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

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).

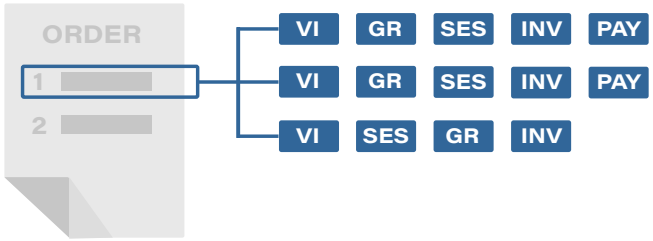
The goal of this report is to help this company understand their processes and potentially help them prevent faults or inefficiencies that would otherwise have gone unnoticed. Special care is given to compliancy, as it is given that the process owner has compliance questions.

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.

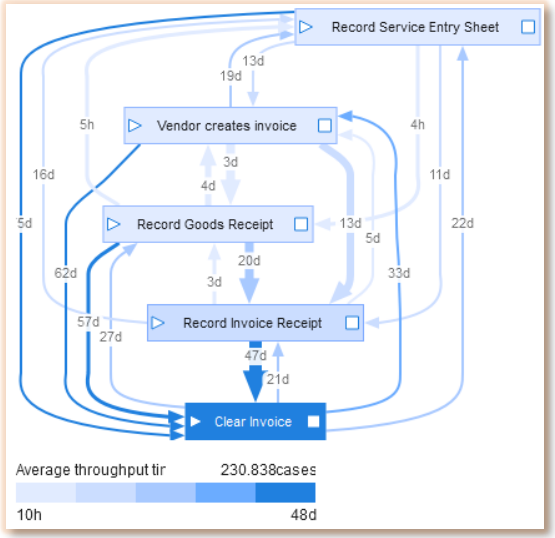


Figure 2: Invoicing process throughput times

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.

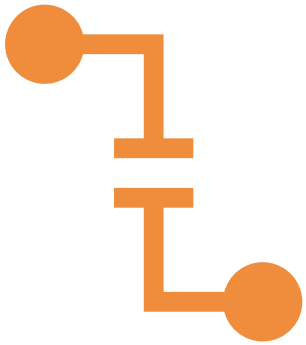
During our analysis we consistently use the PO amount as a reality check for our conclusion. We found out that for the other amounts (for instance clear invoice) the total amounts were much higher, as a result of the many iterations. This influenced our decision on how to approach this challenge. Normally 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.

