steps which are often done by user 602 603, pending processes
~~Vendor creates invoice~~	Here follows the activity on clear invoice and record invoice receipt

Table 15: Filtering for start events of the 3-way-matching after GR sub process

Name	Remarks
Create Purchase Order Item	
Create Purchase Requisition Item	Ok but always before *Create Purchase Order Item*
Vendor creates debit memo	Part of time overlapping processes
Vendor creates invoice	

Table 16: Filtering for end events of the 3-way-matching after GR sub process

Name	Remarks
Record Service Entry Sheet	
Record Goods Receipt	Pending and not yet finished processes
Record Invoice Receipt	Pending?
~~Create Purchase Order Item~~	Pending and not yet finished processes

Clear Invoice	
~~Remove Payment Block~~	Mostly after record invoice receipt
~~Change Price~~	Pending processes
~~Change Quantity~~	Pending processes at the end of the timeframe
~~Cancel Goods Receipt~~	Pending processes
~~Cancel Invoice Receipt~~	Pending processes
~~Vendor creates debit memo~~	Pending processes
Delete Purchase Order Item	Ok, but to less cases
Change Delivery Indicator	Pending processes, but also potential end if indicator has been set to 0
Set Payment Block	Ok but to less cases
Change Approval for Purchase Order	Only after the deletion of a purchase order item
Cancel Subsequent Invoice	Only after *clear invoice*
~~Reactivate Purchase Order Item~~	To less cases
~~Change Final Invoice Indicator~~	To less cases, but possible end event

Table 17: Filtering for start events of the 3-way-matching before GR sub process

Name	Remarks
Create Purchase Order Item	
Create Purchase Requisition Item	Ok but always before *Create Purchase Order Item*
~~Vendor creates debit memo~~	Part of time overlapping processes
Vendor creates invoice	

Table 18: Filtering for end events of the 3-way-matching before GR sub process

Name	Remarks
~~Block Purchase Order Item~~	Pending process, often after *Change Approval for Purchase Order*
Cancel Goods Receipt	Ok -> Abort
Cancel Invoice Receipt	Ok -> Abort
Cancel Subsequent Invoice	Ok -> Manual Billing process
~~Change Approval for Purchase Order~~	Pending process
~~Change Currency~~	Pending process
Change Delivery Indicator	ok
~~Change Final Invoice Indicator~~	Ok, to less cases
Change Price	Ok, possible post Invoice adaption
Change Quantity	Ok, possible post Invoice adaption

~~Change Storage Location~~	Pending process
Change payment term	Pending process
Clear Invoice	
~~Create Purchase Order Item~~	Pending process
Delete Purchase Order Item	
~~Reactivate Purchase Order Item~~	Pending process
~~Receive Order Confirmation~~	Pending process
~~Record Goods Receipt~~	Pending process
Record Invoice Receipt	
~~Record Subsequent Invoice~~	Ok, but less cases
~~Release Purchase Order~~	Pending process
Remove Payment Block	
~~Update Order Confirmation~~	Pending process
Vendor creates debit memo	
Vendor creates invoice	

Appendix B: Detailed table of the filtering process for the consignment sub log

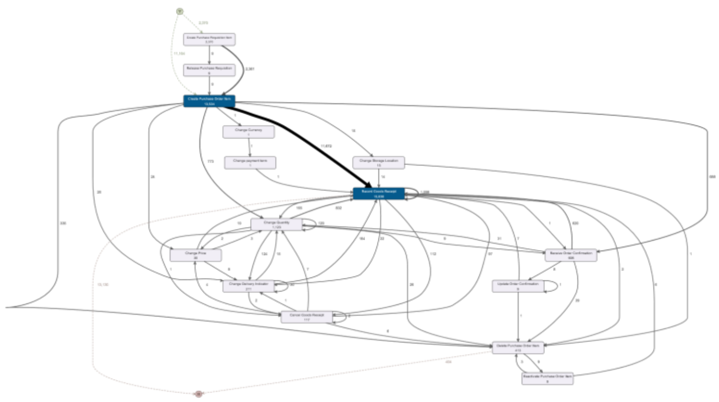


Figure 21: Process Discovery of the manually filtered consignment sub log using DISCO (100% of all activities and paths)

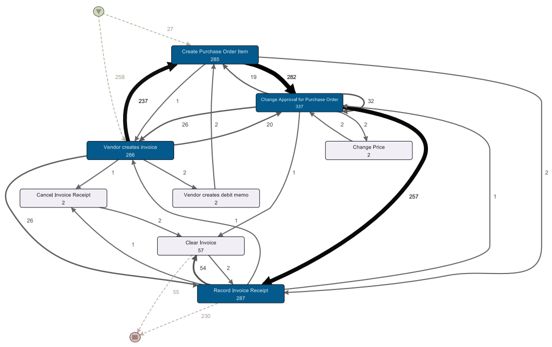


Figure 22: Process Discovery of the manually filtered 2-way-matching sub log using DISCO (100% of all activities and paths)

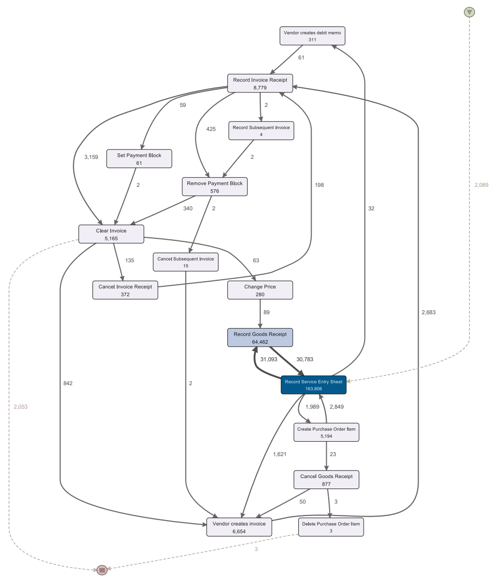


Figure 23: Process model discovered for 3-way-after service line items

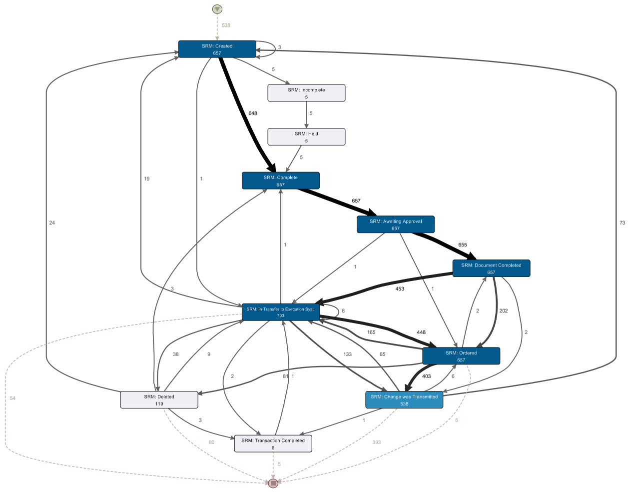


Figure 24: Process Discovery of the manually filtered 3-way-matching after GR SRM sub log using DISCO (100% of all activities and paths)

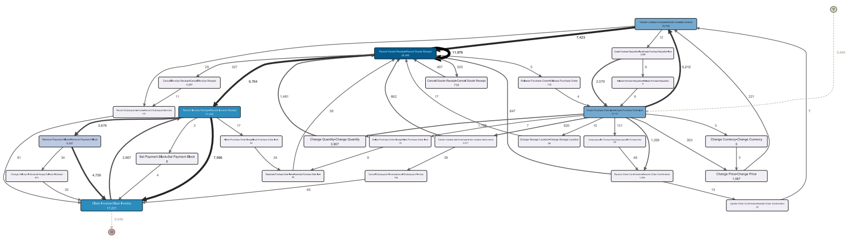


Figure 25: Discovered process model for 3-way-before-items with multiple invoices and GR

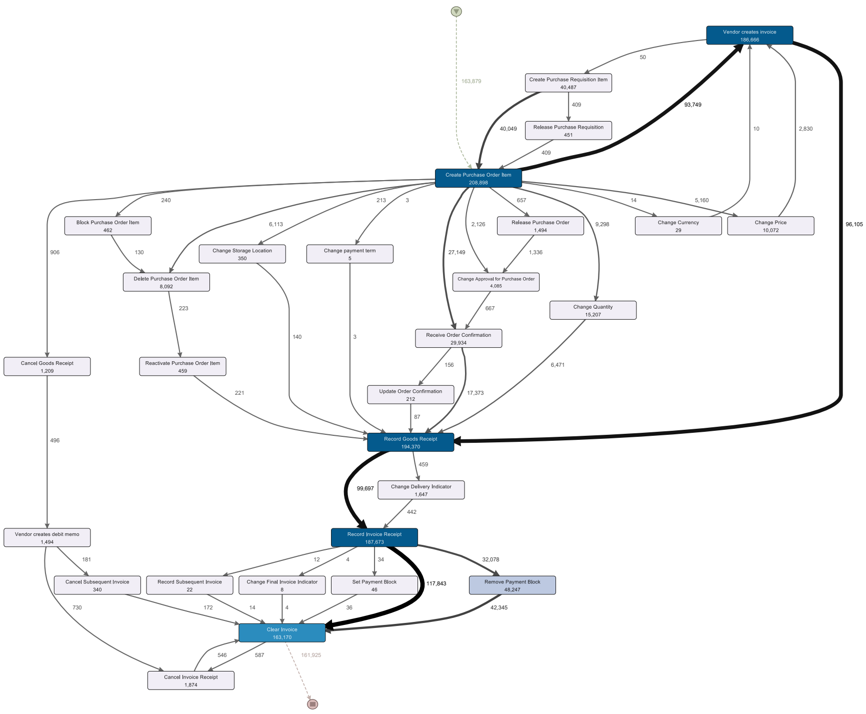


Figure 26: Discovered process model for 3-way-before items with one invoice and one GR

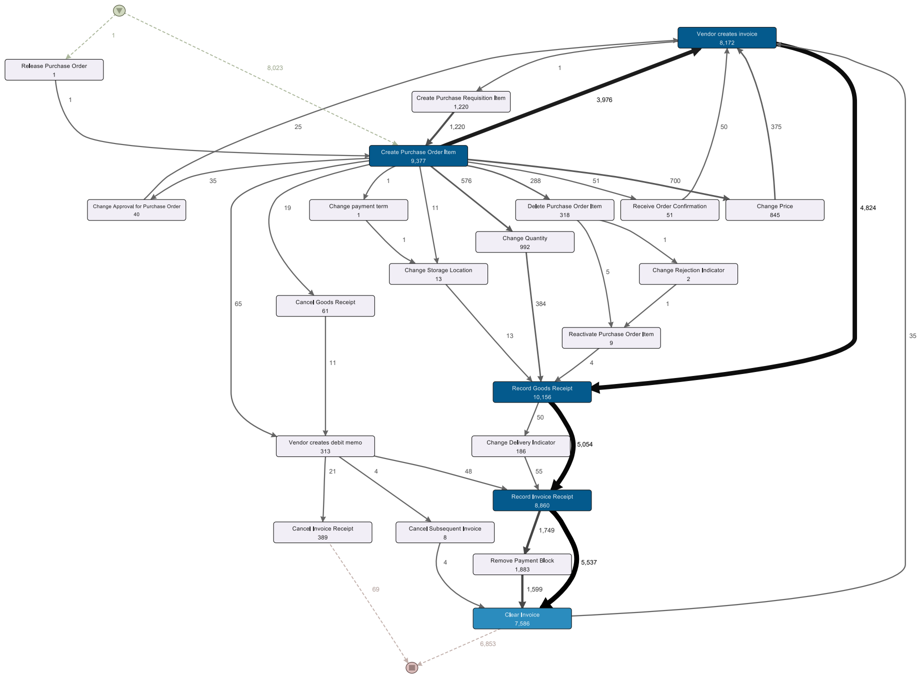


Figure 27: Process model discovered for 3-way-after non-service items

NINTH INTERNATIONAL BUSINESS PROCESS INTELLIGENCE CHALLENGE

Submission in The Non-Student Category

Aiming on completeness of analysis and usefulness for the purpose of a real-life process mining setting.

submitted by

Albert Kisjes

akisjes@agilos.nl

+31 6 5585 3729

Jordy Bekker

jbekker@agilos.nl

+31 6 1192 0648

Process Mining Conference 2019

1st International Conference on Process Mining, June 24-26, 2019, Aachen, Germany

ICPM

Contents

1. Abstract

For the ICPM Process Mining contest an anonymized dataset is available. This dataset has the structure of a cases and events table and is from a large multinational company operating from the Netherlands. The field of expertise of this company is coatings and paints and their data contain information of 4 of their 60 subsidiaries. Related companies are AkzoNobel and PPG paints.

With the Agilos Analytics 4 Improvement (A4I) Process Mining application (powered by ProcessGold) we conducted an analysis on this data about their procure to pay process (P2P).

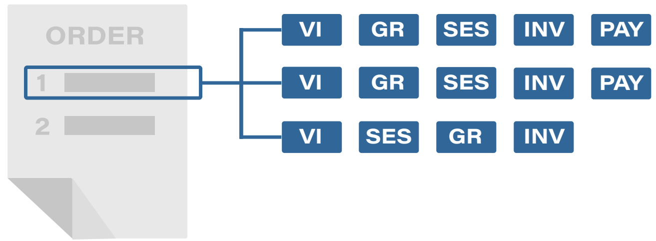
The goal of this report is to help this company understand their processes and potentially help them prevent faults or inefficiencies that would otherwise have gone unnoticed. Special care is given to compliancy, as it is given that the process owner has compliance questions.

To effectuate beforementioned goal, three main topics of interest are reviewed.

1. The process models that can be distinguished from the data.
2. The throughput of the invoicing process.
3. Deviations from the standard process.

1. The process models that can be distinguished from the data

We performed both process intelligence and business intelligence on the data set. This allowed us to identify the discriminating properties that together result in a set of process models, which together assist in understanding and improving the overall process.



Process Intelligence

Throughout this analysis we have worked with three different case ID's (as shown above) to analyze the process from different angles, namely:

- Purchase order, which holds all the purchasing line items.
- Purchase order line item, which for this contest is defined as Case ID.
- Invoicing process iteration, which is created by us to determine the throughput of the invoicing process.

As indicated before, we decided upon incorporating Business Intelligence (BI) and Process Intelligence (PI) to construct different process models. We used the beforementioned invoicing process iteration, to identify the process models that comprise of more than one iteration. The Agilos Analytics 4 Improvement app "A4I App" powered by ProcessGold enables us to easily switch from Case ID and to combine the PI-view with the BI-view.

Not only have we identified the models with multiple iterations. We also divided the process into only 'primary' activities and with 'other' activities, which we also used for the business intelligence view. The following primary activities are identified.

1. Create Purchase Requisition Item
2. Create Purchase Order Item
3. Vendor Creates Invoice
4. Record Invoice Receipt
5. Record Goods Receipt
6. Record Service Entry Sheet
7. Clear Invoice

We have divided our process models by iterations and by exclusively following the primary activities, but what we also did is determine whether the process started with a purchase requisition (PR) and combine all this with the item type field. The result is the process models as shown below.

▲ Process model	Total
Consignment	978
Limit	95
Other orders, 1 iterations with other activities	414.591
Other orders, 1 iterations without other activities	406.824
Other orders, >1 iterations with other activities	77.897
Other orders, >1 iterations without other activities	17.701
Other PR orders, 1 iterations with other activities	69.309
Other PR orders, 1 iterations without other activities	70.014
Other PR orders, >1 iterations with other activities	12.479
Other PR orders, >1 iterations without other activities	3.816
Service	81.011
Total	1.154.715

Figure 1: Activities per process model (filtered by closed cases)

2. The throughput of the invoicing process

To analyze the invoicing process, we only focused on the relevant activities. As mentioned before we created an invoicing process iteration that we can use as Case ID.

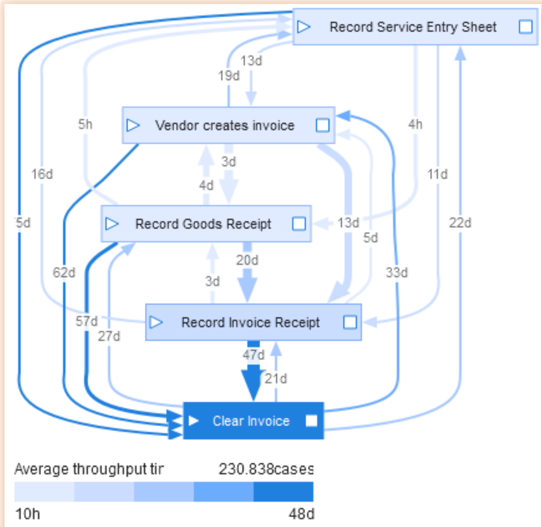


Figure 2: Invoicing process throughput times

With this we created the following graph, which shows the average throughput times of the largest process model.

We filtered out the cases with a wrong starting activity and the incomplete cases, this is explained in chapter 3.

To gain insight in the process models defined as before, we use BI to see the average throughput times of each of the models. As shown on the right, the orders with PR (depicted as 'Other PR orders…') are on average considerably faster than their non-PR counterpart!

An important conclusion is that where the process flow only contains the seven primary activities (depicted by '…without other activities'), the throughput time is lower. This means the Lean mantra First Time Right proves to pay off from this perspective!

