BPI Challenge 2019

Tommaso Romanò
University of Milan

Abstract

The report deals with the detailed analysis of this log of a Netherlands company, 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. Then, conformance checking and process performance focused on the throughput time and rework.

https://colab.research.google.com/drive/1zoqSu5zu-wHsZM8qDhL6WUHA0vSFs6GM?usp=sharing

1 Data Understanding

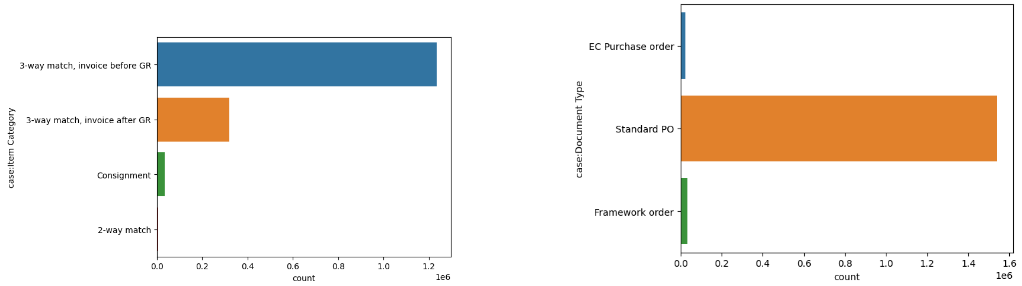
1.1 Data

The log contains 1,595,923 events, belonging to 251,734 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:concept:name (Case ID) A combination of PO ID and item ID as the case identifier
- concept:name (Activity) The name of the activity that the events refer to
- time:timestamp of activity completion
- org:resource (User) the resources recording the activity

Majority of attributes, however, are on the case level recorded for each item including:

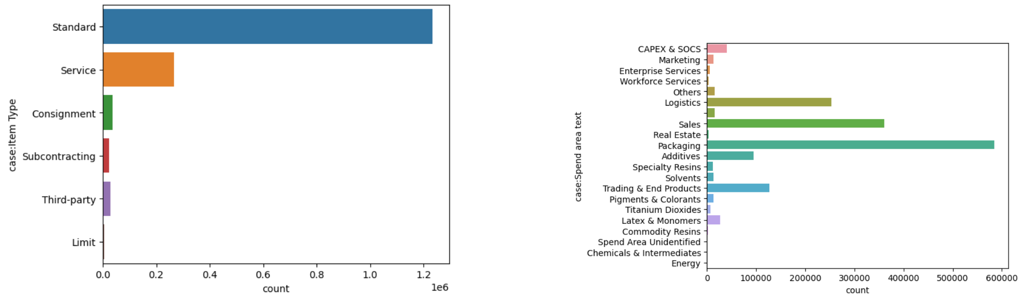
- case:Company The anonymized ID of the respective subsidiary that the case relates to
- case:Name Anonymized name of the vendor
- case:Spend area text The purchasing area
- case:Cumulative net worth The value of the item in Euro
- case:Document type The type of purchasing document
- case:Item category The invoicing procedure is determined based on this attribute
- case:Item type: the type of the item



(a) Countplot of Item Category

(b) Countplot of Document Type

Figure 1



(a) Countplot of Item Type

(b) Countplot of Spend Area Text

Figure 2: Caountplots of different attributes

The most important attributes that will be used during the challenge are: case:concept:name, case:Item Category, case: Document Type. In particular, the level of analysis is this: Item Category > Document Type > Item Type > Spend Area Text > Sub Spend Area Text. There are 4 different Item Category:

- 3-way match, invoice before GR (221,010 and 88% of cases): Purchase Items that do require a goods receipt message, while they do not require GR-based invoicing. For such purchase items, 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.
- 3-way match, invoice after GR (15.182 and 6% of cases): 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.
- 2-way matching (1,044 and 0.5% of cases): For these items, the value of the invoice should match the value at creation, but there is no separate goods receipt message required.
- Consignment (14,498 and 5.5% of cases): 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.

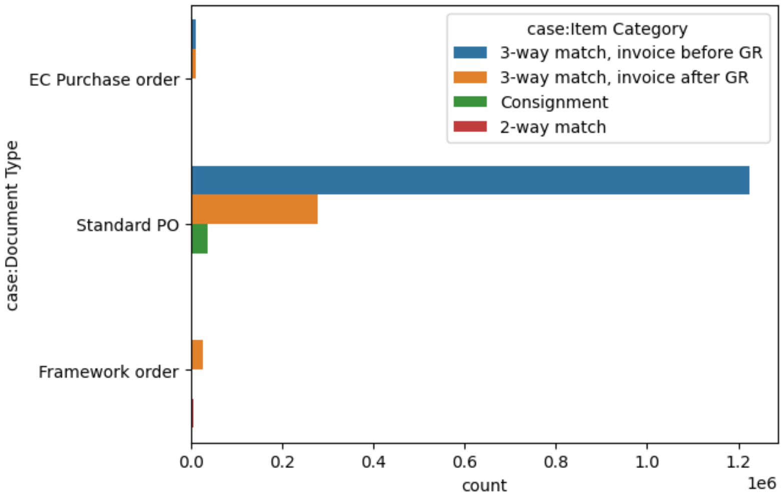


Figure 3: Countplot relation between Item Category and Document Type

The relation between these attributes is that the major of the document types is Standard PO with the 3-way match, invoice before GR. It is important to notice that the Framework Order is the only document type for the 2-way match.