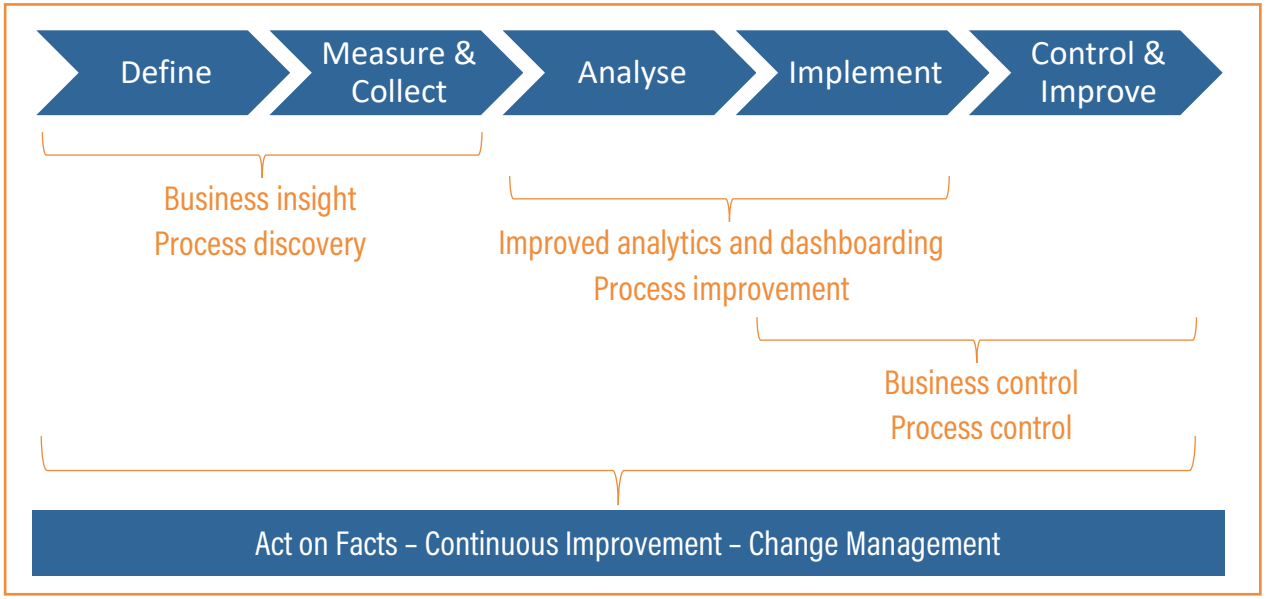


Figure 5: DMAIC applied in the Analytics for Improvement approach

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.

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)

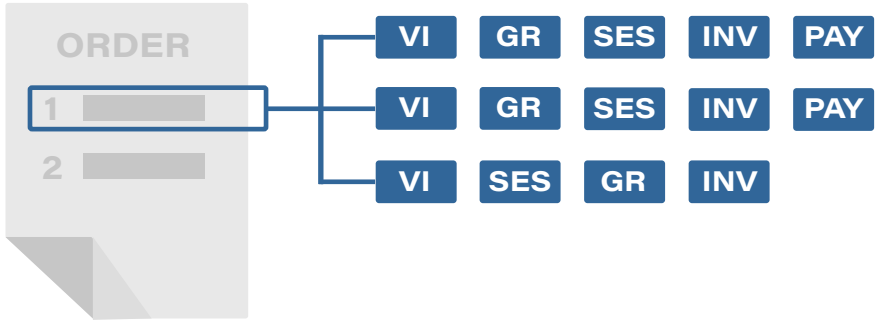


Figure 7: Structure of a purchase order

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.