▲ Process model	Average throughput	Iteration Count	Sum iteration Vendor Invoice amount
Consignment	0s	8	€ 0
Limit	34d	17	€ 282.418
Other orders, 1 iterations with other activities	76d	63.388	€ 250.909.168
Other orders, 1 iterations without other activities	64d	81.365	€ 179.130.980
Other orders, >1 iterations with other activities	57d	18.775	€ 84.873.291
Other orders, >1 iterations without other activities	40d	7.576	€ 21.766.611
Other PR orders, 1 iterations with other activities	63d	8.827	€ 45.213.386
Other PR orders, 1 iterations without other activities	53d	11.669	€ 26.864.529
Other PR orders, >1 iterations with other activities	42d	2.883	€ 7.977.647
Other PR orders, >1 iterations without other activities	35d	1.518	€ 3.607.603
Service	18d	34.812	€ 635.461.163
		230.838	€ 1.256.086.796

Figure 3: Average throughput time per process model

The following conclusions can be drawn from the table above

- Processes that start with a PR have a much lower throughput time than processes without a PR. We note that the deployment of PR way of working (which started in September 2018 within this company) paid off from this perspective
- The throughout time is significantly lower when no additional activities have to be carried out (like price changes, payment blocks, etc.)
- Service orders have a significant lower throughput time than non-service orders

3. Deviations from the standard process

We have successfully identified various process models and invented a technique to analyze the invoicing process. Now, deviations can be easily identified and considered for further analysis and improvement. Our philosophy is nonetheless, analytics for improvement.

During our analysis we consistently use the PO amount as a reality check for our conclusion. We found out that for the other amounts (for instance clear invoice) the total amounts were much higher, as a result of the many iterations. This influenced our decision on how to approach this challenge. Normally we align the amounts and check for variances, this was not as effective with the dataset used for this challenge.

Normally in a Process Mining project, intensive communication with the client is required to be successful. However, during this contest no communication with the client was possible or allowed. During the analysis of the dataset a lot of questions and potential issues surfaced, and this list is too long to include in this report. We therefore decided to mainly focus on the issues where we think it potentially has the most impact for the company.

The main areas covered in chapter 5 of this report are:

- (Potential) circumvention of segregation of duties
- Anomalies in the process flows
- Anomalies in the event amounts
- Data Quality

Dataset used: van Dongen, B.F., Dataset BPI Challenge 2019. 4TU.Centre for Research Data. https://doi.org/10.4121/uuid:d06aff4b-79f0-45e6-8ec8-e19730c248f1

2. Methodology

For this assignment we deployed the Agilos Analytics 4 Improvement methodology which is powered by the Agilos A4I Application. This application is customized on the ProcessGold Data Visualization & Process Mining platform.

On the right the DMAIC approach is displayed where the different steps are brought into the perspective of the Analytics for Improvement approach.

This allows for a combined business insight (pivoting on the output e.g. on PO Values, or on other value dependent on the quality of the data preparation) and process performance insight through a combination of the extensive process mining capabilities.

The A4I App also includes a dashboard which both shows the actual values of the filtered data as well as a footer with the total values of the unfiltered dataset. This footer also shows the percentage, see below.

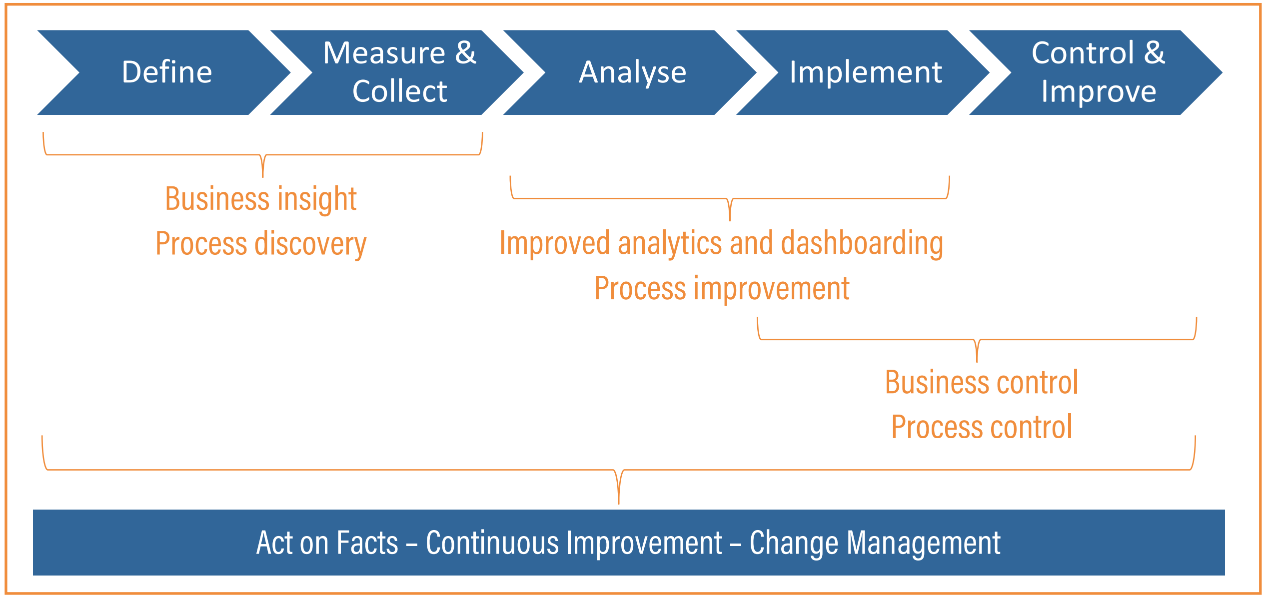


Figure 5: DMAIC applied in the Analytics for Improvement approach

PO's	PO lines	Process iterations	Events	PO amount	Activities	Variants	Vendors	Items	Users
53.846	186.277	230.838	1.154.715	€ 670.231.046	40	6.833	1.489	322	587
Σ 76.349 | 71%	Σ 251.734 | 74%	Σ 422.605 | 55%	Σ 1.595.923 | 72%	Σ € 996.818.747 | 67%	Σ 42 | 95%	Σ 11.973 | 57%	Σ 1.975 | 75%	Σ 490 | 75%	Σ 628 | 93%

Figure 4: A4I footer. Closed purchase order line items filter applied

3. Process models

PO's	PO lines	Process iterations	Events	PO amount	Activities	Variants	Vendors	Items	Users
76.349	251.734	422.606	1.595.923	€ 996.818.747	42	11.973	1.975	490	628
Σ 76.349 | 100%	Σ 251.734 | 100%	Σ 422.606 | 100%	Σ 1.595.923 | 100%	Σ € 996.818.747 | 100%	Σ 42 | 100%	Σ 11.973 | 100%	Σ 1.975 | 100%	Σ 490 | 100%	Σ 628 | 100%

Figure 6: A4I footer. All purchase order line items

1. Getting to know the data

To understand the volume and the complexity of the dataset an overview is needed. An overview is generated by the A4I application and shown above, which allows us to get a general idea about the dataset. It tells us that we are looking at 42 different process steps performed with almost 1.6 million events that were performed by 628 unique users. This concerns more than 76 thousand purchase orders that were placed at almost 2 thousand vendors for 490 unique items, totaling an amount of almost 1 billion euros. This process was performed in almost 12 thousand different ways, as calculated from more than 250 thousand purchase order line items.

We now have a general idea about the data, let's look at how the purchase order is built up. See the illustration below. The purchase order is a document which is linked to a vendor and consist of line items. The line items hold information about individual goods or services that are up for order. Each line calculates the total amount by multiplying the number of units with the price per unit. Especially in the case of services, a purchase order line item follows the same process or almost the same process multiple times. This is what we call a process iteration.

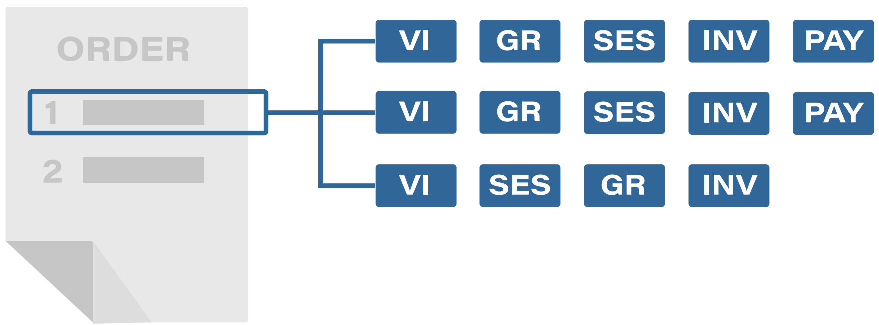


Figure 7: Structure of a purchase order

When we are talking about an invoicing process, we should consider scoping our activities to only the relevant ones for this process. We identified the following activities:

- Vendor creates invoice (VI)
- Record Goods Receipt (GR)
- Record Service Entry Sheet (SES)
- Record invoice receipt (INV)
- Clear invoice (PAY)

Below the five biggest variants are shown. A variant is a sequence of process steps as performed for each purchase order line item that belong to that variant. The five below are all different sequences. Together, they cover 45% of all the cases. Note that a small portion of sequences explain how the majority of the purchase order line items are handled.

For this analysis, a purchase order line item is followed through the procure to pay process. For each of the five variants you can see how a purchase order line item is being processed. Variant 1 is the biggest sequence the dataset that covers around 1/5th of all the purchase order line items. It accounts for almost €120 million in orders.

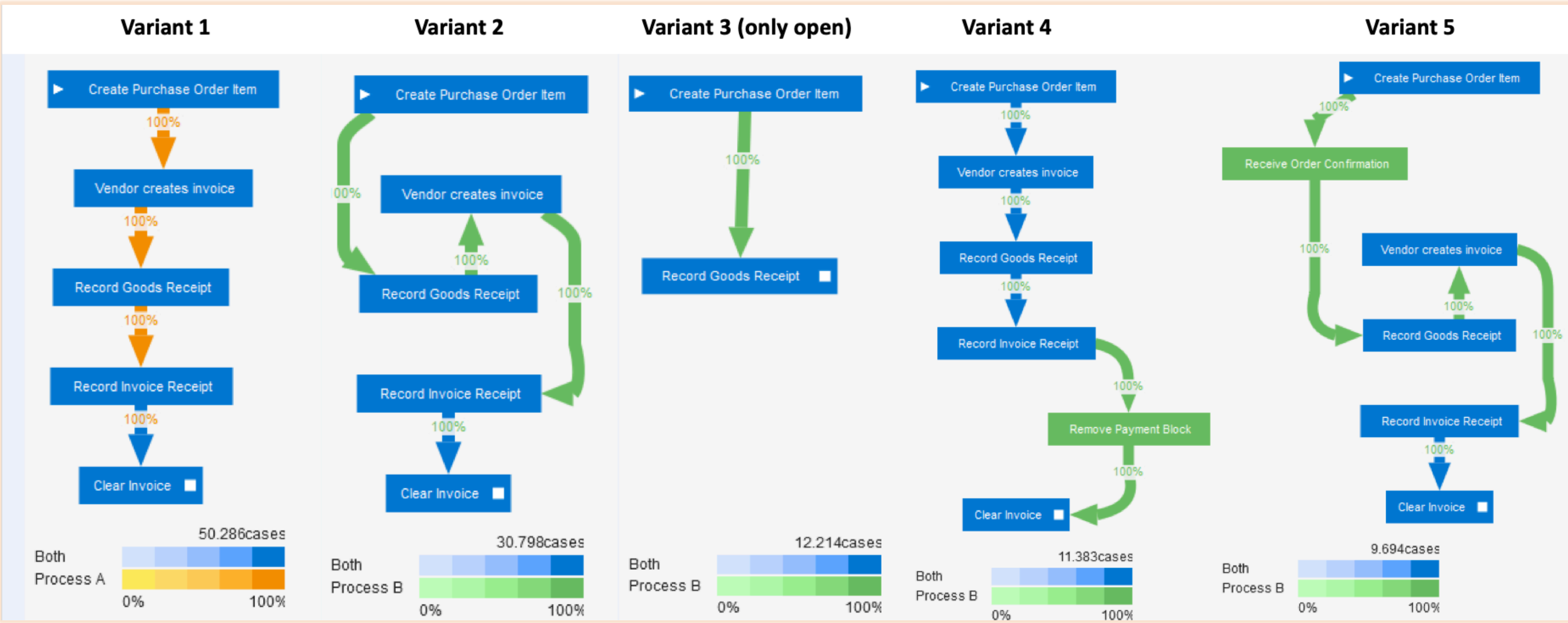


Figure 8: Graphs of the 5 most frequent process variants

2. Finding the closed cases

Processes should have a clear beginning and ending, because to assess which different process models there are, we should only look at the processes which have both a beginning and ending. This means we want to filter the data to only see the processes that are executed from cradle to grave, so we can effectively use them to model the different processes. Other processes are either incomplete or have started incorrectly.

To realize this filtering a technique is used to categorize processes in either open cases, closed cases and incorrect start cases. This is done by identifying firstly the possible activities from which the process could logically start, and secondly by identifying at which activities the process comes to a natural ending.

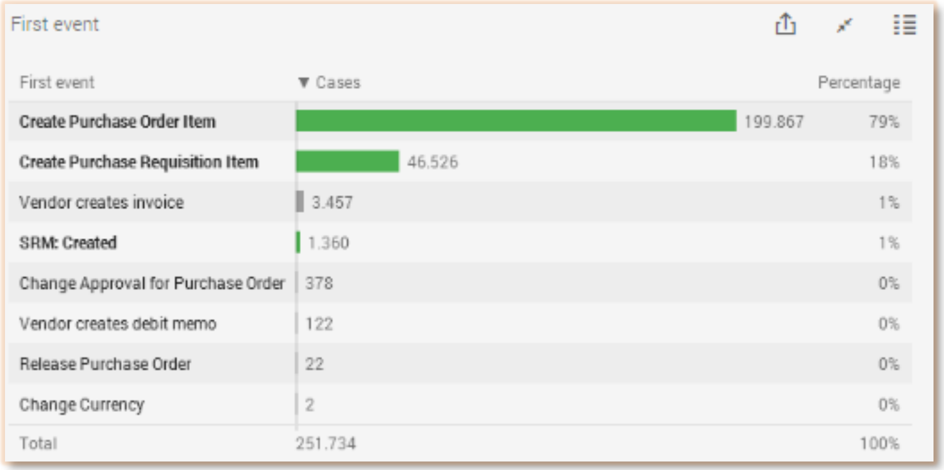


Figure 9: First activity of all purchase order line items

By analyzing the complete dataset, we are able to see which activities started the process. This start is based on the timestamp, logically. Above an overview is shown of the starting activities. It says that 79% of the purchase orders start with the activity "Create Purchase Order Item". In 18% of the cases this is "Create Purchase Requisition Item". Although this covers almost the entire dataset, we found out that for a small portion of the purchase orders a SRM-process is in place which precedes the creation of a purchase order. The rest (2% of the cases) is regarded as an incorrect start case.

We therefore defined the starting activities as such, which can be seen on the image above by the green color and the bold font that has been used.

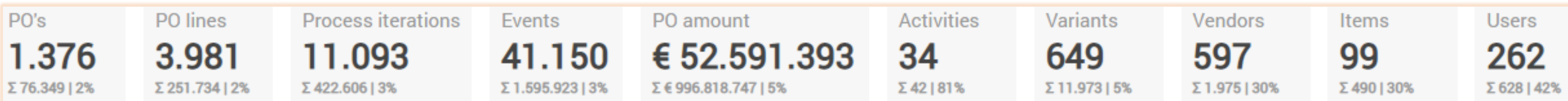


Figure 10: A4I footer. Purchase order line item with incorrect start activity

When we look at our last activity within a purchase order line item the data shows us the picture at the right. This picture only shows the top 10 end activities, in reality the list is longer. We see that 72% of the purchase order line items end with "Clear Invoice", which is a reasonable end definition for this process. When a purchase order line item is made, but later deleted because of some reason a "Delete Purchase Order Item" also is an understandable last step in the process. Even though this step covers just 3% of the cases. The rest (24% of the cases) is regarded as open case, these processes are simply not done yet.

The abovementioned activities are defined as end activities, can be seen on the image on the right above by the red color and the bold font that has been used.

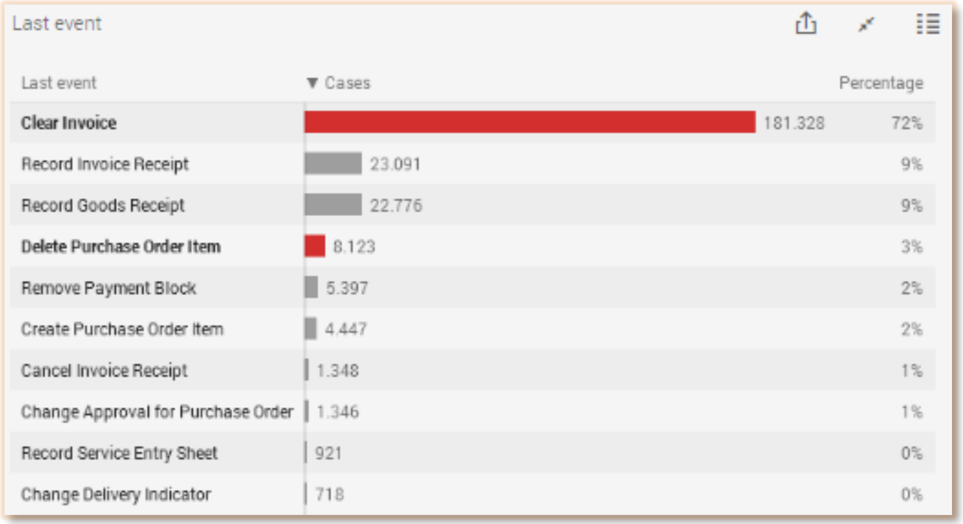


Figure 11: Top 10 last activity of all purchase order line items

We have now identified the purchase order line items with incorrect start activity and the open purchase order line items, what remains are the line items with the abovementioned start activity and end activity. We call these closed purchase order line items and, on the right we can see how many of those are found in each subsidiary.