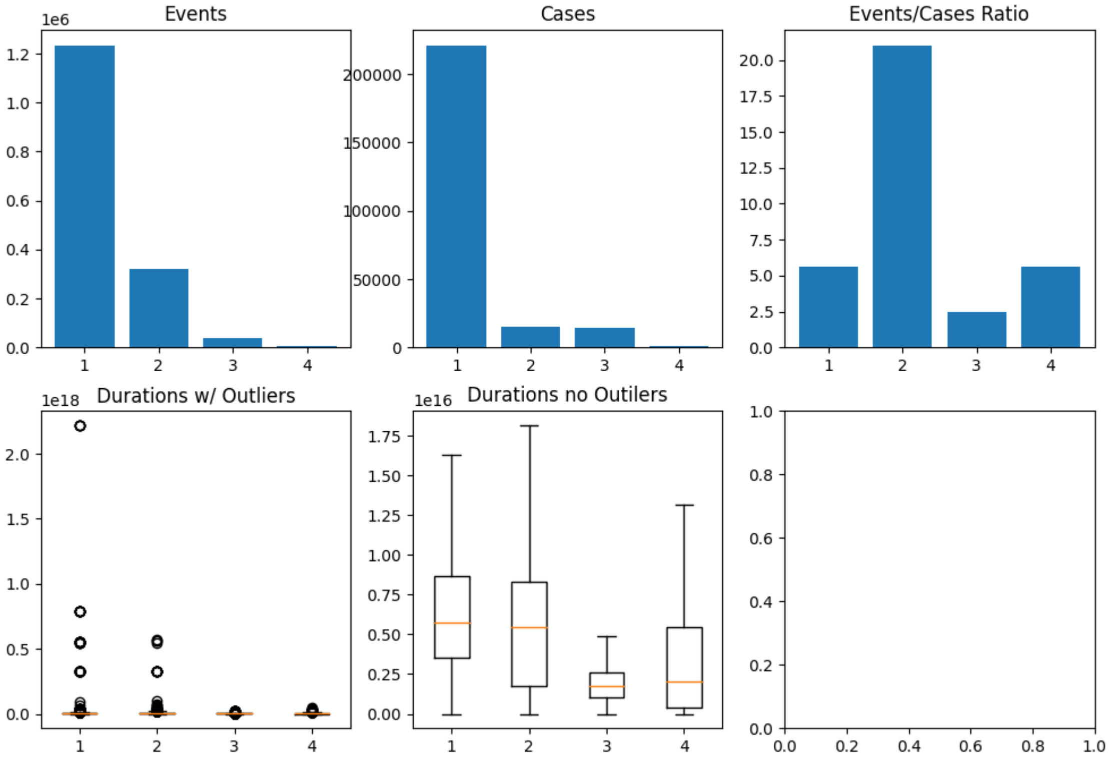


Figure 4: Comparing count and duration of Item Category

1.2 Variant

From the raw event log, 11,973 unique variants and the top 10 variants cover more than 60% of all the cases. The median duration is of 71 days. Process Variant Analysis is an important technique to analyze event logs of the processes to understand the differences between process variants to improve a business process.

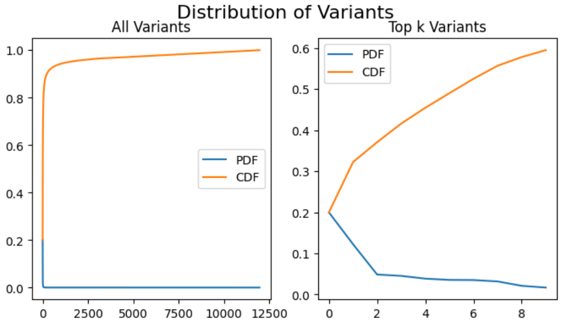


Figure 5: Distribution of Variants

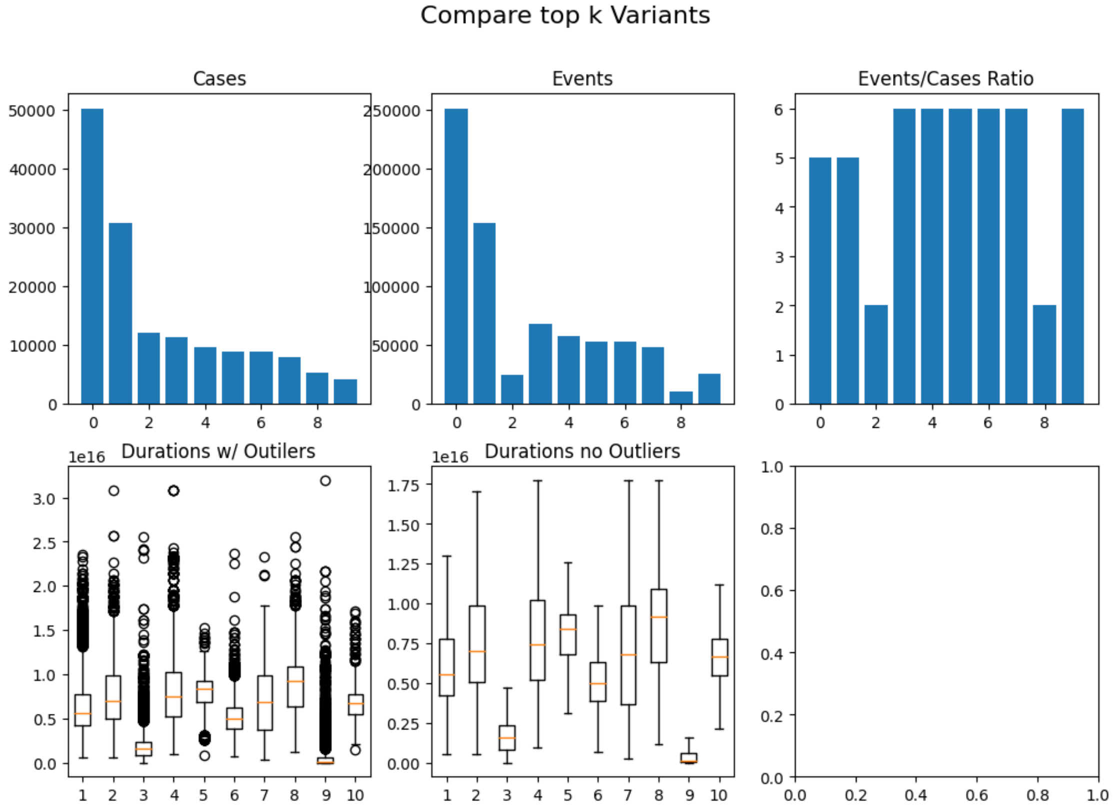


Figure 6: Comparing count and duration of top 10 Variants

1.3 Filter

An important step before the Process Discovery and further analysis are to filter noise. The aim of this process is to review data quality issues and choose carefully how to handle them.