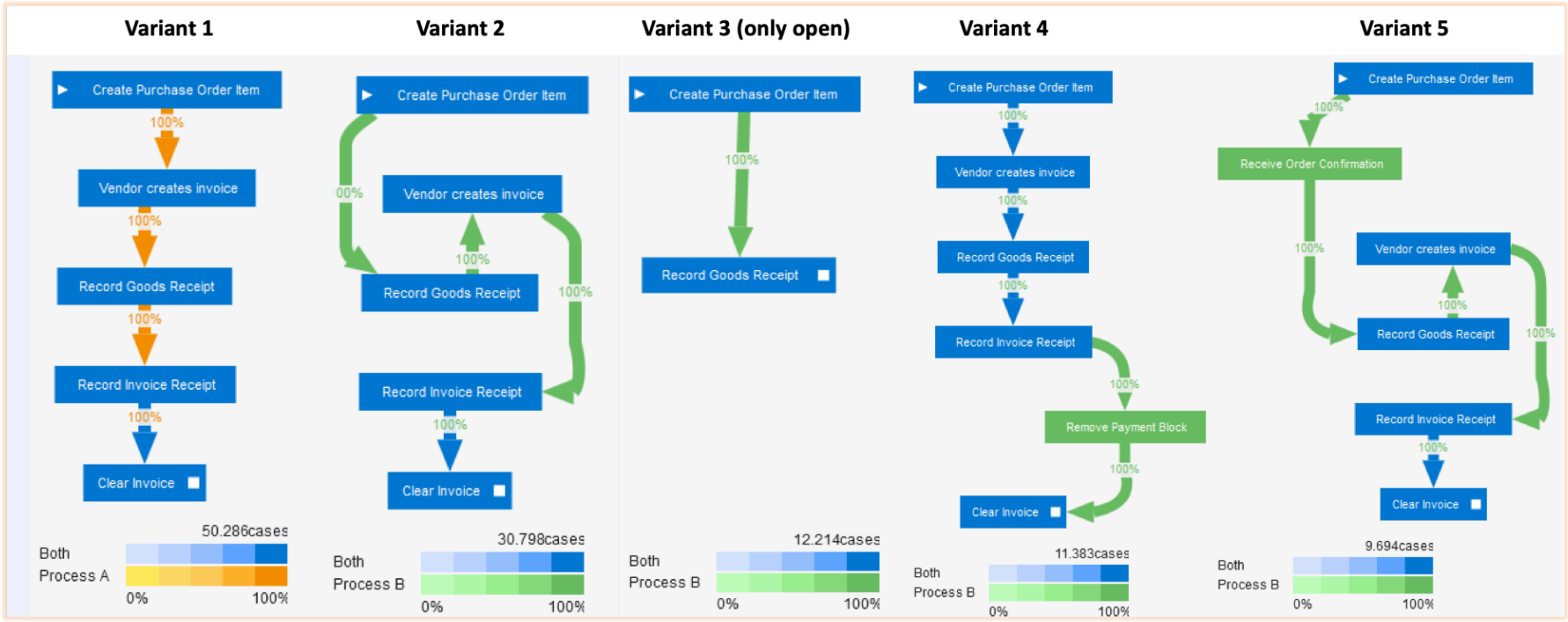


Figure 8: Graphs of the 5 most frequent process variants

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.

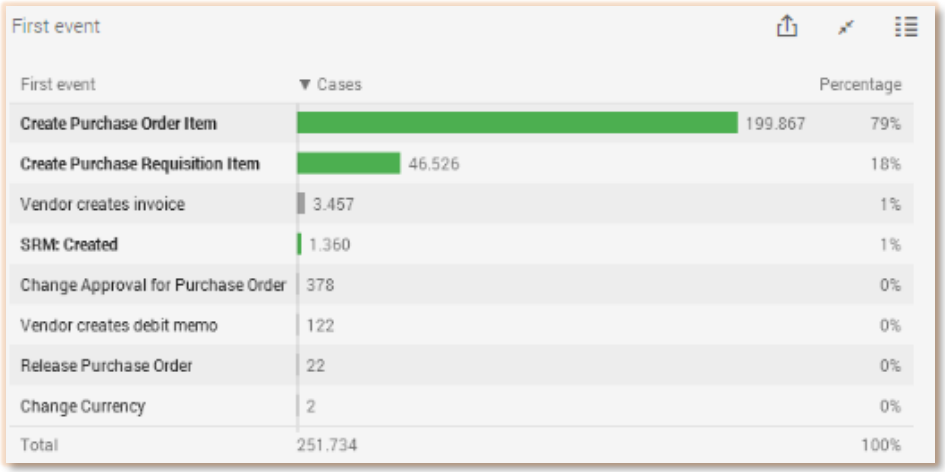


Figure 9: First activity of all purchase order line items

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.

PO's	PO lines	Process iterations	Events	PO amount	Activities	Variants	Vendors	Items	Users
1.376	3.981	11.093	41.150	€ 52.591.393	34	649	597	99	262
Σ 76.349 | 2%	Σ 251.734 | 2%	Σ 422.606 | 3%	Σ 1.595.923 | 3%	Σ € 996.818.747 | 5%	Σ 42 | 81%	Σ 11.973 | 5%	Σ 1.975 | 30%	Σ 490 | 30%	Σ 628 | 42%

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.