▲ Company	Closed	Incorrect start	Open	Total
companyID_0000	186.255	3.226	61.205	250.686
companyID_0001	2	0	0	2
companyID_0002	2	0	0	2
companyID_0003	18	755	271	1.044
Total	186.277	3.981	61.476	251.734

Figure 13: Closed purchase order line items per subsidiary

PO's	PO lines	Process iterations	Events	PO amount	Activities	Variants	Vendors	Items	Users
24.091	61.476	180.676	400.058	€ 273.996.308	39	4.491	1.309	438	538
Σ 76.349 | 32%	Σ 251.734 | 24%	Σ 422.606 | 43%	Σ 1.595.923 | 25%	Σ € 996.818.747 | 27%	Σ 42 | 93%	Σ 11.973 | 38%	Σ 1.975 | 66%	Σ 490 | 66%	Σ 628 | 86%

Figure 12:A4I footer. Open purchase order line items

PO's	PO lines	Process iterations	Events	PO amount	Activities	Variants	Vendors	Items	Users
53.846	186.277	230.839	1.154.715	€ 670.231.046	40	6.833	1.489	322	587
Σ 76.349 | 71%	Σ 251.734 | 74%	Σ 422.606 | 55%	Σ 1.595.923 | 72%	Σ € 996.818.747 | 67%	Σ 42 | 95%	Σ 11.973 | 57%	Σ 1.975 | 75%	Σ 490 | 75%	Σ 628 | 93%

Figure 14: A4I footer. Closed purchase order line items

3. Confronted with a difficult choice

When only looking at the closed line items, we are left with almost 7 thousand different sequences in which the activities were performed. Because of our filtering we have 43% less variants to look at. These 6.8 thousand variants should be divided in different process models.

Of the 6.8 thousand variants, the top 4 (based on purchase order line item amount) are summarized in the below table on the right. It also shows multiple item types (as shown in the columns) follow the same process. As we can see, the item category " 3-way match, invoice before GR" and " 3-way match, invoice after GR" are both in variant 1 and 2. In our understanding this attribute should show whether the activity "Record Invoice Receipt" is either before or after "Record Goods Receipt". This would mean that both variants should have the attribute "3-way match, invoice after GR".

Variant 1 and variant 2 together are responsible for 26% of the closed purchase order line item amount, the graphs are shown on the right.

We had to make a difficult choice between:

- Process models based on available attributes (e.g. item type), to later see which variants (based on the available process steps) do not comply; or
- Process models based on available process steps (e.g. create purchase requisition item), to later see which attributes do not fit in this model.

We did both.

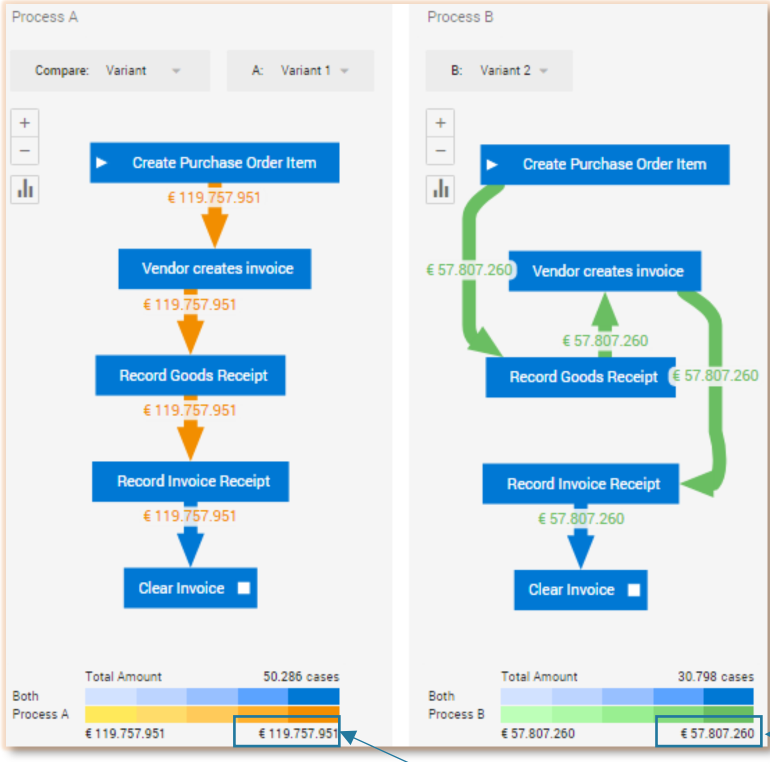


Figure 15: Two biggest variants compared (based on PO amount)

Variant	2-way match	3-way match, invoice after GR	3-way match, invoice before GR	Consignment	▼ Total
Variant 1	€ 0	€ 7.926.881	€ 111.831.070	€ 0	€ 119.757.951
Variant 2	€ 0	€ 1.606.868	€ 56.200.392	€ 0	€ 57.807.260
Variant 4	€ 0	€ 3.833.462	€ 28.036.916	€ 0	€ 31.870.378
Variant 10	€ 0	€ 321.463	€ 29.241.367	€ 0	€ 29.562.830

Figure 16: Top 4 (closed) variants in terms of PO amount

4. Combine process intelligence with business intelligence

Based on the insight of both business intelligence as well as process intelligence we decided to define a collection of process models based on both attributes as well as the frequency and existence of certain activities. And we realized this split could not be used for variant clustering as some variants within one or more process models as defined by us. The table on the right summarizes the process models identified and is based on the views provided by the A4I app.

The summary of this table on the right provides some interesting insights: 69.7 % of the Non-PR cases with 1 iteration is responsible for 54,7 % of PO value but is processed with only 7% of the cases. On the other hand, 57,2 % of the Non-PR cases with more iterations only accounts for 4,4 % of the PO amount value. Consignment Po's have no PO value.

				# of events								
Sort	Process model	Sum of # cases	Sum of PO Amount	Sum of 1. Create Purchase Requisition Item	Sum of 2. Create Purchase Order Item	Sum of 3. Vendor creates invoice	Sum of 4. Record Invoice Receipt	Sum of 5. Record Goods Receipt	Sum of 6. Record Service Entry Sheet	Sum of 7. Clear Invoice	Sum of Other activities	Sum of Total
1.1 Non PR	Other orders, 1 iterations without other activities	96.717	€ 211.202.388	0	2.355	3.188	3.410	8.727	0	2.604	0	20.284
1.2 Non PR	Other orders, 1 iterations with other activities	78.703	€ 333.931.314	0	8.787	12.131	16.424	16.398	0	14.467	23.542	91.749
2.1 Non PR	Other orders, >1 iterations without other activities	2.355	€ 19.216.837	0	96.717	89.680	90.785	94.169	0	83.239	0	454.590
2.2 Non PR	Other orders, >1 iterations with other activities	8.787	€ 66.963.911	0	78.703	68.586	68.847	69.683	0	64.732	108.167	458.718
3.1 PR	Other PR orders, 1 iterations without other activities	21.303	€ 70.368.403	627	627	845	791	2.545	0	512	0	5.947
3.2 PR	Other PR orders, 1 iterations with other activities	20.259	€ 103.438.892	1.603	1.603	2.199	2.816	3.249	0	1.974	3.661	17.105
4.1 PR	Other PR orders, >1 iterations without other activities	627	€ 5.219.961	21.303	21.303	18.050	18.055	20.048	0	11.694	0	110.453
4.2 PR	Other PR orders, >1 iterations with other activities	1.603	€ 9.523.692	20.259	20.259	16.971	16.970	18.188	0	8.851	25.996	127.494
5.	Service	5.838	€ 163.740.191	0	5.838	7.580	9.967	65.631	164.975	6.013	7.597	267.601
8.	Limit	1.044	€ 13.213.158	0	1.044	689	695	0	0	307	3.163	5.898
9.	Consignment	14.498	€ 0	2.800	14.498	0	0	15.459	0	0	3.327	36.084
Grand Total		**251.734**	**€ 996.818.747**	**46.592**	**251.734**	**219.919**	**228.760**	**314.097**	**164.975**	**194.393**	**175.453**	**1.595.923**

Figure 17: The process models in relation with the primary activities

Relation between # cases. PO amounts and # events		**Sum of # cases**	**Sum of PO Amount**	**# events**	**% of total cases**	**% of total amount**	**% of # events**
1.1 and 1.2	Non PR: 1 iteration	175.420	€ 545.133.702	112.033	69,7%	54,7%	7,0%
2.1 and 2.2	Non PR more interations:	11.142	€ 86.180.748	913.308	4,4%	8,6%	57,2%
3 and 3	PR	43.792	€ 188.550.948	260.999	17,4%	18,9%	16,4%
5.	Service	5.838	€ 163.740.191	267.601	2,3%	16,4%	16,8%
8.	Limit	1.044	€ 13.213.158	5.898	0,4%	1,3%	0,4%
9.	Consignment	14.498	€ 0	36.084	5,8%	0,0%	2,3%
		251.734	**€ 996.818.747**	**1.595.923**	**100,0%**	**100,0%**	**100,0%**

Figure 18: Condensed overview of the process models

4. Throughput

PO's	PO lines	Process iterations	Events	PO amount	Activities	Variants	Vendors	Items	Users
53.054	183.293	247.173	1.199.622	€ 736.467.307	41	8.286	1.653	320	589
Σ 76.349 | 69%	Σ 251.734 | 73%	Σ 422.606 | 58%	Σ 1.595.923 | 75%	Σ € 996.818.747 | 74%	Σ 42 | 98%	Σ 11.973 | 69%	Σ 1.975 | 84%	Σ 490 | 84%	Σ 628 | 94%

Figure 19: A4I footer. Purchase order line items with invoice receipt and payment

1. Double invoice and payment?

For each purchase order line item where an invoice and payment has been registered, the throughput can be analyzed. So, to only look at the purchase order line items where the activity "Record Invoice Receipt" and "Clear Invoice" are present we applied a filter in our A4I application. In above overview it is shown that for 73% of the purchase order line items we can analyze the invoicing process.

Before we dive into the throughput analysis, a check is done to see if certain process steps are repeated within a specific purchase order line item. If we expand above filter by adding the prerequisite that either an invoice receipt or a payment activity has been performed more than once within a specific purchase order line item, 2% of the purchase order line items remain as shown in below overview.

The easiest thing to do is to exclude these cases from the throughput analysis. Even though these purchase order line items are just a fraction of the total, the total net value of the orders are quite sizable with 15% of the total.

Because of the sizable net amount, it is not feasible to exclude the abovementioned process variants from our throughput analysis. So as stated in the challenge a technique is sought to match the events within a line item. To understand what is happening a short recap is in place. We have almost 6 thousand purchase order line items, divided over almost 3 thousand purchase orders. Within these purchase order line items multiple process iterations occur, in total more than 23 thousand. Visually this looks like the illustration on page 8.

We have looked for a way to identify the different process iteration, and here we are only interested in the invoicing process. To reduce the complexity, we only looked at the following activities:

- Vendor creates invoice;
- Record Goods Receipt;
- Record Invoice Receipt;
- Record Service Entry Sheet; and
- Clear Invoice.

PO's	PO lines	Process iterations	Events	PO amount	Activities	Variants	Vendors	Items	Users
2.941	5.946	23.117	99.516	€ 144.746.267	37	2.874	639	93	367
Σ 76.349 | 4%	Σ 251.734 | 2%	Σ 422.606 | 5%	Σ 1.595.923 | 6%	Σ € 996.818.747 | 15%	Σ 42 | 88%	Σ 11.973 | 24%	Σ 1.975 | 32%	Σ 490 | 32%	Σ 628 | 58%

Figure 20: A4I footer. Purchase order line items that have multiple invoices and payments

The next step in identifying the different process iterations is numbering each first occurrence of the beforementioned activities individually as 001, the second as 002, the third as 003, etc. Together with the purchase order line item we now created a process iteration ID.

2. First time right

After applying the identifyer for each iteration it is possible to analyze the throughput of each process model. See below. Important conclusion is that where the processflow only contains the seven primary activities, the throughput time is lower.

▲ Process model	Average throughput	Iteration Count	Sum iteration Vendor Invoice amount
Consignment	0s	8	€ 0
Limit	34d	17	€ 282.418
Other orders, 1 iterations with other activities	76d	63.388	€ 250.909.168
Other orders, 1 iterations without other activities	64d	81.365	€ 179.130.980
Other orders, >1 iterations with other activities	57d	18.775	€ 84.873.291
Other orders, >1 iterations without other activities	40d	7.576	€ 21.766.611
Other PR orders, 1 iterations with other activities	63d	8.827	€ 45.213.386
Other PR orders, 1 iterations without other activities	53d	11.669	€ 26.864.529
Other PR orders, >1 iterations with other activities	42d	2.883	€ 7.977.647
Other PR orders, >1 iterations without other activities	35d	1.518	€ 3.607.603
Service	18d	34.812	€ 635.461.163
		230.838	€ 1.256.086.796

Figure 21: Activities per process model (filtered by closed cases)

This means that the Lean mantra First Time Right proves to pay off from this perspective!

Below the average throughput time is shown together with the number of cases for the different process models. As an example, we also included a graph of the throughput time analysis of the largest process model (meaning that it contains the most purchase order line items).

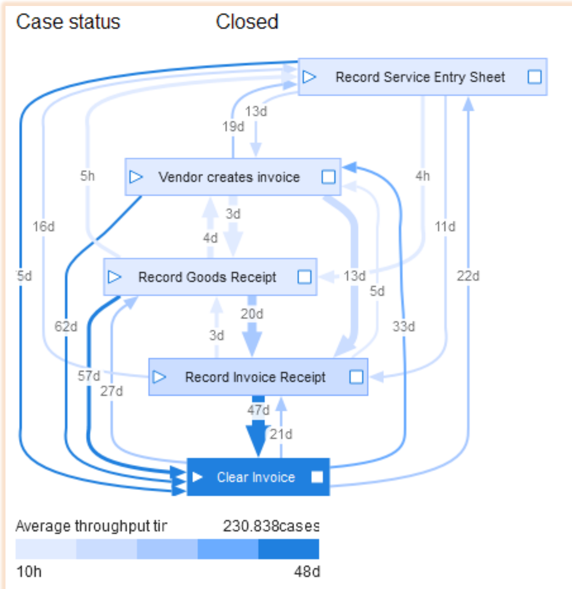


Figure 22: Invoicing process throughput times

The following conclusions can be drawn from the table below on the left:

- Processes that start with a PR have a much lower throughput time than processes without a PR. We note that the deployment of PR way of working (which started in September 2018 within this company) paid off from this perspective
- The throughput time is significantly lower when no additional activities have to be carried out (like price changes, payment blocks, etc.)
- Service orders have a significant lower throughput time than non-service orders

5. Deviations

In this chapter various deviations are laid out. Please note that not every deviation is reported, just some points of focus which are in our opinion worth investigating.

1. (Potential) circumvention of segregation of duties

Certain users perform certain activities within a case. Some activities should not be done by the same user. The table on the right contains an example of users that executed both the invoice receipt and the payment within the same case ID flow.

User	User activities per case	-	▼ Total
user_013	Cancel Invoice Receipt, Clear Invoice, Record Invoice Receipt	€ 1.534.265	€ 1.534.265
user_004	Cancel Invoice Receipt, Clear Invoice, Record Invoice Receipt	€ 1.204.825	€ 1.204.825
user_015	Cancel Invoice Receipt, Clear Invoice, Record Invoice Receipt	€ 1.161.943	€ 1.161.943
user_012	Cancel Invoice Receipt, Clear Invoice, Record Invoice Receipt	€ 918.687	€ 918.687
user_019	Cancel Invoice Receipt, Clear Invoice, Record Invoice Receipt	€ 745.482	€ 745.482
user_020	Cancel Invoice Receipt, Clear Invoice, Record Invoice Receipt	€ 545.946	€ 545.946
user_007	Cancel Invoice Receipt, Clear Invoice, Record Invoice Receipt	€ 529.589	€ 529.589
user_001	Cancel Invoice Receipt, Clear Invoice, Record Invoice Receipt	€ 372.741	€ 372.741
user_006	Cancel Invoice Receipt, Clear Invoice, Record Invoice Receipt	€ 72.947	€ 72.947
	Total	€ 7.086.425	€ 7.086.425

Figure 23: Users who perform multiple activities for the same case

2. Anomalies in the process flows

In this section deviations from the standard process are laid out.

Activities logged under wrong purchase order line item

For 4 purchase orders, activities are logged under one line item while these should probably have been logged under the other. In the dashboard below is seen that the purchase order has the same sum of purchase order as clear invoice, but on line item level this is not the case.

PO's	PO lines	Process iterations	Events	PO amount	Activities	Variants	Vendors	Items	Users
4	8	13	57	€ 32.138	12	8	4	6	20
Σ 76.349 | 0%	Σ 251.734 | 0%	Σ 422.606 | 0%	Σ 1.595.923 | 0%	Σ € 996.818.747 | 0%	Σ 42 | 29%	Σ 11.973 | 0%	Σ 1.975 | 0%	Σ 490 | 0%	Σ 628 | 3%

▲ Purchasing Document	Case ID	1. Create Purchase Requisition ...	2. Create Purchase Order Item	3. Vendor creates invoice	4. Record Invoice Receipt	5. Record Goods Receipt	7. Clear Invoice	Other activities	Total
4507029880	4507029880_00010	€ 0	€ 11.278	€ 22.556	€ 22.556	€ 11.278	€ 22.556	€ 11.278	€ 101.502
	4507029880_00020	€ 0	€ 11.278	€ 0	€ 0	€ 11.278	€ 0	€ 0	€ 22.556
4507034588	4507034588_00010	€ 0	€ 3.469	€ 6.938	€ 6.938	€ 3.469	€ 6.938	€ 10.407	€ 38.159
	4507034588_00030	€ 0	€ 3.469	€ 0	€ 0	€ 3.469	€ 0	€ 3.469	€ 10.407
4508055222	4508055222_00070	€ 62	€ 62	€ 124	€ 124	€ 62	€ 124	€ 124	€ 682
	4508055222_00080	€ 62	€ 62	€ 0	€ 0	€ 62	€ 0	€ 0	€ 186
4508059067	4508059067_00010	€ 1.260	€ 1.260	€ 0	€ 0	€ 1.260	€ 0	€ 6.300	€ 10.080
	4508059067_00040	€ 0	€ 1.260	€ 2.520	€ 2.520	€ 1.260	€ 2.520	€ 2.520	€ 12.600
	Total	€ 1.384	€ 32.138	€ 32.138	€ 32.138	€ 32.138	€ 32.138	€ 34.098	€ 196.172

Figure 24: A4I dashboard. Activities that might be registered under the wrong purchase order line item

No changes after debit memo

When a debit memo is created, we expect either "Change Price" or "Change Quantity". There are almost 6 thousand purchase order line items where the activity "Vendor creates debit memo" occurs, as shown below on the first footer.