- Time: there are few cases with events dating back to outside the scope of the event log (Max 25670 days) or to a future date. As most of the cases show regular behavior, this is only considered for the performance analysis.
- Incomplete, Cancelled, Open: analyzing the start and end activities, there are

cases that don't end with ”clear invoice” and many ends with ”Cancel Goods Receipt, Cancel Invoice Receipt, Cancel Subsequent Invoice, Delete Purchase Order Item” that means canceled. This leads to a total of Complete: 181,328 (72%), Incomplete: 70,406 (27.9%), Canceled: 11,198 (4.4%), Open: 59,208 (23.5%)

2 Process Discovery

Process Discovery aims to construct process models based on event logs, and here it is focused on discovery control-flow. The discovered process models should capture the behavior recorded in the event log considering the trade-off between desired and observed behavior. In this challenge has been roughly followed these steps for each of the different 4 Item Categories:

- Filter by complete, incomplete, open, canceled cases
- Filter by Document Type
- Filter by frequent variants
- using inductive miner to generate the BPMN process diagram, Process Tree, and Petri Net

2.1 3-way match, invoice before GR

This is the category with the most cases of the log. It has around 77.7% of complete cases, and consequently, the log was filtered by only complete cases. The Document Type frequency table showed that 99.1% of documents are Standard PO and so only that category was considered. From the distribution of the variants, the top 10 cover 80% of the total cases, and these variants were used for the inductive miner.

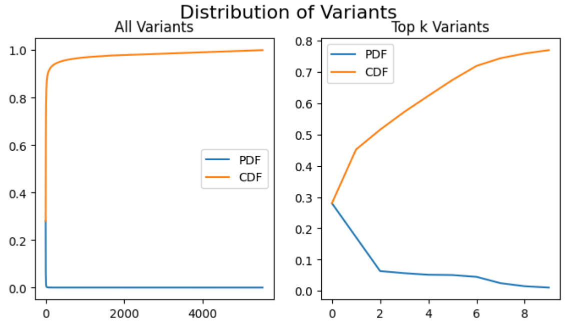


Figure 7: 3-way match, invoice before GR, distribution of variants

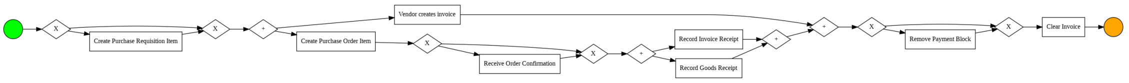


Figure 8: 3-way match, invoice before GR, BPMN inductive

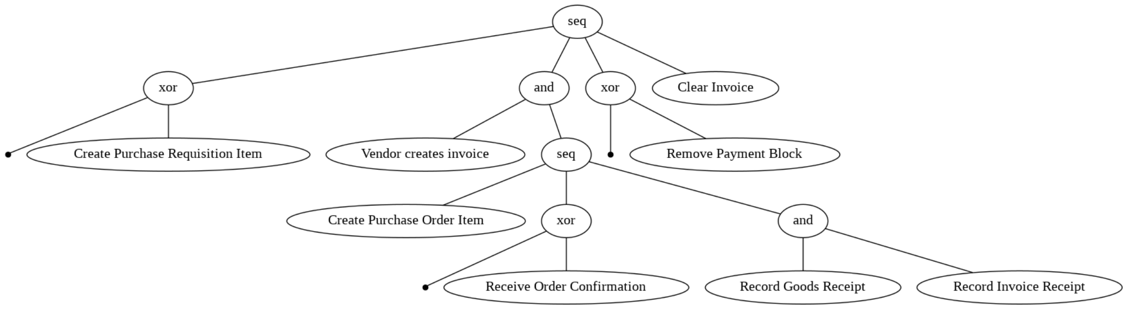


Figure 9: 3-way match, invoice before GR, process tree inductive

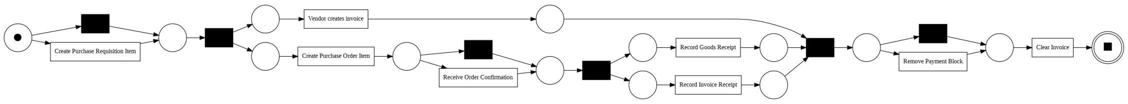


Figure 10: 3-way match, invoice before GR, petri net inductive

2.2 3-way match, invoice after GR

This category represent around the 6% of the total cases. Applying the filters: 60.9% complete, 39% of incomplete, 5% canceled, 33% open. This means that almost all the incomplete cases are open and the remaining are canceled. Using the complete cases and filtering by document type it is better to distinguish between EC purchase orders and all others because it is handled by a supplier relationship manager (SRM).