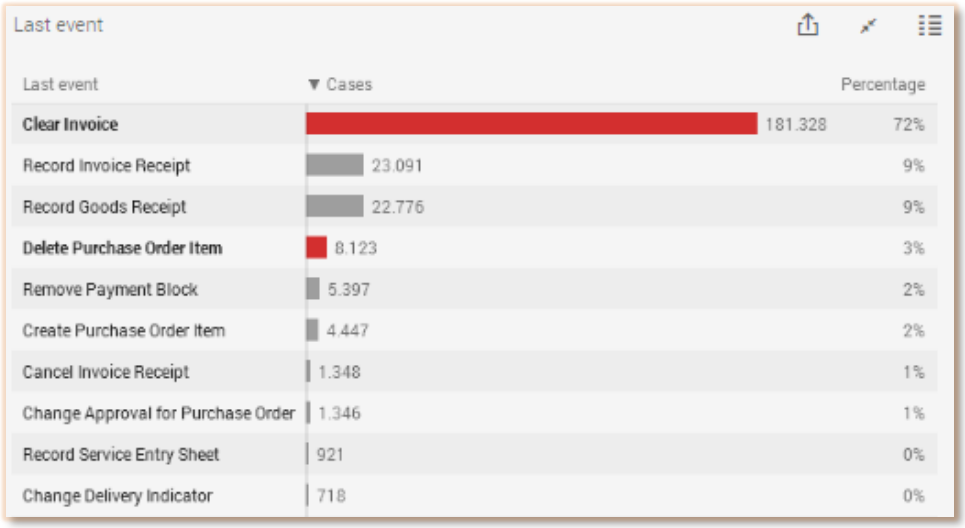


Figure 11: Top 10 last activity of all purchase order line items

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.

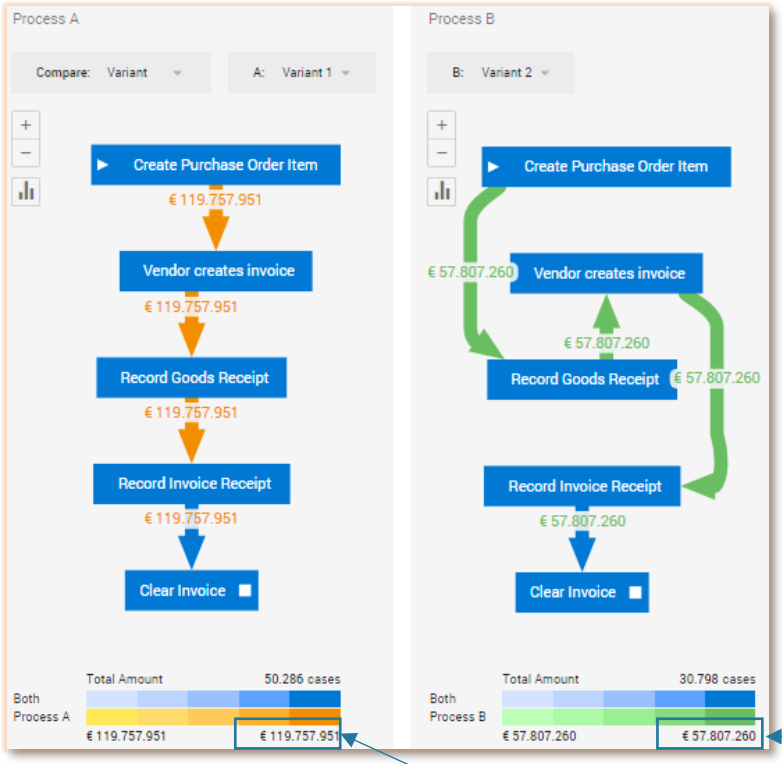


Figure 15: Two biggest variants compared (based on PO amount)

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

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.

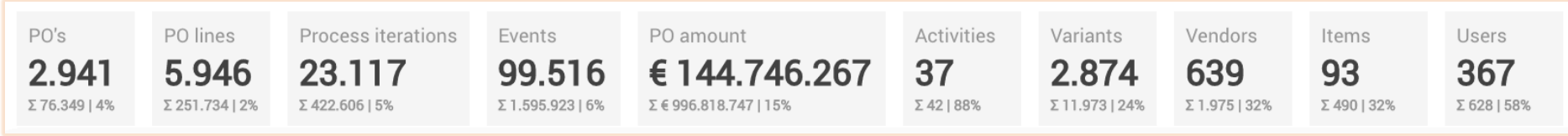


Figure 20: A4I footer. Purchase order line items that have multiple invoices and payments

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).