Below is seen that in almost 5 thousand purchase order line items, the abovementioned changes were not recorded. This means that, according to the dataset, a debit memo did not lead to a change in price or quantity.

PO's	PO lines	PO amount	Events	Activities	Variants	Vendors	Items	Users
2.569	5.988	€ 103.228.342	79.782	35	2.379	655	179	373
Σ 76.349 | 3%	Σ 251.734 | 2%	Σ € 996.818.747 | 10%	Σ 1.595.923 | 5%	Σ 42 | 83%	Σ 11.973 | 20%	Σ 1.975 | 33%	Σ 490 | 33%	Σ 628 | 59%

Figure 25: A4I footer. Purchase order line items which contain the event "Vendor creates debit memo"

PO's	PO lines	PO amount	Events	Activities	Variants	Vendors	Items	Users
2.032	4.831	€ 45.984.057	55.322	34	1.769	517	90	333
Σ 76.349 | 3%	Σ 251.734 | 2%	Σ € 996.818.747 | 5%	Σ 1.595.923 | 3%	Σ 42 | 81%	Σ 11.973 | 15%	Σ 1.975 | 26%	Σ 490 | 26%	Σ 628 | 53%

Figure 26: A4I footer. Purchase order line items which contain the event "Vendor creates debit memo", as well as "Change Price" or "Change Quantity"

3. Anomalies in the event amounts

PO with 0 value

Within the data we also found that certain purchase orders have a PO-value of €0,-. Below table shows that for every item type there are purchase order line items valued at zero. See the first two columns.

16.385 PO line items have a 0-value, including 14.498 consignment orders. Other than the consignment orders, this is probably waste. Further investigation is required.

	0	<>0	Total Count o	0	<>0	Total Sum of Total
Row Labels	**Count of Case II**	**Count of Case ID**		**Sum of Total**	**Sum of Total**	
Consignment	14498		14498	€ 0,00		€ 0
Limit	1	1043	1044	€ 0,00	€ 13.213.158	€ 13.213.158
Service	29	5809	5838	€ 0,00	€ 163.740.191	€ 163.740.191
Standard	1435	218751	220186	€ 0,00	€ 792.621.567	€ 792.621.567
Subcontracting	411	4267	4678	€ 0,00	€ 12.500.465	€ 12.500.465
Third-party	11	5479	5490	€ 0,00	€ 14.743.366	€ 14.743.366
Grand Total	**16385**	**235349**	**251734**	**€ 0,00**	**€ 996.818.747**	**€ 996.818.747**

Figure 27: 0-valued purchase order line items per item type

Vendor	PO top 20	invoice top 20	Sum of 2. Create Purchase Order Item	Sum of 7. Clear Invoice
vendorID_0396	PO top 20	invoice top 20	€ 9.007.795	€ 142.879.524
vendorID_0234	PO top 20	invoice top 20	€ 16.305.609	€ 115.441.179
vendorID_0151	PO top 20	(blank)	€ 8.960.426	
vendorID_0176	PO top 20	(blank)	€ 9.450.894	
vendorID_0197	PO top 20	(blank)	€ 9.884.171	
vendorID_1023	PO top 20	(blank)	€ 9.908.207	
vendorID_0259	PO top 20	(blank)	€ 10.069.100	
vendorID_0193	PO top 20	(blank)	€ 10.378.017	
vendorID_1085	PO top 20	(blank)	€ 11.549.505	
vendorID_0939	PO top 20	(blank)	€ 13.696.972	
vendorID_0277	PO top 20	(blank)	€ 15.331.375	
vendorID_0963	PO top 20	(blank)	€ 19.319.387	
vendorID_0147	PO top 20	(blank)	€ 20.054.044	
vendorID_0183	PO top 20	(blank)	€ 20.411.055	
vendorID_0479	PO top 20	(blank)	€ 21.939.532	
vendorID_0166	PO top 20	(blank)	€ 22.243.010	
vendorID_0184	PO top 20	(blank)	€ 23.153.203	
vendorID_0106	PO top 20	(blank)	€ 24.691.895	
vendorID_0159	PO top 20	(blank)	€ 30.251.481	
vendorID_0104	PO top 20	(blank)	€ 32.104.521	
vendorID_0207	(blank)	invoice top 20		€ 49.251.777
vendorID_0204	(blank)	invoice top 20		€ 38.549.237
vendorID_0034	(blank)	invoice top 20		€ 25.980.760
vendorID_1442	(blank)	invoice top 20		€ 25.871.650
vendorID_0213	(blank)	invoice top 20		€ 37.083.404
vendorID_0395	(blank)	invoice top 20		€ 33.255.360
vendorID_0854	(blank)	invoice top 20		€ 35.932.388
vendorID_0885	(blank)	invoice top 20		€ 48.104.922
vendorID_0397	(blank)	invoice top 20		€ 28.578.836
vendorID_0509	(blank)	invoice top 20		€ 57.833.162
vendorID_0877	(blank)	invoice top 20		€ 45.467.452
vendorID_0201	(blank)	invoice top 20		€ 57.776.474
vendorID_0020	(blank)	invoice top 20		€ 56.170.401
vendorID_0472	(blank)	invoice top 20		€ 46.089.856
vendorID_0040	(blank)	invoice top 20		€ 53.838.985
vendorID_0330	(blank)	invoice top 20		€ 95.869.008
vendorID_0388	(blank)	invoice top 20		€ 223.304.784
vendorID_0053	(blank)	invoice top 20		€ 228.473.945
Grand Total			**€ 338.710.199**	**€ 1.445.753.104**

Figure 30: Top 20 purchase order line items and top 20 vendor invoices

Payment to vendor in relation with purchase orders

Row Labels	Sum of 2. Create Purchase Order Item
PO top 20	€ 338.710.199
(blank)	€ 658.108.548
Grand Total	**€ 996.818.747**

Figure 28: Top 20 PO value

Row Labels	Sum of 7. Clear Invoice
invoice top 20	€ 1.445.753.104
(blank)	€ 1.045.273.332
Grand Total	**€ 2.491.026.436**

Figure 29: Top 20 Clear Invoice (payment) value

In the vendor top 20 from a PO value perspective and from a clear invoice perspective, only 2 out of 20 belong to both categories. This is very strange and probably not caused by reality but due to the way the process log was created.

Service orders

After taking a closer look at the service order some observations can be made:

- The amounts processed in the events look unrealistic and could be due to the way the event log is preprocessed by the organizers of the challenge. Normally we would expect the values of service entry sheets , invoices and payment to be much more in the range of the original PO. If the values are realistic, this organization should clearly have a closed look at the way the order values for PO values are estimated and agreed with vendors.
- For > 50% of the service order with one or more service entry sheets the invoice has not yet been cleared.
- 1063 events belong to a variant where there is a service entry sheet but no good receipt event.

▲ Variant clusters	Cases	Events	2. Create Purc...	3. Vendor crea...	4. Record Invoi...	5. Record Goo...	6. Record Servi...	7. Clear Invoice	Other activities	Total
noPR-GR-SE-INV-noVI-noPayment	37	176	€ 39.468	€ 0	€ 40.045	€ 40.042	€ 39.919	€ 0	€ 17.739	€ 177.213
noPR-GR-SE-INV-VI-noPayment	263	7.563	€ 5.961.444	€ 8.489.337	€ 11.971.544	€ 53.296.673	€ 15.859.851	€ 0	€ 22.548.129	€ 118.126.978
noPR-GR-SE-INV-VI-Payment	2.596	111.974	€ 123.889.775	€ 2.323.830.382	€ 5.144.100.121	€ 5.918.503.694	€ 4.787.904.658	€ 1.831.158.540	€ 2.412.330.555	€ 22.541.717.725
noPR-GR-SE-noINV-noVI-noPayment	2.707	146.822	€ 10.799.987	€ 0	€ 0	€ 434.691.014	€ 78.623.311	€ 0	€ 82.720.135	€ 606.834.447
noPR-noGR-noSE-noINV-noVI-noPayment	234	1.063	€ 23.049.276	€ 0	€ 0	€ 0	€ 0	€ 0	€ 27.761.559	€ 50.810.835
noPR-noGR-SE-noINV-noVI-noPayment	1	3	€ 241	€ 0	€ 0	€ 0	€ 482	€ 0	€ 0	€ 723
Total	5.838	267.601	€ 163.740.191	€ 2.332.319.719	€ 5.156.111.710	€ 6.406.531.423	€ 4.882.428.221	€ 1.831.158.540	€ 2.545.378.117	€ 23.317.667.921

Figure 31: Total amounts processed within various variant clusters

4. Data quality

Old events

When looking at the table below, the amount of events are shown from before 2018. These were found in the dataset and seem to be mostly related to data quality issues.

Year	#events
1948	10
1993	9
2001	22
2008	45
2015	3
2016	6
2017	223
Total	**318**

Figure 33. Amount of events from before 2018

Item anonymization

After analyzing the data further we found that, through the anonymization performed, the same items numbers exist in the data for the various item types. The only way to make them unique is by joining the item type and item number, but this would be an assumption that we have not performed for our analysis. Recommendation is to create a unique item number when anonymization in the future.

PO's	PO lines	PO amount	Events	Activities	Variants	Vendors	Items	Users
97	264	€ 0	318	3	85	81	76	4
Σ 76.349 | 0%	Σ 251.734 | 0%	Σ € 996.818.747 | 0%	Σ 1.595.923 | 0%	Σ 42 | 7%	Σ 11.973 | 1%	Σ 1.975 | 4%	Σ 490 | 4%	Σ 628 | 1%

Figure 32: A4I footer. Purchase order line items with events from before 2018

Vendor ID's and names not consistent

126 vendor ID's have the same vendor name. The top three vendor names (in terms of cases) are illustrated on the right. It shows that they have multiple vendor ID's. This could indicate that a new vendor is created for the same organization, but for another entity (with the same trade name). It could also mean that due to reluctancy or human error the vendor is created as a double.

Another option is that other subsidiaries or delivery addressed are applied, but still this should also be documented and labelled in order to prevent the wrong comparisons and conclusions.

Although the ID is the differentiator, unique naming for vendors is recommended. This is mainly to improve the vendor-based business intelligence and to reduce the errors of selecting the wrong vendor.

We recommend using a standard naming convention for every vendor, for example: [COUNTRY] | [ENTITY_NAME] ([TRADE_NAME]).

Vendor Name	Vendor	Closed	Incorrect start	Open	▼ Total
vendor_0104	vendorID_0104	8.074	0	1.743	9.817
	vendorID_0803	4	0	1	5
vendor_0164	vendorID_0166	868	2	235	1.105
	vendorID_0184	565	4	99	668
	vendorID_0619	2	0	76	78
	vendorID_1078	7	0	13	20
	vendorID_1559	1	0	0	1
vendor_0143	vendorID_0144	266	0	36	302
	vendorID_0479	126	1	108	235
	vendorID_0357	38	0	194	232
	vendorID_0146	16	0	150	166
	vendorID_0458	5	0	108	113
	vendorID_0892	90	3	18	111
	vendorID_0223	49	0	31	80
	vendorID_0286	51	0	9	60
	vendorID_0616	42	1	17	60
	vendorID_0926	21	1	8	30
	vendorID_0359	10	0	2	12

Figure 35: Top 3 vendor names with the various vendor ID's

PO's	PO lines	Process iterations	Events	PO amount	Activities	Variants	Vendors	Items	Users
10.695	22.002	22.043	126.490	€ 192.501.504	31	1.553	126	124	318
Σ 76.349 | 14%	Σ 251.734 | 9%	Σ 422.606 | 5%	Σ 1.595.923 | 8%	Σ € 996.818.747 | 19%	Σ 42 | 74%	Σ 11.973 | 13%	Σ 1.975 | 6%	Σ 490 | 6%	Σ 628 | 51%

Figure 34: All cases which have vendor ID's with the same vendor name

6. Improvement

1. Original Data Set

The data made available for this challenge shows some very strange patterns, for instance: log records without user ID's, mostly identical amounts within case ID's, 50% of iteration-ID's with a throughput time of zero (with no specific vendor, users, period/time), anonymization of item ID is such a way that the same item number is in use for different item types, etc.

For future challenges we would appreciate to receive the raw data as we believe data preparation for process mining determines 50% of the final value for the customer. Now we continually have to ask ourselves whether our observations are due to the way the data preparation took place or that this company has a real root cause in the processes that resulted in this data.

2. Next Level Process mining

We believe the next level of process mining will be enabling Business Process Improvement through a combination of (integrated) business intelligence and process intelligence and the flexibility to work within the same dataset with more than one case ID.

3. Process Model improvement: First Time Right and Elimination of Waste

We believe the company can significantly reduce process costs by deep diving into the process models identified & further scrutinizing the root cause of both additional activities without iterations and the reason for the many iterations. First Time Right should be the goal!

For consignment orders we advise the company to consider changing the process in such a way, that for this order type also values are recorded (as the stock is within the premises of the company) and -where possible- establish a connection between consignment orders and the corresponding real orders for those items.

We also see a lot of case iteration ID's where the throughput time is zero (50% of iterations) and would not be surprised if these could be eliminated as they do not provide value. However, when this is due to the data preparation, this obviously should be fixed.

We believe approval activities should never be performed by a batch user. In case a vendor creates the invoice in a cloud application, we believe the application should log the user ID of a natural person at that company.

To perform even better checks on the authorization, we suggest adding the department code as an attribute for each user and to define a rule set with conflicting activities.

By making our app available to the users in the departments of the company and periodically adding data they can monitor and improve the processes themselves to both get clean and stay in control.

4. Throughput improvement

Further analysis should provide insight if the root cause for the high average throughput time is only due to the payment terms or that there are other reasons for the high throughput times.

5. Data Quality

Application default settings and continuous data quality monitoring are the ways to bring the quality of the data within the process on a much higher level. In any case, the consistency of vendor ID's and names should be assured, the strange amounts in the events should be fixed (e.g. if they are due to copying it from the wrong (currency) field or related to another unit of measurement?) and the user names should be logged by default.

Thanks for reading!

Jordy Bekker

IT Auditor at Mazars

+31 611 920 648
Jordy.bekker@gmail.com

Albert W. Kisjes

Partner at Agilos

+31 665 853 729
akisjes@agilos.nl

Balancing Efficiency and Risk in Procure to Pay: Safely Realizing Cost Savings Using Process Mining Techniques

Adrien Porter, David Masse, Nuss Visatemongkolchai, Jithendra Seneviratne, Tanvee Deokule, Nicholas Hartman

CKM Analytix, 200 West 41st Street, New York, NY 10036, USA
{aporter, dmasse, nvisatemongkolchai, jseneviratne, tdeokule, nhartman}@ckmanalytix.com

Abstract. For the 2019 BPI Challenge, we use process mining techniques to explore and analyze procure to pay event logs from a large multinational paints and coatings company. Suboptimal procurement processes can lead to increased costs and operational risks for businesses. Within the company's data, we identify a range of opportunities for process optimization, including:

1. Invoices are frequently paid at only fixed pre-defined intervals. These intervals, which vary by vendor, often introduce long delays in payments while the process awaits the next available payment window. We model that clearing invoices more dynamically could better optimize cash flow management and allow for negotiation of discounts based on faster payments.

2. We apply algorithms to scan for noncompliance against published process requirements and recommend user re-training and/or system adjustment for users or vendors associated with concentrations of such behaviors.

3. We analyze the existing use of automation throughout the process, which identifies several activities handled via a mix of human resources and automated systems. We model the impact of a potential increase in the saturation of automation and make recommendations on where such automation is best targeted.

4. We rank vendors using a custom two-dimensional complexity metric that identifies which vendors most commonly cause common process inefficiencies. By flagging the least consistent and most time-consuming vendor-specific processes, future process efficiency efforts can be better targeted.

5. A social network analysis reveals that certain users perform more than one essential cross-checking step necessary for payment processing, which may raise the potential for fraud or errors by limiting key checks and balances. We also identify vendors and categories of goods/services that are particularly prone to such behaviors.

6. We identify vendors and categories of goods/services with frequent change events (e.g. "Change Quantity" or "Change Approval for Purchase Or-

der”). Modeling of the impact of these activities identifies that such rework introduces significant delays within the overall process and increases the amount of human labor required to complete the process.

By leveraging ongoing advanced process mining to monitor the impact of actions against the above opportunities, the subject company will be able to realize material cost savings and efficiency gains while also better monitoring and managing risks associated with the procure to pay process.

Keywords: Process Mining, Process Discovery, BPIC 2019, Process Improvement, Event Logs, Conformance Checking, Social Network Analysis, Automation, Procurement, SAP, Materials Management, Machine Learning

1 Introduction

Procurement is a critical business activity for any enterprise. Businesses commonly utilize purchase orders to acquire the goods and services necessary for their operations. Modern businesses make use of enterprise resource planning (ERP) systems to digitally track the progress of each order. These systems produce valuable data that can be mined to answer a multitude of questions about the functioning of the procurement process. In this report, we explore and analyze the BPI Challenge 2019 dataset to optimize the process, identify and manage risks and provide additional leverage for future pricing negotiations. We perform analyses on compliance, automation, throughput, payment times, process complexity, and social networks.