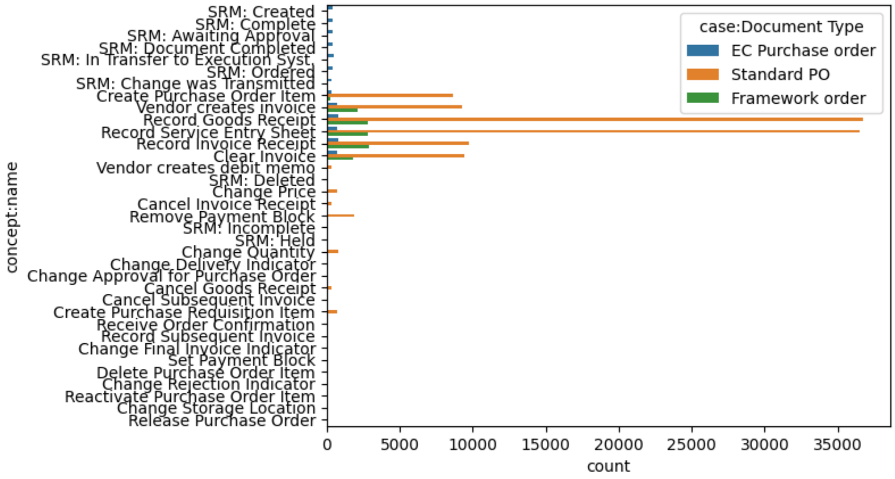


Figure 11: 3-way match, invoice after GR, compare Document Type and Activities

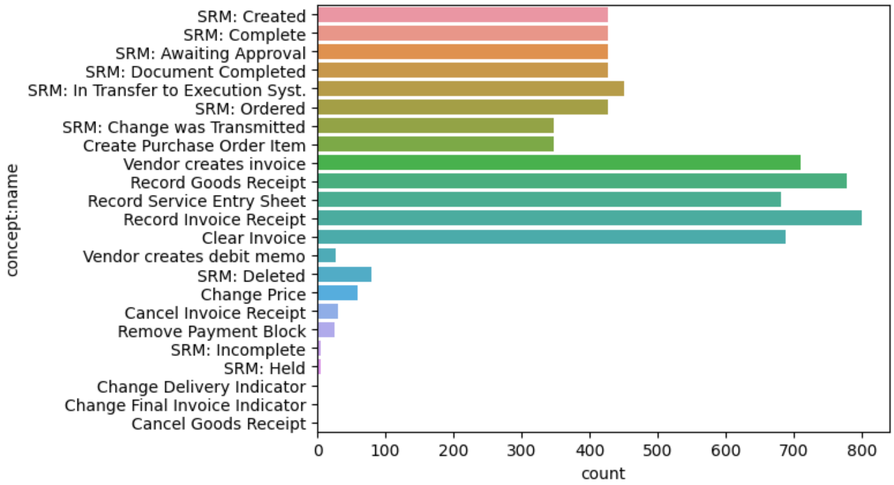


Figure 12: 3-way match, invoice after GR, EC activities

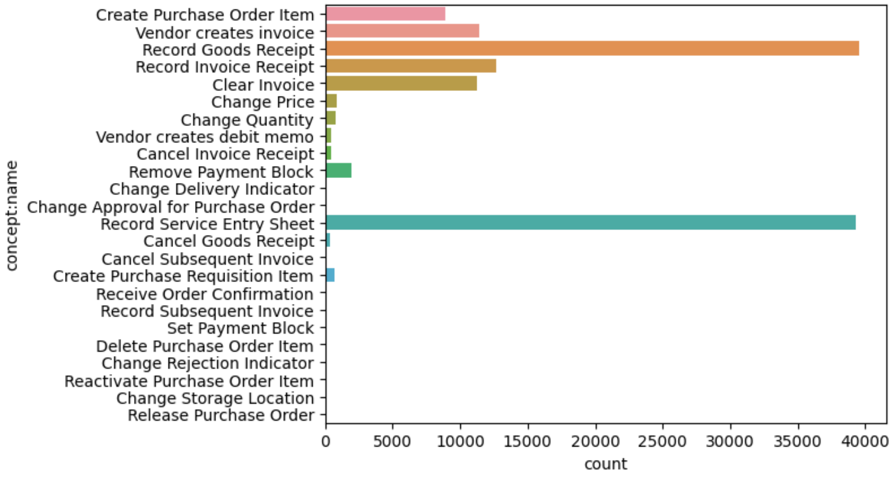


Figure 13: 3-way match, invoice after GR, not EC activities

For each different Document Type (EC Purchase order and all others) was analyzed the distribution of variants and used the inductive miner. From the BPMN, Process Tree and Petri Net, emerges that there's a huge difference in the process between these two Document Category, as the Activities plots suggested.