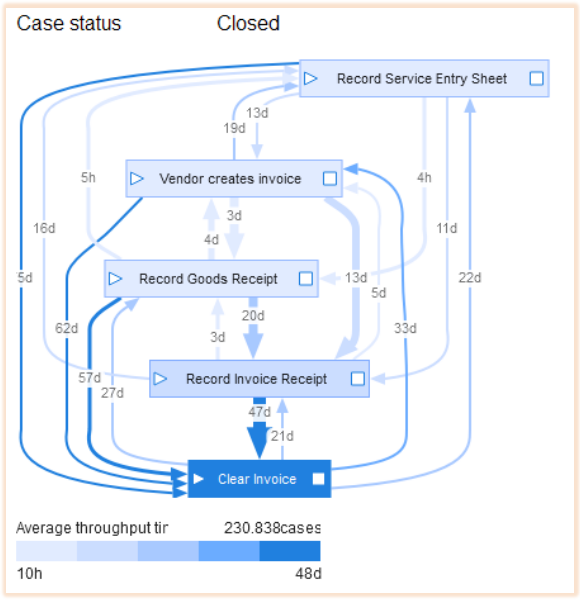


Figure 22: Invoicing process throughput times

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 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

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.

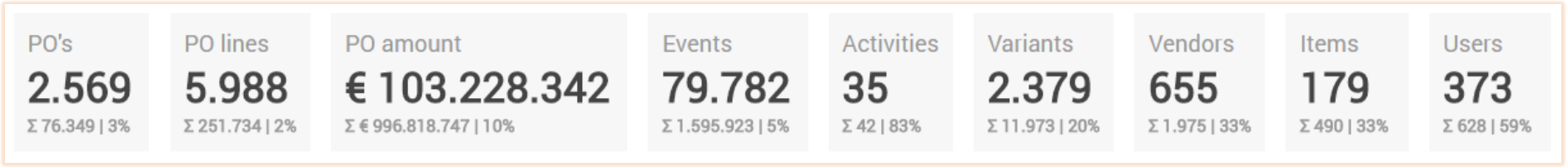


Figure 25: A4I footer. Purchase order line items which contain the event "Vendor creates debit memo"

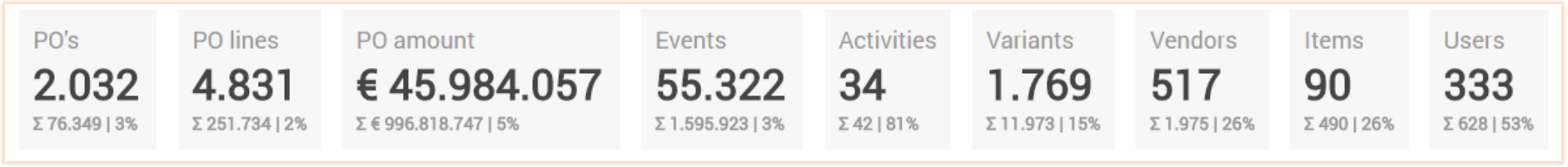


Figure 26: A4I footer. Purchase order line items which contain the event "Vendor creates debit memo", as well as "Change Price" or "Change Quantity"

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.

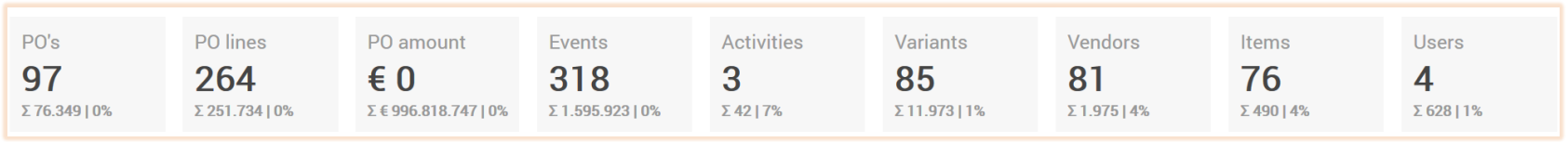


Figure 32: A4I footer. Purchase order line items with events from before 2018

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