2 Overview of the Data

2.1 The Data

The BPI Challenge 2019 dataset comprises just over one year’s worth of purchase order data from a large multinational paints and coatings company. The raw dataset contains 1,595,923 events distributed across 251,734 cases [1]. Cases are defined as a combination of a purchase order and purchase item. Each time-stamped event contains one of 42 activities such as “Create Purchase Order Item.” In addition, each event has an associated set of informational attributes which includes the vendor, the value, the categorizations of the purchased item, the ERP system user, and the type of purchase order. Based on the terminology present in the data, we infer that the system in use is a commercial product made by SAP.

The dataset contains some cases with event timestamps that fall outside of the date range specified in the challenge instructions. We apply a date filter to only allow cases that start and end between the beginning of 2018 (2018-01-01 00:00:00) and the publication date of the dataset for the competition (2019-01-27 23:59:59). This reduces the dataset to 1,587,802 events across 251,463 cases. 99% of the data is retained after this filter is applied. We present detailed descriptions, statistics, and observations about the time-filtered data in the appendix.

The data was processed and analyzed using a combination of Fluxicon Discovery, ProM, and a process-mining tool internally developed at CKM Analytix, as well as custom Python analytics code.

Several sources of the event log were published for the competition. We noticed a time-zone discrepancy (events shifted by five or six hours) between the CSV data from the BPI Challenge web page and the .dsc file provided by Fluxicon. All analyses discussed in this report are based on the raw CSV data.

Table 1 shows summary statistics around case completion. Because the dataset is a slice in time, it is important to distinguish between cases that can be considered complete (i.e. contain a "Clear Invoice" event or "Record Goods Receipt" in Consignment cases) and cases that may have been in process when the sample was taken. Every case includes exactly one "Create Purchase Order Item" event, so the beginning of each case is present (i.e. no case is included in the log that started before the timeframe of the log and continues in the log). Time-related aggregates (e.g. median case duration) do not apply to incomplete cases as an incomplete case may have just started but could last any number of days into the future.

Table 1. Case Completion Statistics

	Count	% by Count	Value (EUR millions)[1]	% by Value	Mean Value (EUR)	Median Value
Completed cases	196,881	73.2%	711.6	78.3%	3,615	491
Non-completed cases	54,582	26.8%	260.1	21.7%	4,766	565
Total	**251,463**	**100%**	**971.8**	**100%**	**3,864**	**508**

2.2 Description of the Four Archetypal Processes

An important aspect of the data in this challenge is the existence of sub-processes explicitly specified in the challenge statement and codified in the data based on certain attributes (see "case Item Category" in the appendix). Below we present an overview of each process and summarize associated key statistics in **Table 2**.

[1] Sum of values recorded for "Create Purchase Order Item," which occurs exactly once per case.

Table 2. Summary Statistics for the four archetypal processes

	3-Way After	3-Way Before	2-Way	Consignment
Events	312,554[2]	1,233,410	5,758	36,080
Cases	15,129	220,810	1,027	14,497
Complete Cases[3]	9,624	173,503	289	13,465
Unique Activities	38	39	11	15
Unique Process Variants[4]	5,297	8,591	144	301
Median Complete Case Duration (days)	80	77	7	20
Mean Complete Case Duration (days)	89	81	20	24
Median Case Value (EUR)	402	594	6,174	0

The process maps below, for clarity, show only the most common and essential steps for compliance. The numbers between the straight arrows at the left show median position ("rank") within the cases for each activity. The numbers next to the curved arrows show the number of transitions from one activity to another in the event log (thickness of the arrows is scaled by this figure).

3-way Match, Invoice After Goods. We refer to this process as "3-way-after" throughout the rest of the report. Invoice receipts should be entered only after goods are received (activity "Record Goods Receipt") and are matched against the goods receipt and PO creation. We include the Service subprocess as part of "3-way-after" in the summary statistics but isolate and analyze it in compliance section of the report.

2 This figure includes the Service sub-process, which comprises 261,016 events and 5,800 cases within the main 3-way after process.

3 Complete Cases: Cases are considered complete if they include both "Create Purchase Order Item" and "Clear Invoice" activities ("Record Goods Receipt" for Consignment). Complete case durations are based on the time elapsed between the first instance of a "Create Purchase Order Item" activity and the last instance of a "Clear Invoice" activity.

4 Calculation of variants: we group together otherwise similar cases with different counts of repeated events at the same moment. For instance, a case with event sequence A → B (recorded 3 times at the exact same timestamp) → C would be considered the same variant as a case with sequence A → B (recorded 5 times) → C.

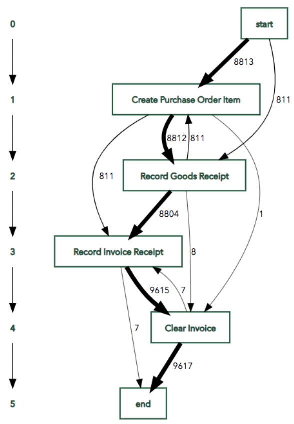


Fig. 1. First cycle (first instance of activities up to the first "Clear Invoice") of complete cases

3-way Match, Invoice Before Goods. We refer to this process as "3-way-before" throughout the rest of the report. In this process, an invoice receipt may be entered prior to the entry of a goods receipt, but any payment is blocked until the goods receipt is entered and matched against the invoice received and PO created. We define "cycle" as the series of activities leading up to an individual instance of "Clear Invoice," i.e. a payment to a vendor (see compliance section 3.2 for a description of how corresponding events are identified). For 91% of the cases in this group, the goods receipt occurs before or at the same time as the invoice receipt (as in "3-way match, invoice after goods"), at least in the first cycle. For compliance, these cases do not require "Remove Payment Block," but 21% of the time, this extra step is performed, increasing the median time between "Create PO Item" to "Clear Invoice" from 72 to 91 days.

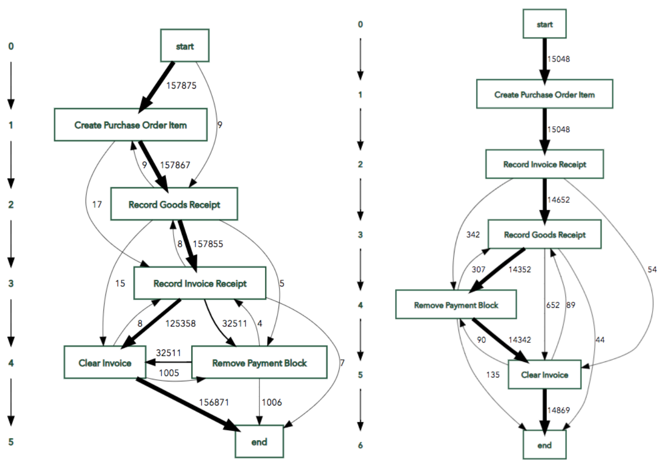


Fig. 2. First cycle of complete cases where "Record Goods Receipt" happens before or simultaneously with "Record Invoice Receipt" (left) and where "Record Goods Receipt" happens after "Record Invoice Receipt (right)

Approximately 95% of cases where "Record Invoice Receipt" happens before "Record Goods Receipt" do have "Remove Payment Block" as required for compliance and take a median of 73 days to complete.

2-way Match. We refer to this process as "2-way" throughout the rest of the report. Invoices received are simply matched to the initial purchase-order value.

Consignment. We refer to this process as "Consignment" throughout the rest of the report. Goods receipts are matched against initial purchase-order value only as there are no invoices associated with this sub-process.

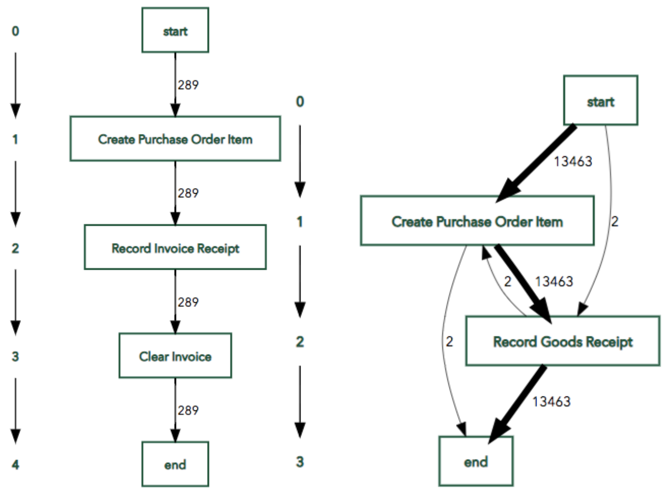


Fig. 3. First cycle of complete cases for 2-way (left) and Consignment (right)

3 Compliance Checking

3.1 Key Findings and Recommendations

Compliance Analysis. We deem a process to be compliant if it follows a set chronological sequence of events in conjunction with matching EUR amounts. These sequences are determined for each process type, as defined in the challenge statement. Cases belonging to the "Service" Item Type (subprocess of 3-way-after) exhibit low compliance rates (43.7% by number) compared to other processes (>80%). We recommend adjustments to the manner in which the system is utilized, either by reprogramming batch resources and/or retraining human users in order to maintain better adherence to standard PO process guidelines, thereby reducing financial and operational risk.

Least Compliant Vendors. Our analysis reveals a set of high-value vendors associated with low compliance rates. These processes should be scrutinized, as they could present a substantial financial risk to the company. The highest-value vendors that comply less than 90% of the time are ID_0183, ID_0479, ID_0234, ID_1023 and ID_0404.

3.2 Compliance Analysis

In order to determine which cases have been handled properly, we must define a compliance pattern for each sub-process in the event log. This pattern is a sequence of activities that must occur in time order and with matching EUR amounts for a case to be called compliant. "Create Purchase Order Item" occurs once and only once per case throughout the log, but other activities can occur zero, one or more times per case. For a case to be compliant, we require a correct sequence of matching events to precede every instance of "Clear Invoice" in a case, i.e. every cycle must be compliant. The one "Create Purchase Order Item" event is used for all cycles, but to match up other events, we number each instance of a particular activity with its time order, i.e. "Record Goods Receipt 1" then "Record Goods Receipt 2" and so on. Thus "Record Goods Receipt 2" is checked against "Record Invoice Receipt 2" and possibly "Remove Payment Block 2" for proper sequencing and EUR value. We cannot judge whether an incomplete case will ultimately reach a compliant outcome, so we only apply our compliance test to completed cases.

Compliance by Process. Below we present the patterns used for each of the four main sub-processes. Events with a EUR value that does not match the EUR value of the item at PO creation are addressed separately. Simultaneous events are considered "in compliant order" regardless of their positions in the event log. A summary is shown in **Table 3**.

Three-way matching, invoice after goods. Each instance of "Clear Invoice" must be preceded by the following sequence in order: "Create Purchase Order Item," "Record Goods Receipt," "Record Invoice Receipt." 3-way-after cases have been further split into cases with Item Type Service and cases with non-Service Item Types (see discussion below).

Three-way matching, invoice before goods. Cases that are compliant using the pattern for "3-way-before" are deemed compliant. In addition, "Record Invoice Receipt" is allowed to occur before "Record Goods Receipt"; but when this happens, compliant cases need to have a "Remove Payment Block" event after "Record Goods Receipt" and before "Clear Invoice." This is to ensure the invoice is not paid before the goods ordered are confirmed as received.

Two-way matching. Each instance of "Clear Invoice" must be preceded by the following sequence in order: "Create Purchase Order Item," "Record Invoice Receipt."

Consignment. "Create Purchase Order Item" followed by at least one instance of "Record Goods Receipt."

Table 3. Compliance for the four archetypal processes and Service sub-process

	3-Way After	Non-Service	Service	3-Way Before	2-Way	Con-sign.
Compliance by number	80.5%	93.8%	43.7%	96.8%	100%	100%
Compliance by EUR value	80.2%	83.1%	79.4%	95.4%	100%	n.a.

Analysis. To address the markedly low Service compliance numbers, our analysis uncovers two primary driving forces behind noncompliance for such processes.

Mismatched EUR values. We note that 1.1% of completed cases have at least one event with a EUR value that does not match the EUR value of the PO item at creation. Among cases with Item Type Service, this figure jumps to 53%. The non-matching EUR values for "Record Goods Receipt" and "Record Invoice Receipt" are often near-multiples of the expected value. We assume that the fact that the multiples are not exact arises from rounding errors during the anonymization of the data as discussed on the challenge description page [2], or from tax, shipping or other add-on charges.

Since we do not have the quantity for each PO item, we could not determine with certainty which cases were handled properly. These non-matching values may indicate 1) confusion around when to use unit prices and when to use aggregate values in the system and/or 2) a problem with the way service entry sheets, upon approval, are translated into goods receipts (typically done by batch_06).

We noticed that the median time between PO item creation, service entry sheets and goods receipts for Service cases is zero, so this may be a case of automating a process that should not occur. The system should check that the number of these events makes sense and prohibit extraneous events or those with incorrect EUR values.

Lack of standardized usage of the system across Service processes. We grouped together similar processes based on event sequences using K-modes clustering. This method is chosen due to its performance on categorical data and scalability for large datasets. Four clusters emerged as a model that best describes the data as determined by the silhouette method, which finds the optimal separation of data based on their similarities and differences.

Of these four groups, one appears to be the standard process, with core activities with no repetition (881 cases, of which 67.2% compliant). The remaining three groups, however, see significantly lower compliance rates with an average of 27.3% across a combined 1,677 cases. Processes within these groups see distinct deviation from the standard process, such as with 'Record Goods Receipt' and 'Record Service Entry Sheet' appearing with N repetitions simultaneously. EUR value recorded at invoice receipt may also be an N multiple of the value at purchase order creation.

These different groupings highlight potentially inconsistent usage of the system by Service vendors. We therefore recommend standardizing procurement system practices where possible to simplify compliance detection and reduce risk.

Data Exploration using a Random Forest Classifier. In parallel with compliance analyses discussed above, a random forest classifier was applied to the dataset to help uncover patterns behind compliance. We utilized the classifier as a data exploration tool find meaningful data segmentations (e.g. vendors, spend areas) given compliance outcomes by examining characteristics of a case or event that most heavily influence the decision tree. While direct results from this analysis were omitted to respect the limited length of this report, we note that the methodology has been an integral tool in extracting insights from this dataset. An elaboration of how the model works can be found in the appendix.

3.3 Least-Compliant Vendors

Having established a framework for compliance, we are able to identify which vendors are associated with a high degree of non-compliant processes. We look at 3-way-before and 3-way-after cases together since vendors are present across different subprocesses, while setting aside 2-way and Consignment cases since they are 100% compliant. The bubble plot in **Fig. 4** further highlights the low compliance of Service processes, as discussed in the prior section. Additionally, we notice a higher degree of compliance among higher volume vendors. Nonetheless, we also see a certain number of vendors that stand out based on their lower levels of compliance. The figure also shows that the number of cases does not necessarily correlate with EUR value of the vendor's cases (sum of PO item creation event values, shown as circle size below). We therefore extracted the vendors who are responsible for the highest EUR value in items but were compliant for under 90% of their cases. We chose these vendors due to the fact that non-compliance for high value purchase orders presents a substantial risk to the company.

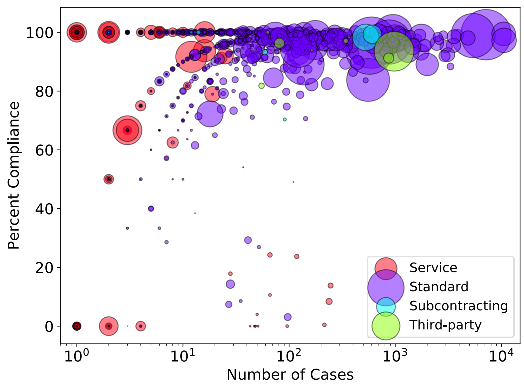


Fig. 4. Vendor compliance with number of cases. Bubble size indicates cumulative value of items from a vendor. Note the logarithmic scale on the horizontal axis.

We present a set of vendors that warrant further investigation in **Table 4**.

Table 4. Highest-value vendors that comply less than 90% of the time

Vendor	Number of Cases	Compliance %	EUR (millions)	Doc. Type	Item Type	Spend Class	Spend Area
ID_0183	555	83.8	17.0	Standard PO	Standard	PR	Latex & Monomers
ID_0479	137	89.1	15.6	Standard PO	Standard	PR	Titanium Dioxides
ID_0234	3	66.7	7.8	Framework order	Service	NPR	Logistics
ID_1023	18	72.2	6.3	Standard PO	Standard	PR	Titanium Dioxides
ID_0404	2001	89.1	4.7	Standard PO	Standard	NPR	Sales

Case Study: vendorID_0183. A deeper look at this particular vendor reveals that non-compliance tends to occur as a result of non-compliant activity occurring after a compliant first cycle (from "Create Purchase Order Item" to "Clear Invoice"). We identify a number of cases where a "Cancel Invoice Receipt" event occurs over 120 days after the initial "Clear Invoice" event, which would conclude a compliant cycle. Following "Cancel Invoice Receipt," we see a new "Record Invoice Receipt", and a new "Clear Invoice", all happening within a few minutes of each other. We suspect that this sequence is a result of some kind of necessary price/quantity change that could be benign. However, we highlight the fact that a "Clear Invoice" event has already occurred (and presumably funds have been transferred to the vendor), and that this post-completed transaction work and payment do not appear to follow a robust process. The system records no information about what could be causing this, making audit difficult. Either the user has not been educated on how to perform these types of actions, or the system is not set up to allow a process which should be able to take place.

4 Automation

4.1 Key Findings and Recommendations

Automation Opportunities. 49% of events in the log show potential for automation, representing an opportunity for the client to reduce payroll costs. Analysis reveals that automated activities are presently concentrated in certain spend areas and vendors. We recommend that the company begin investigating these processes to further automate its workflow.

4.2 Automation Analysis

We explored activities by resource type: human users or batch systems, which are described by the challenge as automated processes. Several activities that could be automatically executed by a batch resource were also completed by human users (**Fig. 5**).