2.2.1 EC Purchase order

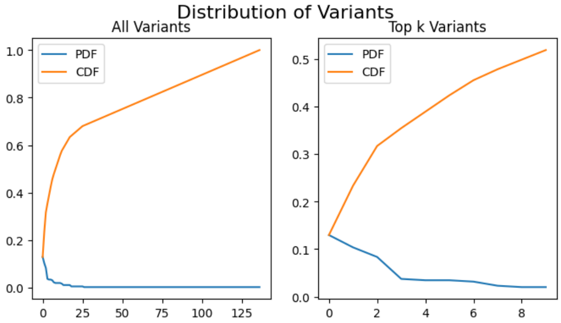


Figure 14: 3-way match, invoice after GR, EC distribution of variants

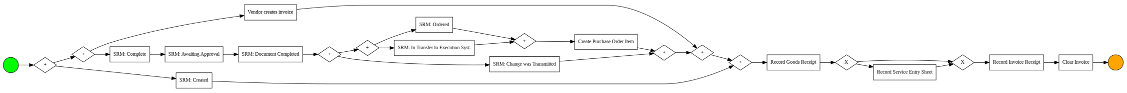


Figure 15: 3-way match, invoice after GR, EC BPMN

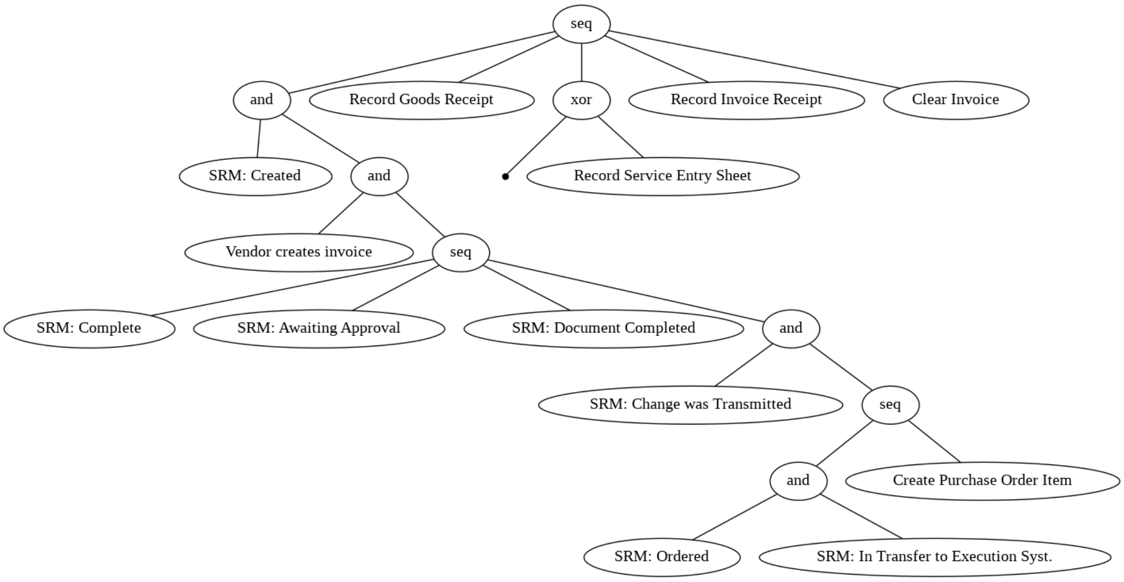


Figure 16: 3-way match, invoice after GR, EC process tree

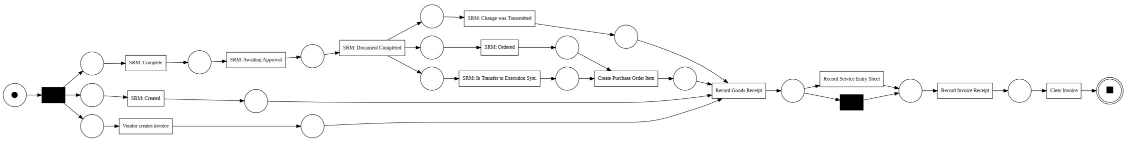


Figure 17: 3-way match, invoice after GR, EC, petri net

2.2.2 No EC Purchase order

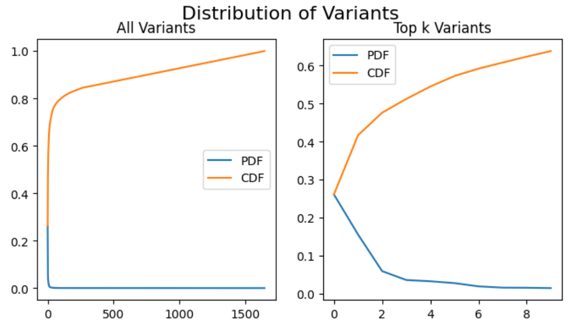


Figure 18: 3-way match, invoice after GR, NO EC distribution of variants

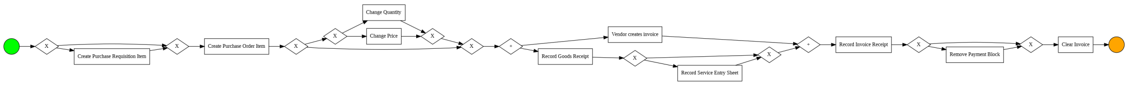


Figure 19: 3-way match, invoice after GR, NO EC BPMN

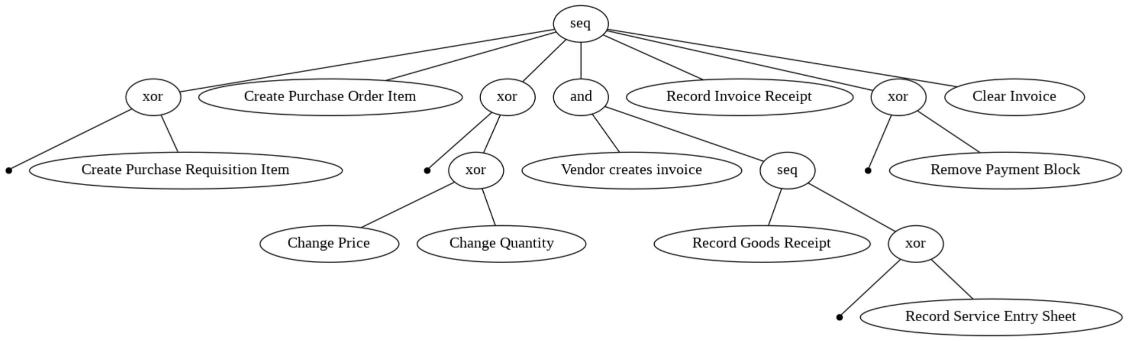


Figure 20: 3-way match, invoice after GR, NO EC process tree

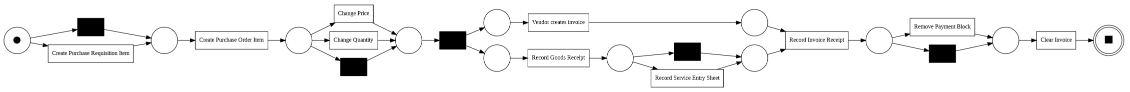


Figure 21: 3-way match, invoice after GR, NO EC petri net

2.3 2-way match

This category has the majority of data quality issues: the event ”Change Approval for Purchase Order” is occurring multiple times for each item. That’s because no goods receipts message is received, and because after the creation of the item it is added to its trace resulting in several ”Change Approval for Purchase Order” as both start activities and end activities.

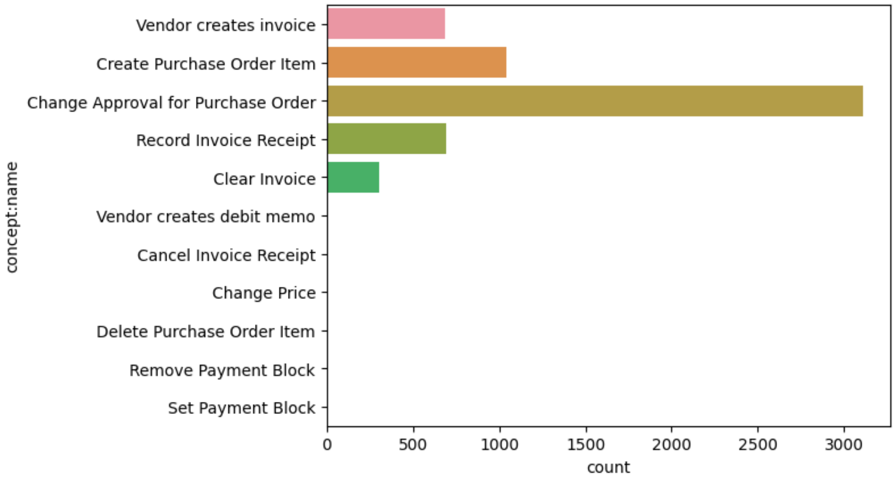


Figure 22: 2-way match, count of activities

To solve this issue, filter out this event type out from the traces as it indicates wrong sequences, the incomplete cases and select the top variants.

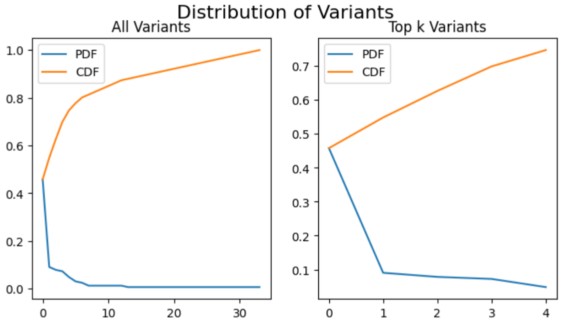


Figure 23: 2-way match, distribution of variants

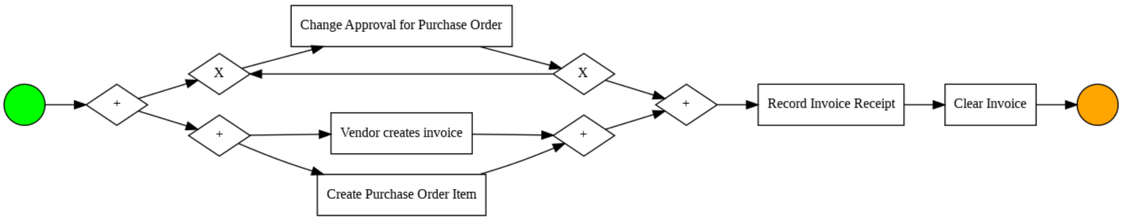


Figure 24: 2-way match, BPMN

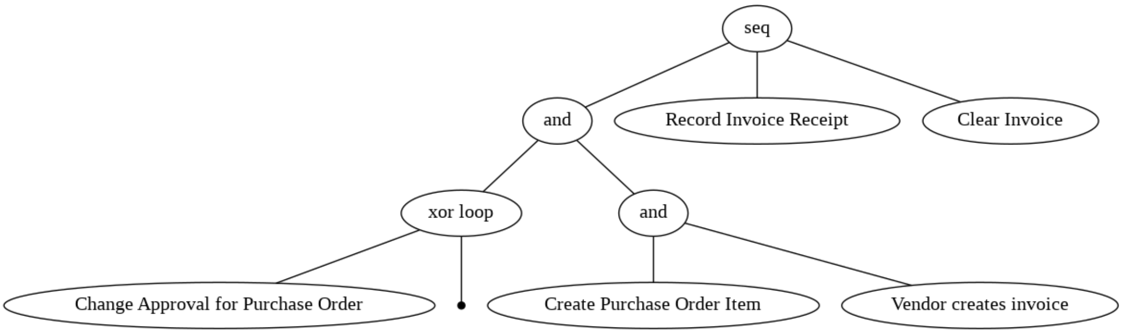


Figure 25: 2-way match, process tree

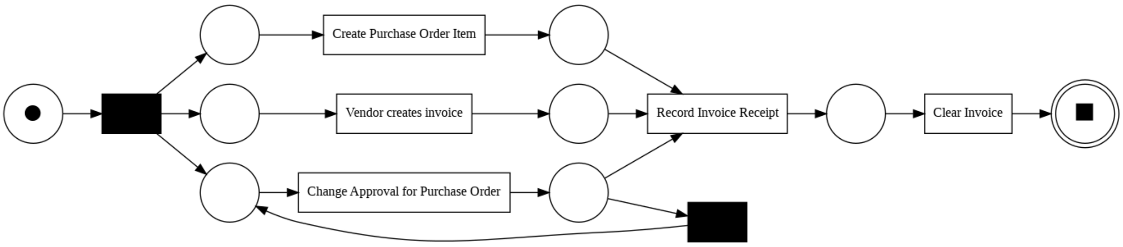

Create Purchase Order Item
Vendor creates invoice
Change Approval for Purchase Order
Record Invoice Receipt
Clear Invoice

Figure 26: 2-way match, petri net

2.4 Consignment

The consignment process consists in receiving a good in the warehouse and paid after usage, and the product usage is not part of the process log. In fact, this category doesn't have the "Clear Invoice" and activity. This results in a 100% of incomplete cases.

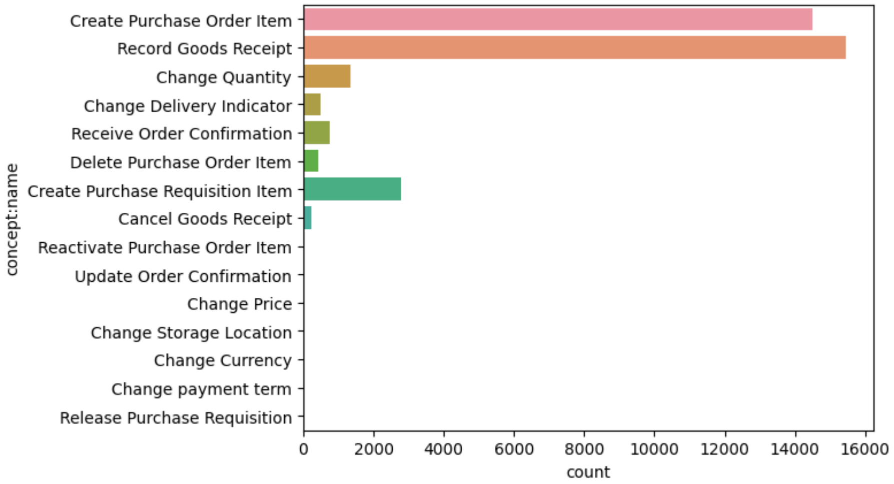


Figure 27: Consignment, count of activities

To use this category process, we filter out the complete cases and select the top variants.

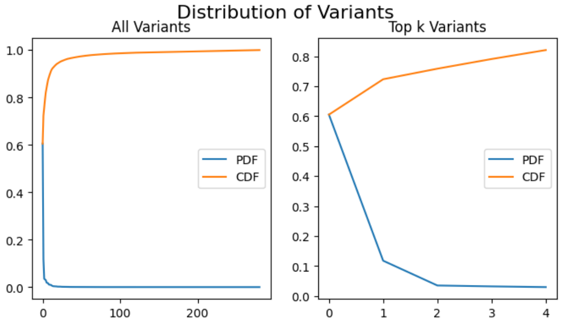


Figure 28: Consignment, distribution of variants

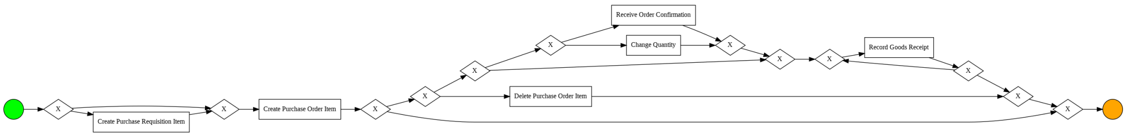

Create Purchase Requisition Item
Create Purchase Order Item
Receive Order Confirmation
Change Quantity
Delete Purchase Order Item
Record Goods Receipt
X