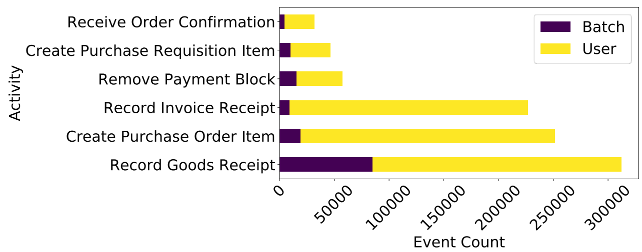


Fig. 5. Top six automatable events by count.

Human-user-executed instances of these the top six events by count represent 49% of events in the dataset[5]. Without additional information from the client at this stage, we assume that this represents a significant opportunity to further automate the workflow in order to reduce costs. Why were such events logged as done manually? Could they have been instead carried out by batch processes?

A key assumption in this analysis is that 'user' and 'batch' resources are synonymous with manual and automated. Degrees of automation differ widely across 1,961 vendors. We noticed instances of events that were carried out by users but occurred on a very consistent schedule (e.g. every 12 am on Sunday). These occurrences are relatively rare.

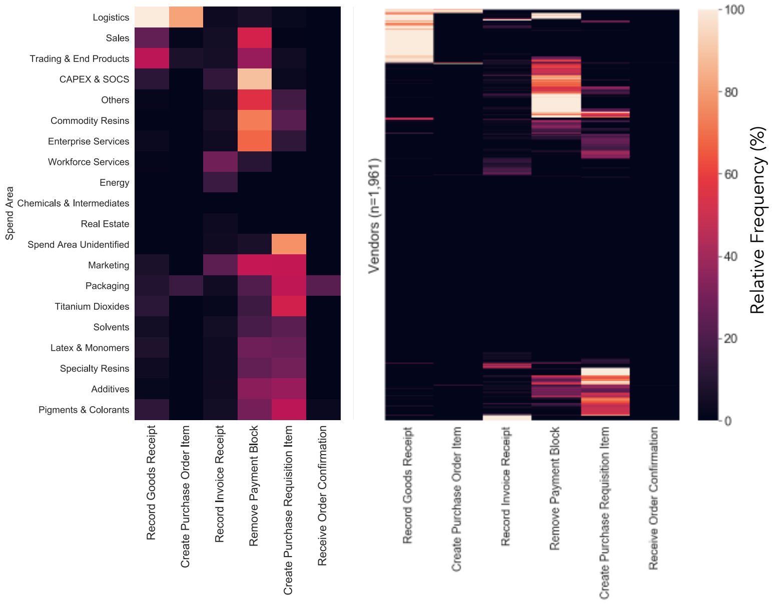


Fig. 6. Percentage of automation by spend area (left) and vendor (right)

5 An automatable event is defined as events within one of the 6 high-count activities shown in Fig. 5 executed by a 'user' resource (yellow). We assume any efforts to transform processes towards automation is best focused on frequently-occurring events.

Fig. 6 shows the distribution of such batch events by spend area and vendor. Brighter-colored tabs correspond to a higher percentage of events executed via batch processes by category. The following insights may be derived:

1. Automation is concentrated within certain spend areas and vendors. For instance, batch executions of ‘Record Goods Receipt’ are associated with vendors dealing in Logistics, Trading & End Products, and Sales spend areas. Processes in categories such as Real Estate and Chemical & Intermediates are relatively manual.

2. High automation rate for one activity is not necessarily correlated to a high automation rate for other activities across the same spend area or vendor. Cases in which goods receipts are consistently batch-recorded do not correspondingly see the same level of automation in purchase order creation, invoice receipt, or other events. Multiple separate efforts may need to be integrated in order to achieve a fully automated process.

4.3 Case Studies

We present the following vendor case studies to further illustrate how automation is utilized by the company.

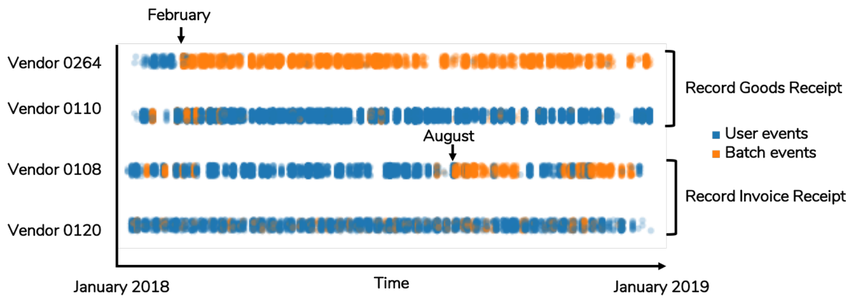


Fig. 7. ‘Record Goods Receipt’ and ‘Record Invoice Receipt’ events over time

Case Study 1: Vendor 0110 & 0264 (Record Goods Receipt). Of 3,802 instances of goods receipt records for items associated with vendor 0110, 13.7% were executed by batch processes, distributed randomly over the course of the event log (**Fig. 7**). The lack of discernible differences between cases with user- and batch-executed events suggests that automation of goods receipt record should be explored with this vendor in order to potentially reduce resource time required/payroll costs.

On the other hand, there appears to be a temporal pattern in user- and batch-executed for items associated with Vendor 0264, with a shift to batch goods receipt in February 2018. It appears that the process associated with this vendor underwent a change through the introduction of automation. We believe that further investigation into this vendor could reveal insights into how other vendor processes could have their goods receipts further automated as well.

Case Study 2: Vendor 0108 (Record Invoice Receipt). Similarly, invoices from Vendor 0108, which are exclusively of the Sales spend area, began being recorded largely by batch systems beginning in August 2018. Insights beyond the scope of this dataset would allow for further understanding of automation (e.g. paper-based invoices may need to be entered manually into the procurement system in the absence of a functional image-recognition algorithm).

Case Study 3: Vendor 0299 (Create Purchase Order Item). Some purchase order items from this vendor were created manually, while others were generated by batch processes. Approximately 4% of items were deleted by users soon after they were batch-created. This is may represent a notable inefficiency in the process, and therefore should be investigated.

Additionally, the presence of a batch user in a case typically reduces the number of human users involved by 0.8 with high statistical significance (p-value = 0). We surmise that there may be multiple obstacles preventing the automatability of an event. We recommend that the client first investigate processes associated with vendors that are already partially automated, and transition towards automation where viable in order to potentially reduce payroll costs.

5 Throughput Analysis: Backlogs, Payment Timing and Process Complexity

5.1 Key Findings and Recommendations

Backlog Analysis. "Record Invoice Receipt" is the activity most in need of increased automation, or more resources, in order to speed up the completion of a three-way or two-way match.

Payment Terms. We recommend that the company transform its payment process so that invoices can be paid at least weekly, thereby enabling it to pay invoices in a more optimized manner. This could lead to the following benefits:

1. Reduced costs by taking advantage of early-payment discounts. This represents EUR9.16 million in potential savings to the company.
2. Better overall cash management by optimizing payments up to limits provided in contract terms.
3. Reduced costs from penalties/interest charges for late payments.

Vendor Process Complexity. We tagged vendors using a custom two-dimensional complexity metric that identifies which vendors are most associated with common process inefficiencies. We recommend that processes associated with these vendors be examined further to understand how they might be standardized and streamlined.

5.2 Backlog Analysis

Backlog is a useful metric to determine where a process may be getting stuck. For an activity (e.g. "Record Invoice Receipt"), backlog is defined as the number of cases that are waiting for "Record Invoice Receipt" to happen at any given time. On the process map, these cases are the ones in the process of traversing arrows from other activities to "Record Invoice Receipt." Since the backlog for an activity varies over time, we examine the median and maximum values over the course of the event log time period.

Table 5. Backlog for first cycle of complete cases[6]

Backlog (cases)	3-Way After	3-Way Before, GR before IR	3-Way Before, IR before GR	2-Way	Consign.
GR Median	467	7,802	151	n.a.	876
GR Max.	762	9,802	396	n.a.	1,250
IR Median	949	9,249	739	8	n.a.
IR Max.	1,626	14,785	1,655	16	n.a.
CI Median	1,018	22,136	2,379	9	n.a.
CI Max.	1,640	30,533	4,609	56	n.a.
RPB Med.	n.a.	2,650	184	n.a.	n.a.
RPB Max.	n.a.	5,295	406	n.a.	n.a.

The generally high backlog at "Clear Invoice" – and how to address it – is discussed in depth in the payment-terms section of this report. In this section we highlight the relatively elevated backlogs at "Record Invoice Receipt," especially for the 3-way-after subprocess. Indeed, across the subprocesses that include this activity, median wait times between "Record Invoice Receipt" and the prior step range from 14 to 22 days. Both batch and human users perform this activity throughout the week, so one might wonder about the presence of high wait times and backlogs. One reason is that some cases are waiting for a "Vendor creates invoice" event to move on from "Create Purchase Order Item" or "Record Goods Receipt" to "Record Invoice Receipt." However, in the vast majority of cases "Vendor creates invoice" is already in place, at which point the process should be ready to proceed to "Record Invoice Receipt." We conclude that "Record Invoice Receipt" should be the focus for further automation. More specific opportunities for backlog/time reduction are as follows:

[6] Goods Receipt (GR), Invoice Receipt (IR), Clear Invoice (CI), Remove Payment Block (RPB)

Three-way matching, invoice after goods. Repetitions of "Record Invoice Receipt" and "Record Goods Receipt," which are sometimes extremely quick but can take a couple days, seem to be related to the Service segment (see discussion in compliance section).

Three-way matching, invoice before goods. Potential problem areas for investigation include the transition from "Record Goods Receipt" to "Remove Payment Block." Many unnecessary "Remove Payment Block" actions occur (see description of four subprocesses), but even those that are necessary have a median transition time of 9 hours, and the average transition time is 4 days, meaning that some of these cases get stuck on other activities for several days. This seems like a good candidate for further automation as there should be no obstacle to "Remove Payment Block" once goods have been received. A "Change Quantity" event can add a week to the median 8- to 10-day transition from "Create Purchase Order Item" to "Record Goods Receipt," so we recommend attempts to reduce the frequency of change events (see vendor complexity section).

Two-way matching. The largest median backlog by far (110 cases) sits at the "hidden" activity "Change Approval for Purchase Order." The main onward transition to "Record Invoice Receipt" takes only 3 days on average, so the problem seems to be transitions from "Change Approval for Purchase Order" back to itself, which happens more than once per case and takes 17 days on average, warranting an investigation into how many of these are necessary.

5.3 Payment Terms Analysis

Payment terms are a critical aspect of every purchase order. A purchase order constitutes a contract between the company and the vendor, which specifies the obligations that must be met by the respective parties. In general, once the vendor has received the purchase order from the company, it will deliver the goods or services and send out a corresponding invoice. The company must then send a payment within a period of time specified in the purchase order. Usually the payment deadline is 30, 60 or 90 days. If the payment is not accomplished in time, late fees may be charged to the company.

Cash management is an important consideration for any business, and consequently each side of the payment leg of the transaction has its own interest pertaining to the timeliness of its execution. From the company perspective, payment should be made as close to the deadline as possible to maximize its own cash position. Vendors would like to receive payment as soon as possible. In certain cases, suppliers will discount early payments as an incentive to vendors to pay quickly.

In this section we analyze the elapsed time between events constituting possible triggers of a "countdown clock" to the payment deadline, and the ultimate time at which the payment was made. These analyses consider only "completed" cases. We focus our analysis on the "3-way-before" process because it is the predominant process for transactions in this dataset.

Invoice Clearing Occurs at Set Intervals, While Vendor Invoicing Is Continuous. We consider “Vendor creates invoice” and “Clear Invoice” to be the typical start and end points of the payment timeline. **Fig. 8** shows the distribution of elapsed time between these two events by vendor. The heatmap shows that durations are clustered around certain bands for different vendors.

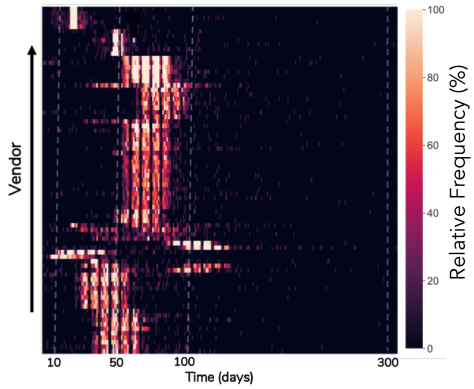


Fig. 8. Distribution of Elapsed time between “Vendor creates invoice” and “Clear Invoice” for 3-way-before, with vendors hierarchically clustered (vendors with at least 100 cases).

Looking at hierarchical cluster maps of “Vendor creates invoice” across time vs. “Clear Invoice” across time, we notice a pattern that explains the spread of durations for each vendor and shows opportunity to improve the process, as shown in **Fig. 9**.

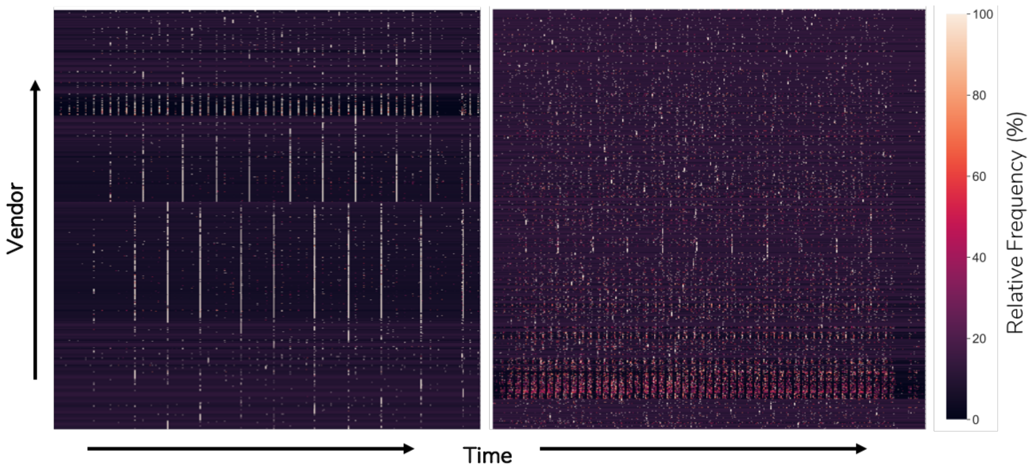


Fig. 9. “Clear Invoice” (left) and “Vendor creates invoice” (right) across time and vendor for 3-way-before, with vendors hierarchically clustered.

For most vendors, “Vendor creates invoice” events occur relatively constantly across weekdays, yet “Clear invoice” events are highly concentrated usually on the first and last Thursday of each month, as displayed by the light streaks. We observe that some vendors clear invoices on a weekly schedule, but overall, it appears that the “Clear Invoice” process is set by an inflexible and pre-determined schedule. This is suboptimal because it means that:

1. In the best case, vendor payments will be made too early, depriving the company of cash they may have been able to use or invest until the due date.
2. In the worst case, late payments if the “Clear Invoice” day is missed.

Below we inspect the distributions at the vendor level as well, which can give some information about typical payment terms for individual vendors.

Case Study: vendorID_0147. Vendor 0147 is chosen as an example case study for the type of analysis that can be performed. This particular vendor exhibits interesting characteristics:

1. Frequency: This vendor has the 37th most cases in the completed “3-way-before” process (952 completed cases).
2. Value: This vendor has a relatively high median PO Item value of EUR12,898 (82nd percentile).

Given this combination of frequency and high value, there is some opportunity for evaluating any potential costs associated with the company’s paying this supplier at any time other than the contracted due date. The data provides limited information about any due dates that would have been specified in the purchase order document, but by looking at the distribution of elapsed time between “Vendor creates invoice” and “Clear invoice,” we can gain some insights into the process, as shown in **Fig. 10**.

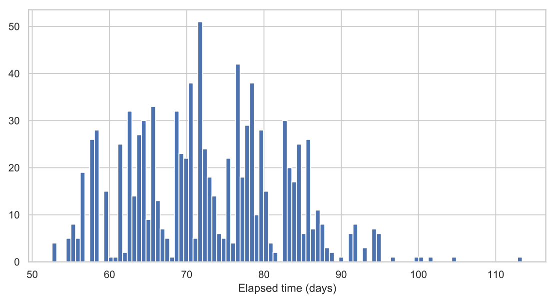


Fig. 10. Elapsed time between “Vendor creates invoice” and “Clear invoice” for completed cases associated with Vendor_0147

The figure shows peaks and valleys that fluctuate over time but do not appear to synchronize with expected payment terms of 30/60/90. Looking more closely at the occurrences of “Vendor creates invoice” and “Clear Invoice” for this vendor (**Fig. 11**), we observe that this action is performed once per month (with a few exceptions). Additionally, “Vendor creates invoice” occurs fairly evenly across weekdays. Consequently, we can conclude that the long amount of time that takes place between payment events can have a significant impact on the completion of a payment to a vendor. This is undesirable because it gives the company less flexibility in paying vendors on time, which could lead to penalties and interest charges. It also provides less ability for the company to use faster payment as leverage in negotiations over pricing or early-payment discounts.

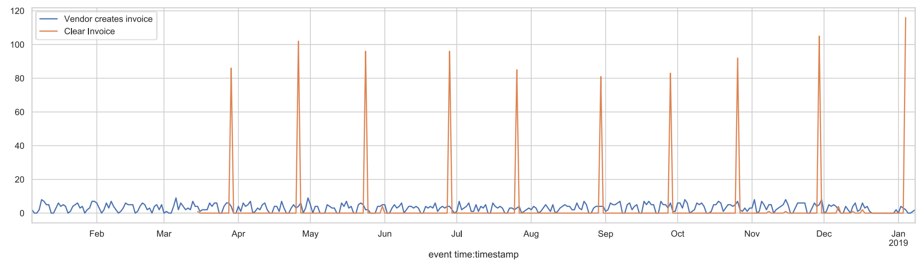


Fig. 11. “Vendor creates invoice” vs “Clear Invoice” across time for 3-way-before for Vendor 0147

Monetary Impact of Dynamic Invoice Clearance. Given insights from **Fig. 9**, revealing that the client currently pays most of its vendors on set dates (e.g. monthly or bi-monthly), we explored an alternative in which the client transitions into a more dynamic process during which an invoice is either cleared:

1. As late as is allowed by payment terms with its vendors, in order to optimize the company’s cash flow.
2. Or, as early as reasonable in order to potentially negotiate an early-payment discount. Industry standard is often a 2% discount for paying within 10 days of the vendor’s transmitting an invoice to the company [7].