Figure 29: Consignment, BPMN

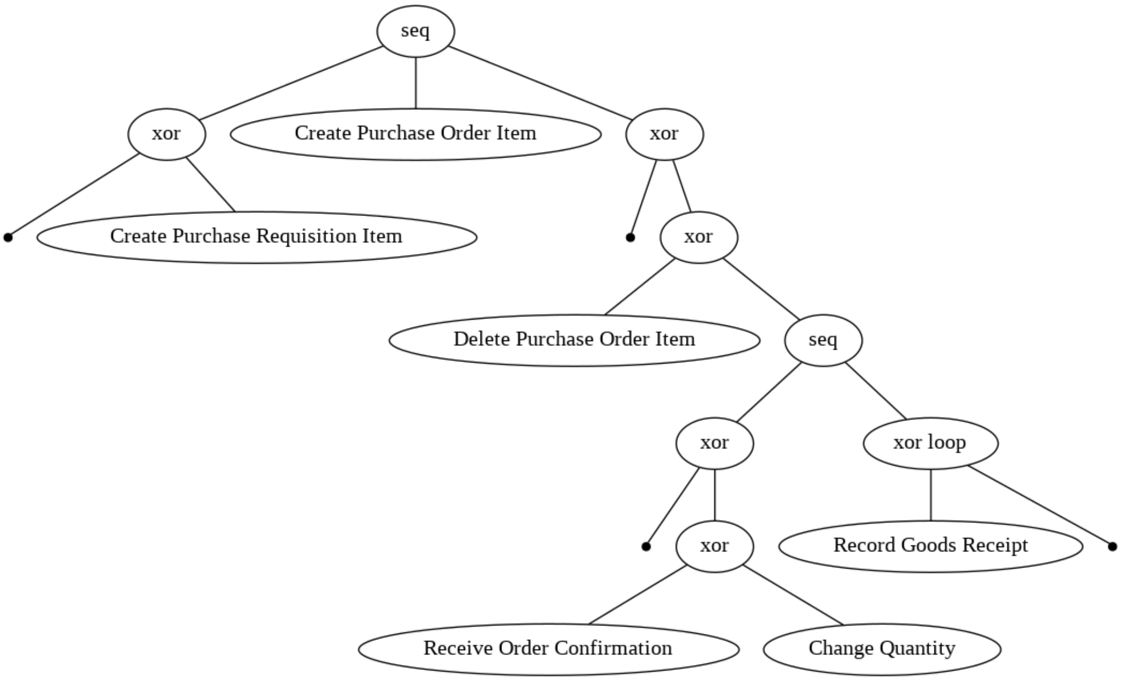

seq
xor
Create Purchase Order Item
xor
Create Purchase Requisition Item
xor
Delete Purchase Order Item
seq
xor
xor loop
xor
Record Goods Receipt
Receive Order Confirmation
Change Quantity

Figure 30: Consignment, process tree

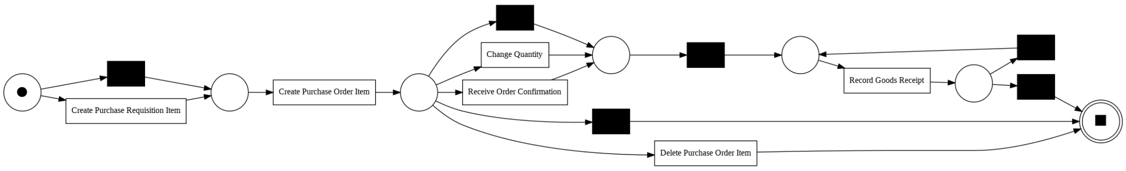

Create Purchase Requisition Item
Create Purchase Order Item
Change Quantity
Receive Order Confirmation
Record Goods Receipt
Delete Purchase Order Item

Figure 31: Consignment, petri net

3 Conformance Checking

Conformance checking is a techniques to compare a process model with an event log of the same process. The goal is to check if the event log conforms to the model, and, vice versa. Two fundamental techniques are used: token-based replay and alignments. These two are used with the petri net of the previous chapter, for each Item Category.

	# traces	% Fitness TBR	% Fitness Alignemnts
Item Category			
3-way match, before	171,191	92	82
3-way match, after with EC	347	58	56
3-way match, after without EC	8911	72	70
2-way matching	166	91	90
Consignment	14,498	96	93

Table 1: Compare Fitness between Item Category

These two techniques lead to the following results: for each Item Category the fitness is of more than 90% except for the "3-way match, invoice after GR".

4 Process Performance

4.1 Throughput Time

I define Throughput Time as the amount of time required to handle single invoice with reference to the times between three main steps: Record Goods Receipt, Record Invoice Receipt and Clear Invoice. I performed my throughput analysis for each of all the categories, subdivided by with and without EC purchase order, except for the Consignment category because it doesn't contain any invoice information. This process consisted in several steps: filtering by item category, filtering by document type, calculating the duration of the Record Goods Receipt to Record Invoice, and then the duration of Record Invoice Receipt to Clear Invoice