We begin by quantifying the cumulative amount of bill-to-pay time historically taken up by simply waiting for invoice clearance. Bill-to-pay time is assumed to be the transition time between a “Vendor creates invoice” in a case log, to the first “Clear Invoice.” Invoice clearance wait time is the duration between invoice clearance and any preceding major event, such as Record Goods Receipt.

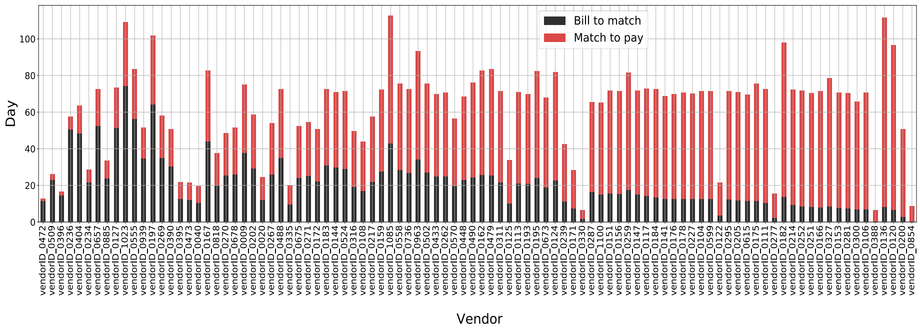


Fig. 12. Match-to-pay wait time (red) as a component of total bill-to-pay time for high-value vendors, sorted by fraction of red days.

Fig. 12 shows the median wait time to invoice clearance as a component of median total bill-to-pay time for high-value vendors with total item values worth over EUR1.5 million. We can see that the company pays certain vendors soon after a potential match is achieved across the purchase order, invoice, and/or goods receipt record. However, transactions with other vendors exhibit match-to-pay taking up the vast majority of process duration.

Since the data provided does not contain any information on payment terms, we cannot quantify the potential benefits of delaying payments to the limit specified by the terms. Instead, we attempt to quantify the theoretical total opportunity for savings if the company were to pay its invoices within 10 days to obtain a 2% discount on the value of the purchase item.

Within the scope of the available dataset, we assume that items that were paid more than 10 days after an invoice is created by the vendor, with goods received before or within 3 days of invoice receipt, would have been eligible for a 2% discount. Such completed cases correspond to 143,056 items, or 57% of items in the dataset. Taking the total value of those cases, a 2% discount represents EUR9.16 million in maximum potential savings to the company.

Note that this analysis could be altered to model more customized early payment schemes, various discount rates, or any opportunity costs to optimize the client's cash flow, if details of payment terms the client has negotiated with its vendors are provided.

We recommend that the client consider a transition towards a higher frequency invoice payment schedule (e.g. weekly, or even daily) where invoice clearance wait time is minimized when possible, or otherwise optimized with respect to the company's cashflow. An ability to pay more frequently may also give the company better overall negotiating power over pricing with vendors. While there are costs associated with such changes (e.g. restructuring the workflow of payroll departments or efforts associated with the computation of optimal invoice clearing), our preliminary calculations show great potential for cost savings from the ability of the company to negotiate and execute on early payment discounts.

5.4 Vendor Process Complexity

Quantifying the complexity of processes associated with certain vendors can lead to insights that may help the company identify opportunities to improve suboptimal processes. In this section we attempt to rank vendors based on their process complexity by looking at a combination of total vendor cases and number of process variants. This analysis looks at complexity in a holistic manner, by encompassing all events that happen in a given case for a given vendor.

Methodology. Once again, we focus on the "3-way-before" process. To evaluate complexity, we perform the following actions.

1. Process variants: We calculate the number of unique activity sequences completed for each vendor (ordered and unordered).
2. Total cases: We count the total number of cases associated with each vendor.
3. Complexity metric: We divide the number of cases by the number of unique variants for each vendor.
4. Subset high-opportunity vendors: We focus on vendors whose total case count is above the 90th percentile to maximize the chances that any process improvement for those vendors will have a meaningful overall impact.

Table 6. Top 5 Most Complex Vendors (90th percentile for Total Cases)

Vendor	Total Cases	Unique Variants	Cases per Variant
vendorID_0673	408	178	2
vendorID_0193	255	87	3
vendorID_0502	422	134	3
vendorID_0144	302	85	3
vendorID_0183	296	82	4

Table 6 shows that for the most complex vendors, there are only approximately 2-4 cases per variant. For comparison, the most consistent vendors in the 90th percentile have a case to variant ratio of 40 or more (with the most consistent vendor, vendorID_ 0550 displaying a ratio of 108) as shown in Fig. 18 in the appendix.

Analysis. Upon inspection of the processes associated with the problem vendors, we notice that many of these processes exhibit certain commonalities. Processes often include many "Change" type events, such as "Change Quantity" or "Change Price." While these activities are perhaps necessary, we believe that managers should investigate to ensure that needless complication is not occurring on the company side. We also note that these events tend to lengthen processes by contributing to backlog, as discussed in the prior section. These events may also be driven by the vendors. Further action could include working with vendors to help minimize these "noise" events, or even switching to vendors who may have more stable and efficient processes.

Fig. 13 explores the relationship between vendor complexity, case processing times, and vendor valuation. Each bubble represents a vendor at the top 90th percentile by case count. Time to resolve (TTR) is assumed to be the time between “Vendor creates invoice” and the activity preceding “Clear Invoice” to avoid the effect of any confounding variables in the analysis (e.g. certain invoices experienced a lot wait time to clear, certain items are created without work being done on it).

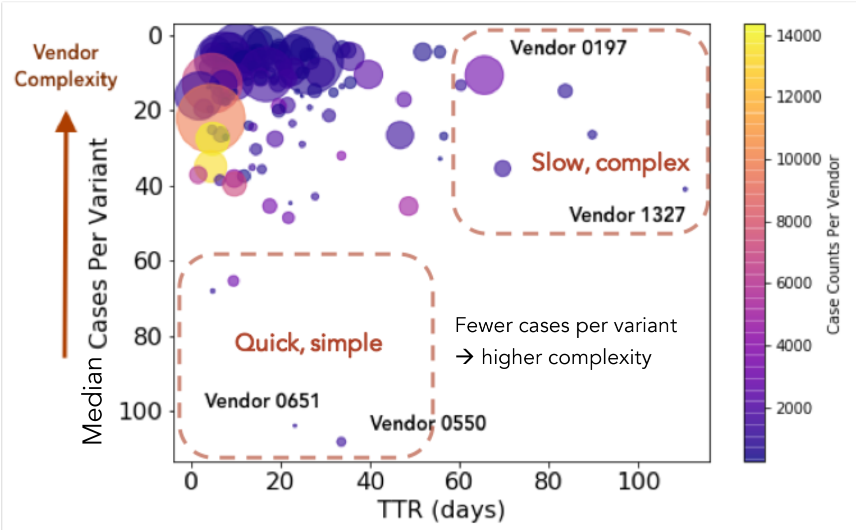


Fig. 13. Vendor complexity, median time to resolve (TTR), vendor valuation (cumulative amount of all purchase order items, represented by the size of the bubble), and case counts per vendor.

Vendor complexity moves inversely with the number of cases per variant (CPV). A higher CPV implies that a vendor has a standardized flow of activities to process a high number of cases and is thus less complex. The following insights can be derived from this analysis:

1. Many high-volume and/or high-value vendors such as vendor 0136 and vendor 0104 tend to be relatively complex. This may be thought of as a natural interpretation of complexity; vendors that are associated with a greater number of cases are likely to have a wider variety of unique process sequences, and items that are valuable may need to be processed with greater caution. We can note, however, that higher complexity does not necessarily translate to longer processing times, as seen towards the top left quadrant of the plot.

2. Cases from vendors such as 0197 may be problematic, given the relatively high complexity, time, case count, value, and resolution time. In relation to an earlier analysis on the monetary impact of dynamic invoice payments, high processing times may have an additional opportunity cost of lost potential to negotiate early payment discounts with vendors.

6 Social Network Analysis

6.1 Key Findings and Recommendations

User Relationship Analysis. A similar-task miner leads to the insight that, in 3,384 complete cases worth EUR61.4 million (1.7% of complete cases by number but 8.6% by value), the same human user performs "Create Purchase Order Item" as well as at least one instance of "Record Goods Receipt." This may raise the potential for fraud or errors by limiting checks and balances. Other combinations of compliance-related steps by the same human user are also present and may warrant further investigation. In order to direct the company's attention to other potential anomalies, we also highlight human users who seldom work together, hand off work or subcontract work.

User Activity Patterns. As expected, resources specialize in particular tasks and work Monday to Friday during business hours. The most prolific users who perform "Clear Invoice" seem to concentrate their activity heavily on Thursdays, which confirms a pattern observed in the payment terms section of this report that may need to be adjusted. We also flag how some human users work at unusual weekday hours or on Saturday mornings while others seem to schedule batch-like activity to occur overnight.

6.2 User Relationship Analysis

Segregation of Duties. Aside from the control-flow perspective on processes, valuable insights can be gained by looking at the resources (human users or batch systems) associated with each event. This is known as social-network analysis ("SNA"). SNA can be performed by using the resource column instead of the activity column in the event log when discovering a process map.

Since the data contains over 600 resources, the resulting diagram is too complicated to reasonably depict the flow of cases among resources. Therefore, we decided to use ProM for two different approaches, focusing on resources that performed more than 3,000 compliance-related events ("Create Purchase Order Item," "Record Goods Receipt," "Record Invoice Receipt," "Remove Payment Block" and "Clear Invoice") in complete cases (since these are the cases where payment was actually made). These 112 out of 627 resources account for 84% of all such events.

Segregation of duties within the purchasing process helps prevent errors and fraud as multiple resources check each other's work on a case. Using SNA, we identify human resources who are performing compliance-related activities that should be segregated, as well as several who operate more on-their-own than most, also a potential red flag. Finally, we note that no resource ("NONE") is recorded for 4% of all "Clear Invoice" events (7% by value). This percentage can be as high as 14% for some common sub-spend areas such as Packaging. "Clear Invoice" resources should always be recorded for better control of payments.

Similar-Task Miner. The first approach involves a group of ProM SNA modules (the second, involving dotted charts, is discussed in section 6.3 below). We began with the similar-task (ST) miner, which organizes groups of resources within "roles" (similar mixes of activities). We found that correlation (rather than Euclidean distance, similarity coefficient or Hamming distance) within the resource-activity matrix provided the best separation of roles.

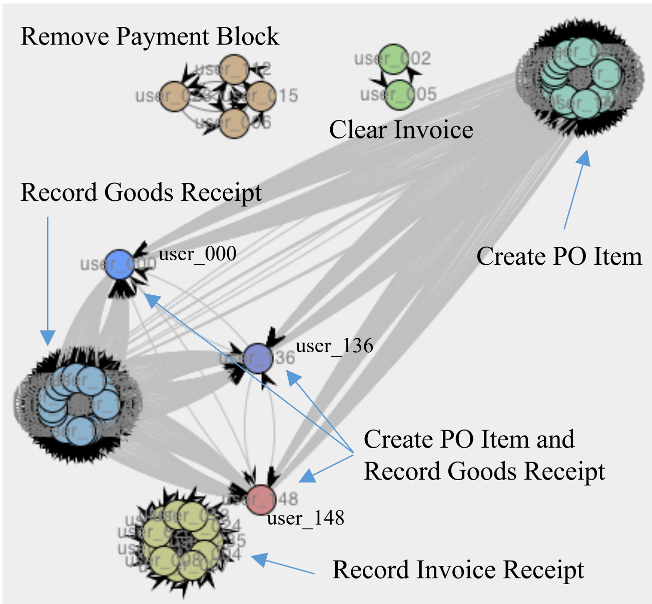


Fig. 14. Similar-task social network of human users performing over 3,000 compliance-related events (all complete cases)

In **Fig. 14**, the isolated circles for user_000, user_136 and user_148 represent users who most frequently perform both "Create Purchase Order Item" and "Record Goods Receipt" in the same case, alerting us to the presence of this potentially risky behavior. There are 44 human resources who do this though some are not active enough to appear in the diagram. As one would expect, the three problematic users are shown between a cluster of users who largely perform "Record Goods Receipt" and those who largely perform "Create Purchase Order Item." The other clusters largely perform "Clear Invoice," "Remove Payment Block" and "Record Invoice Receipt."

Breakdown of Problematic Cases. Of the 3,384 cases where the same human user performs "Create Purchase Order Item" as well as at least one instance of "Record Goods Receipt," 1,256 have Item Type Service, which is 49% of complete Service cases. This compares very unfavorably to the 0.8% of Item Type Standard cases that are problematic. Since all Service cases are 3-way-after, 3-way-after's percentage of problematic cases is also elevated at 14% (vs. 0.1% for 3-way-before).

The Spend Areas with the highest percentage of problematic cases are Workforce Services (79%), Enterprise Services (71%), Logistics (35%), Marketing (14%) and CAPEX & SOCS (9%). The most problematic vendors with more than 100 cases are vendorID_0003 (100%), vendorID_0000 (100%), vendorID_0741 (79%), vendorID_0277 (67%) and vendorID_1466 (19%).

Less-Connected Resources. Other ProM SNA modules that give insight include Hand-over of Work ("HoW"), Working Together ("WT") and Subcontracting ("SC"). These all indicate that resources in general hand off work (activities directly follow), work together (resources working on the same case) and subcontract (perform an activity both before and after another resource) with any and all other resources. Below we show some resources that tend to be less connected in these senses to other users so that the process owner can investigate whether reduced checks and balances with respect to these system users represents a risk (see appendix for graphical representations of these social networks):

- WT: user_005, user_087, user_136, user_171, user_186
- HoW: user_082, user_121, user_157, user_200
- SC: 40 of the 112 most active users do not subcontract at all

6.3 User Activity Patterns

The second approach uses dotted charts to compactly show when various resources perform various activities. This reveals the basic fact that resources tend to specialize in one particular activity while infrequently performing others. It also shows the weekly and daily work pattern where most activities by human users take place during business hours (around 8am-6pm weekdays). Unsurprisingly, batch resources work around the clock and on weekends.

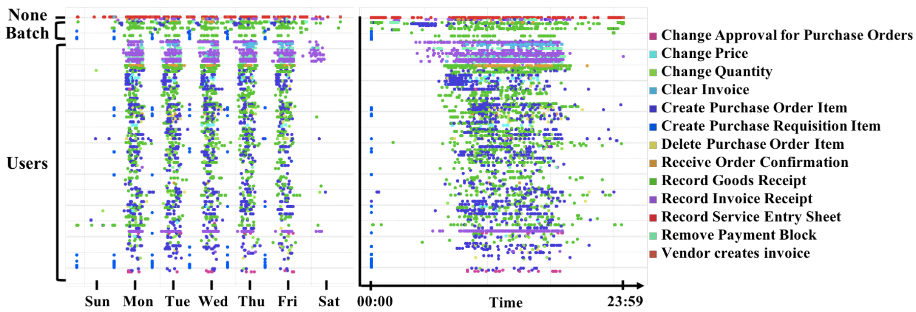


Fig. 15. Timing of most-frequent activities during the week and day by resource (those performing over 3,000 events)

Certain users seem to schedule "Create Purchase Requisition Item" to take place at midnight or 1am on Sunday-Thursday nights. This activity becomes much more common starting in September 2018, with several resources specializing in this activity added at that time. Some users concentrate their activities in the first few hours of the day, tapering off in the afternoon. A few users start work very early in the morning while others tend to arrive later but work into the evening. Finally, other users, often those performing "Record Invoice Receipt," tend to work on Saturday mornings. User_002 and user_005 are by far the most active "Clear Invoice" specialists, but user_002 does not work on Mondays, and user_005 does not work on Mondays or Wednesdays, perhaps contributing to the delays discussed in the payment terms section of this report.

7 Conclusion and Next Steps

In this report we explored a number of lines of inquiry which we believe have the potential to help the company enhance its processes, lower risks, and realize cost savings. We recommend that the process owner take steps to modify its payment processes to enable more dynamic payments to vendors. We also believe the process owner should investigate the noncompliant processes highlighted in the report to address potential risks. Further automation of processes could lead to a reduction in payroll costs. Monitoring vendors associated with high process complexity could lead to simpler and more efficient operations. Finally, our social network analysis points to instances where the segregation of duties is compromised and should be reviewed.

The data provided in the competition certainly gave us an intimate look into the way the subject company manages its purchase orders. However, this dataset only provides part of the picture. Organizations, especially complex ones, generate data in myriad ways. We believe that many of the findings in this report could be significantly enhanced by looking at other kinds of data produced by the subject company, such as employee rosters, inventory data, warehouse shipment delivery records, financial transactions, and even purchase order contracts with various vendors. Bringing disparate data sources together is a powerful way to gain a fuller understanding of what is actually happening within the enterprise at various levels. Additionally, we would also attempt to perform research on how the data is collected by speaking with users of the systems in question. Fundamentally understanding data collection is key to building a solid foundation for any analysis, and it provides solid ground for providing robust recommendations to decision-makers.

Appendix

Data Description.