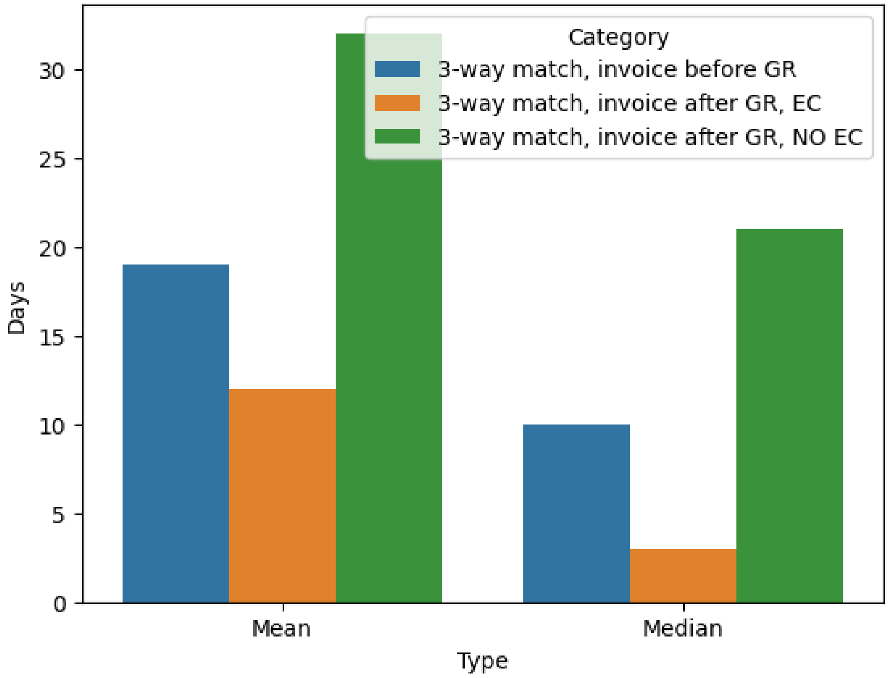


Figure 32: Record Goods Receipt -> Record Invoice Receipt performance

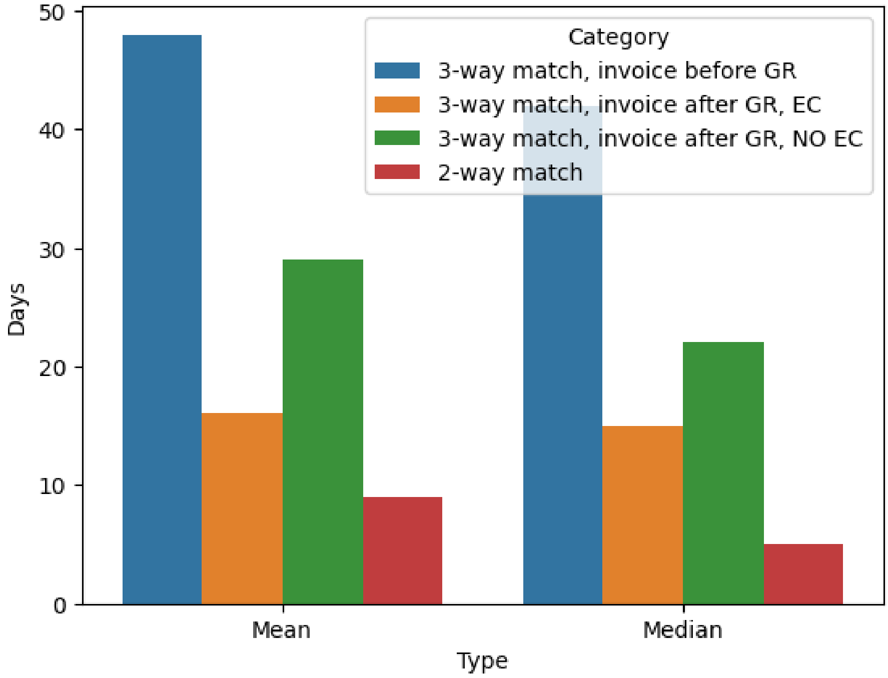


Figure 33: Record Invoice Receipt -> Clear Invoice performance

The results of Record Goods Receipt -> Record Invoice Receipt throughput time is a median of 10 days for 3-way match, invoice before GR, 3 days for 3-way match, invoice after GR, EC, 21 days for 3-way match, invoice after GR, NO EC. On the other hand, The results of Record Invoice Receipt -> Clear Invoice throughput time is a median of 42 days for 3-way match, invoice before GR, 15 days for 3-way match, invoice after GR, EC, 22 days for 3-way match, invoice after GR, NO EC, and finally 5 days for 2-way match.

4.2 Rework

Rework as execution of a change activity, like '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'. The analysis has been hold: count by activity to understand which activities are most relevant for rework, and % by category to understand which category is most relevant for rework.

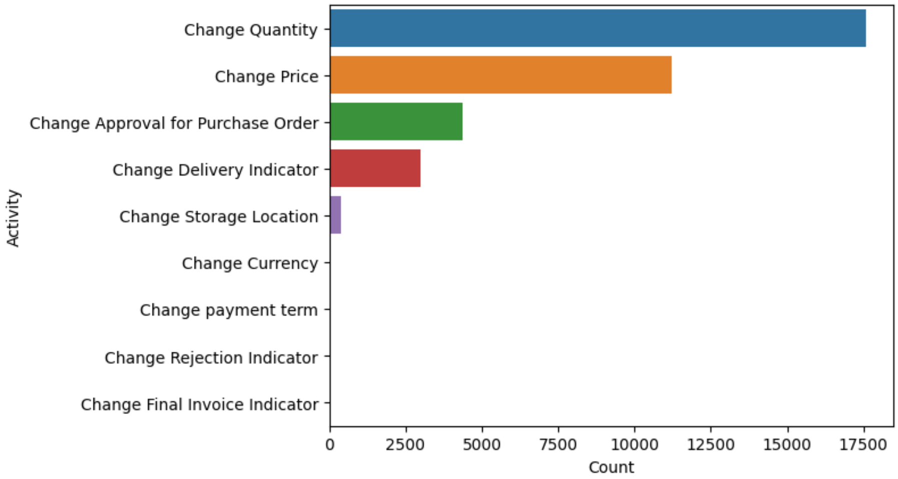


Figure 34: Count of rework cases per activity

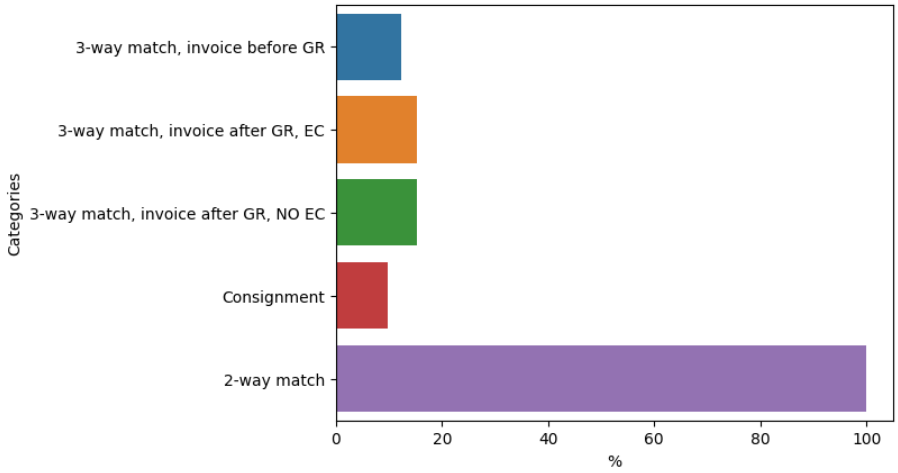


Figure 35: % of rework cases per category

The results are that the most rework is for Change Quantity (17,590 cases) and Change Price (11,224 cases). Also, as we discussed previously, it is not surprise that the 100% of the 2-way match contains Change activities.