Of the twenty-two columns in the dataset, we can immediately set aside five as not adding information:

- "case Purch. Doc. Category name": One value: "Purchase order."
- "case Source": One value, "sourceSystemID_0000."
- "case GR-Based Inv. Verif." and "case Goods Receipt": These True/False indicators, taken together, indicate which of the four main sub-processes a case follows, but "case Item Category" (see description below) already does this directly.
- "case Goods Receipt": See above.
- "event User": This column contains the same values as "event org:resource" (see below).

Of the remaining seventeen columns, five can vary within a case, i.e. they are properties of each event:

- "eventID": 1,587,802 values. Unique identifier for each event. We do not drop events that are duplicates aside from eventID.
- "event concept:name" (also known as "Activity"): 42 values. The action performed. Most common in descending order (10-20% of all events each): "Record Goods Receipt," "Create Purchase Order Item," "Record Invoice Receipt," "Vendor Creates Invoice," "Clear Invoice" and "Record Service Entry Sheet."
- "event org:resource" (also known as "Resource"): 627 values. The person or system who performs the "Activity." 25% of values are "NONE." Twenty non-human "batch" resources, the most common of which is "batch_06" (2.4% of events). 606 human users, the most common of which is "user_002" (10% of events).
- "event time:timestamp" (also known as "Complete Timestamp"): 166,419 values. The time/date when an event happens. Often several events within a case occur at the exact same time, which means that process maps discovered from this event log are accurate only up to a point as it is impossible to determine the "correct" order of simultaneous events.
- "event Cumulative net worth (EUR)" 25,164 values. The monetary value associated with each event. The distribution of these values by case – taken as the EUR value recorded upon a "Create Purchase Order Item" event – is highly skewed to low numbers (including zero for all of the Consignment cases) but ranges up to EUR8.8 million. The median purchase-order item is worth around EUR500. There are entries of amounts up to EUR28,994,530 associated with other activities in the log, but they do not match purchase-order item values and may be incorrect.

Finally, twelve of the columns are case variables, i.e. they are the same for every event in a given case:

- “case concept:name” (also known as “Case ID”): 251,463 values. The identification number for each case/purchase order item. It starts with the number contained in “case Purchasing Document,” followed by an underscore (“_”), followed by the number contained in “case Item.”
- “case Company”: 4 values. The subsidiaries of the coatings/paint company that have purchases in the log. “companyID_0000’ dominates with 99.6% of events. Almost all of the remaining 0.4% of events belong to “companyID_0003.” These 5,758 events (1,027 cases) are precisely those that have “case Item Category” as “2-way match” as well as those that have “case Item Type” as “Limit.” This group of cases has its own exclusive set of vendors while the other three sub-processes have overlap among their vendors. Finally, this group of cases all have “case Document Type” as “Framework order,” but there are many other “Framework order” events not in this group of cases. The few cases for “companyID_0001” (9 events) and “companyID_0002” (6 events) seem to be standard 3-way-match purchase orders.
- “case Document Type”: 3 values. Consistent across all cases/items within a “Purchasing Document.” “Standard PO” for 96.9% of events. “Framework order” accounts for 1.8% of events (see discussion of “companyID_0003” above). The remaining 1.4% of events are “EC Purchase order” and make up precisely the 1,425 cases that involve activities with “SRM” in the name. SRM stands for “Supplier Relationship Management,” an SAP system for efficient production of purchase orders in the field.
- “case Item”: 490 values. The item number within a Purchasing Document (see “Case ID” above). Item numbers tend to follow the convention 00010, 00020, 00030 … 00100, 00110 ... for each consecutive item in a Purchasing Document. In several documents item numbers deviate from this pattern (e.g. 53XXX7 from vendor 0197, 111X9 from vendor 0262, and 0000X from certain Service vendors). We encourage the company and vendors to adhere to a common numbering scheme for simplicity.
- “case Item Category”: 4 values. Indicates which of the four main sub-processes (2-way match et. al.) the case belongs to: “3-way match, invoice before GR” (78% of events), “3-way match, invoice after GR” (20%), “Consignment” (2%), and “2-way match” (0.4%). See compliance discussion for a description of how we break these sub-processes down further.
- “case Item Type”: 6 values. “Standard” for 78% of events. “Service” (16%) seems to be a special sub-process within “3-way match, invoice after GR” (see compliance section). Item Type “Consignment” labels exactly the same cases that have “case Item Category” as “Consignment.” “Limit” is discussed above under “case Company.”
- “case Name”: 1,886 values. This column pertains to vendors (who is selling the items to the company), but for vendors we decided to use the closely related but slightly more detailed column “case Vendor” (described below).
- “case Purchasing Document”: 76,273 values. The first part of the “case concept:name” identifier: unique 10-digit numbers, beginning with 2 (for Document

Type “EC Purchase order”) or 4 (Document Types “Standard PO” and “Framework order”). While 63% of documents contain only one item, a purchase document can have as many as 429 items, which tend to be created either at or around the same time. Within a purchasing document, items can be associated with multiple spend areas or item types but with only one vendor, document type or company. Service orders are largely single-item orders.

- “case Spend area text”: 21 values. The type of goods/services procured, top values being “Packaging,” “Sales” and “Logistics.” 1% blank entries.
- “case Spend classification text”: 4 values. A higher-level classification of goods/services procured. Same 1% of events blank as for the other “Spend” columns. “PR” (56% of events) stands for product-related, items that are raw material for the company’s products (accounted for as Cost of Goods Sold). “PR” comprises ten of the “case Spend area text” values (Packaging, Trading & End Products, Additives, Latex & Monomers, Solvents, Pigments & Colorants, Specialty Resins, Titanium Dioxides, Commodity Resins and Chemicals & Intermediates). “NPR” (42% of events) stands for non-product-related, e.g. capital expenditure or overhead, and comprises eight of the “case Spend area text” values (Sales, Logistics, CAPEX & SOCS, Marketing, Enterprise Services, Real Estate, Workforce Services and Energy). “OTHER” comprises two of the “case Spend area text” values (Others and Spend Area Unidentified).
- “case Sub spend area text”: 136 values. A lower-level classification of goods/services procured, most commonly “Products for Resale” (21% of events). Same 1% of events blank as for the other “Spend” columns.
- “case Vendor”: 1,961 values. Each vendor in the “case Vendor” column maps to exactly one vendor in the “case Name” column, but since there are 75 more values of “case Vendor,” each vendor in the “case Name” column maps to one or more vendor in the “case Vendor” column. The most extreme example of this is “case Name” vendor_0143, which maps to 11 different vendors in the “case Vendor” column.

Compliance Analysis: Random Forest Classifier.

The high dimensional nature of this dataset poses a challenge. Given so many categories of data to focus on, how do we narrow down which areas are worthy of analysis? A random forest classifier was used to help uncover patterns behind compliance. Given an input of event logs and a binary compliance flag, the classifier splits this input into random subsets and send each of them through a decision tree. Each tree then 'votes' on the outcome (i.e. compliant or non-compliant). These results are then aggregated, and the majority vote for each case of event is deemed the final prediction. The result of this classifier can be compared to the real compliance flag, and the model's key features (i.e. the most important characteristics of a case or event that delineates compliance) examined.

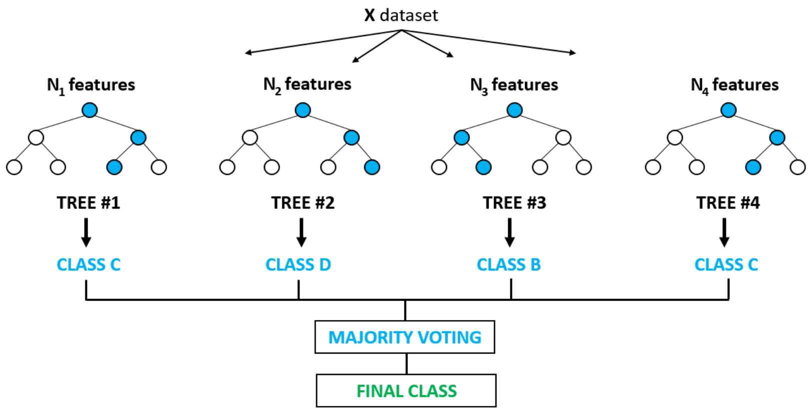


Fig. 16. Random Forest classifier. The dataset, a matrix containing event or case information, is split into *n* decision trees. Each tree votes on the best binary outcome. The votes are then aggregated, and the majority vote determines the final class outcome. [3]

Using a scikit-learn implementation of the random forest classifier [6], feature importance can be explicitly defined as a reduction in Gini impurity, or the likelihood that a datapoint is incorrectly labelled if it was classified according to the distribution of labels in that subset of the data, defined as [4].

$$G = \sum_{i=1}^{J} p_i(1 - p_i)$$

for a dataset with J classes, $i \in [1, 2, \ldots, J]$, and p_i being the fraction of items labeled with class i in the dataset. We want each split in the tree to be the most informative and delineating as possible. Feature importance (FI) is represented by the weighted information gain at each node, defined as [5],

$$FI = \frac{N_t}{N} \times (G_{parent} - \frac{N_{tR}}{N_t} * G_R - \frac{N_{tL}}{N_t} \times G_L)$$

where N is the total number of samples, N_t the number of samples at each node, N_{tR} is the number of samples at the right child, and N_{tL} at the left. As the name suggests, the higher feature importance, the more important and information a characteristic of the data is. We utilized the classifier as a data exploration tool to help uncover meaningful data segmentations given a dependent variable, such as compliance. Note that we chose a random-forest classifier over other models (such as logistic regression) due to its more intuitive nature and higher accuracy.

Figures.

Fig. 17 shows the distribution of median elapsed time between “Vendor creates invoice” and “Clear Invoice” grouped by vendor. Peaks are observed at around 22, 50, and 70 days.

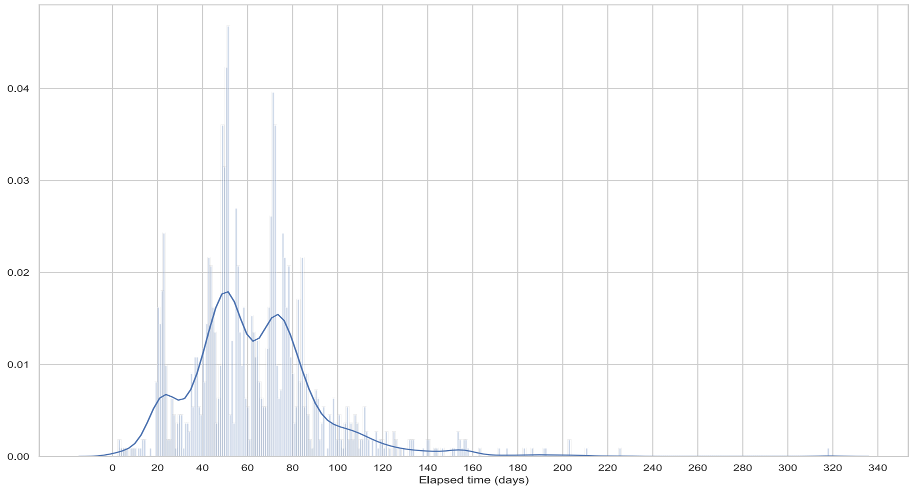


Fig. 17. Distribution of median elapsed time by vendor between “Vendor creates invoice” and “Clear Invoice” events.

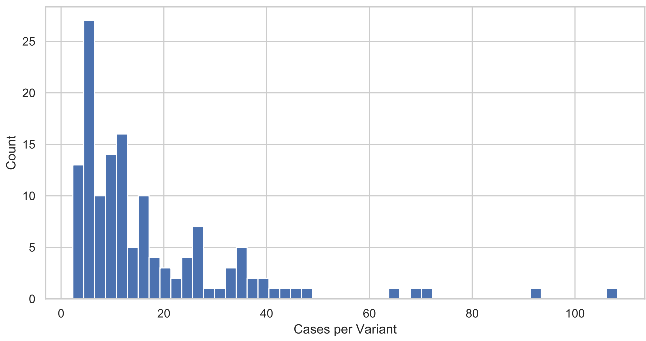


Fig. 18. Histogram of Cases per Variant by Vendor (90th Percentile of Total Case Count)

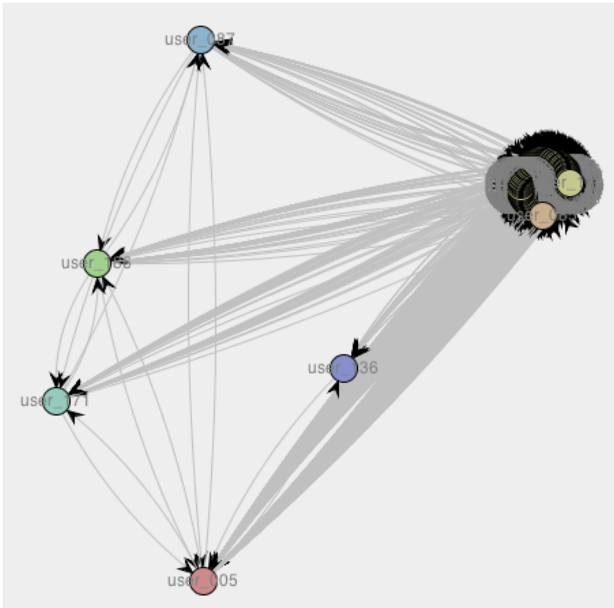


Fig. 19. Working-together social network of human users performing over 3,000 compliance-related events (all complete cases)

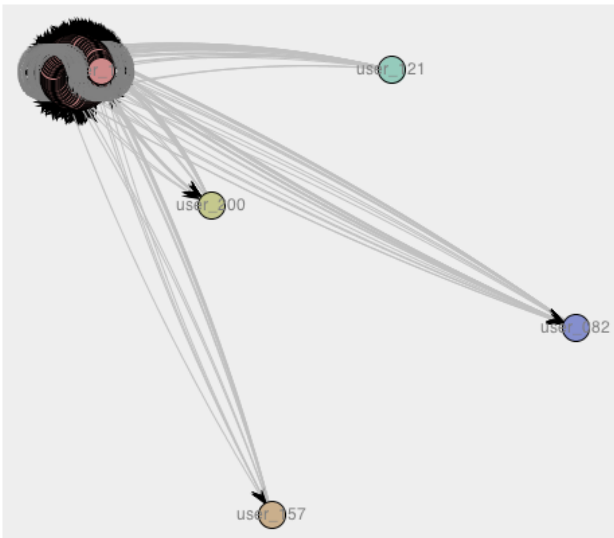


Fig. 20. Handover-of-work social network of human users performing over 3,000 compliance-related events (all complete cases)

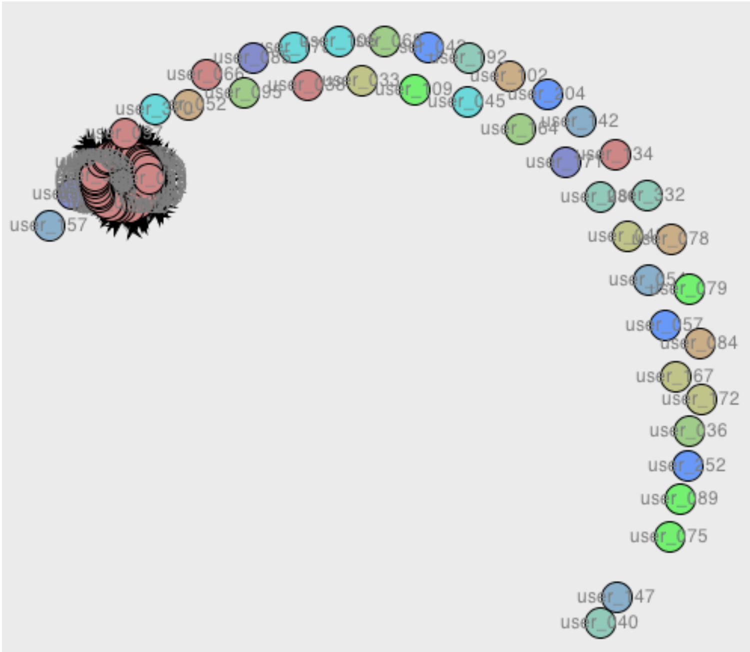


Fig. 21. Subcontracting social network of human users performing over 3,000 compliance-related events (all complete cases)

Tables.

Table 7. Case Completion Statistics (Expanded)

	Number	% by Number	Value (EUR millions)	% by Value	Mean Value	Median Value
Completed cases	196,881	73.2%	711.6	78.3%	3,615	491
Completed, with multiple invoice clearing[7]	6,774	2.7%	126.3	13.0%	18,638	1,251
Non-completed	54,582	26.8%	260.1	21.7%	4,766	565
Non-completed with Delete PO	8,561	3.4%	55.6	5.7%	6,500	1,125
Total	251,463	100%	971.8	100%	3,864	508

7 "Clear Invoice" occurs up to 71 times in a case. These multiple-clear-invoice cases are part of all sub-processes except Consignment and are spread across all document types, item types, spend classifications etc.

References

1. van Dongen, B.F., Dataset BPI Challenge 2019. 4TU.Centre for Research Data. https://doi.org/10.4121/uuid:d06aff4b-79f0-45e6-8ec8-e19730c248f1
2. Process Mining Conference. BPI Challenge 2019. https://icpmconference.org/icpm-2019/contests-challenges/bpi-challenge-2019/
3. Global Software Support. Random Forest Classifier – Machine Learning. https://www.globalsoftwaresupport.com/random-forest-classifier-bagging-machine-learning/, last accessed May 7, 2019
4. Hauskrecht, M., Decision trees. https://people.cs.pitt.edu/~milos/courses/cs2750-Spring03/lectures/class19.pdf, last accessed May 7, 2019
5. Scikit-learn: Random Forest Classifier. https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html, last accessed May 7, 2019
6. Pedregosa et al., Scikit-learn: Machine Learning in Python, JMLR 12, pp. 2825-2830, 2011.
7. Schaeffer MS. Essentials of accounts payable. John Wiley & Sons; October 15, 2002.