BPM Challenge 2020

Vadim Sokolov

Business Information Systems
Prof. Paolo Ceravolo
a.a. 2021 - 2022

Case study description

At Eindhoven University of Technology staff members travel to conferences or to other universities for project meetings and to meet up with colleagues in the field. The university pays for travel expenses.

There are two types of trips, domestic and international. For domestic travel no prior permission is required. An employee can undertake these trips and ask for reimbursement of the costs afterwards.

For international trips, permission is needed from the supervisor. This permission is obtained by filing a travel-permit and this travel permit should be approved before making any arrangements.

To get the costs for a trip reimbursed, a claim is filed. This can be done as soon as costs are actually paid, or within two months after the trip.

The data presented in this challenge

The data is collected from the reimbursement process at TU/e. The files contain data from 2017 (only two departments) and 2018 the full TU/e.

The data is split in datasets presented in the following table:

Table 1 - Data representation

Number	Name	Cases	Events	Size, Mb
1	Domestic declarations	10500	56437	20.5
2	International declarations	6449	72151	29.2
3	Prepaid travel costs	2099	18246	7.8
4	Requests for payment	6886	36796	15.2
5	Travel permits	7065	86581	33.2

Goals

The goal of the research is to analyse the data provided in order to extract useful information. The organisation can use the results of the analysis to reach its goals, improve performance of the processes, reduce bottlenecks and take into account the information about data anomalies.

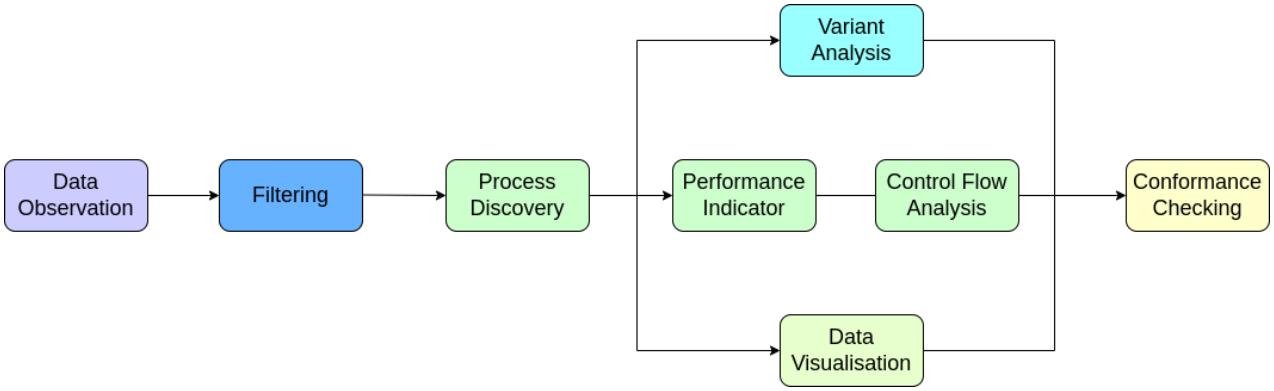


Figure 1 - Goal Diagram

Questions of the challenge

- What is the throughput of a travel declaration from submission (or closing) to paying?
- Is there a difference in throughput between national and international trips?
- Are there differences between clusters of declarations, for example between cost centres/departments/projects etc.?
- What is the throughput in each of the process steps, i.e. the submission, judgement by various responsible roles and payment?
- Where are the bottlenecks in the process of a travel declaration?
- Where are the bottlenecks in the process of a travel permit (note that there can be multiple requests for payment and declarations per permit)?
- How many travel declarations get rejected in the various processing steps and how many are never approved?

Detailed questions

- How many travel declarations are booked on projects?
- How many corrections have been made for declarations?
- Are there any double payments?
- Are there declarations that were not preceded properly by an approved travel permit? Or are there even declarations for which no permit exists?
- How many travel declarations are submitted by the traveller and how many by a mandated person?

- How many travel declarations are first rejected because they are submitted more than 2 months after the end of a trip and are then re-submitted?
- Is this different between departments?
- How many travel declarations are not approved by budget holders in time (7 days) and are then automatically rerouted to supervisors?
- Next to travel declarations, there are also requests for payments. These are specific for non-TU/e employees. Are there any TU/e employees that submitted a request for payment instead of a travel declaration?

Knowledge uplift trail

The first step of the analysis is filtering. It allows us to get rid of noisy data. One of the most important parts of analysis is the search for bottlenecks.

Specification:
Employees, departments represented by administration, budget owners and supervisors are actors in this process. The values in the model are a travel expenses cost and processing time of the declaration.

Two datasets: domestic and international declarations contain the most important information. Therefore a major part of the analysis is held on these two files.

Knowledge uplift was performed using both pm4py library and Disco program. According to figure 1, both these approaches allow us to use data filtering, use process discovery and provide data visualisation, but the most important part of filtering and variant analysis is performed via the pm4py library, which allows us to use more specific methods.
Table 1 - Knowledge uplift trail

Steps	Input	Analytics	Model	Output
Step 1	Raw log	extraction	Descriptive	Event Log
Step 2	Event log	Variant analysis	Descriptive	Variant
Step 3	Data frame / Event log	Filtering	Descriptive	Filtered log/ DF
Step 4	Filtered log	Process Discovery	Prescriptive	Model
Step 5	Model	Conformance check	Prescriptive	Precision and other metrics

Project results

I. Analysis of Domestic Declarations

1. Filtering

First, Domestic Declarations were taken into account for analysis. The first step before extracting knowledge is preprocessing that can be made by filtering. Filtering process was implemented using Pm4Py filtering methods. As we can see from table 1, there are 10500 cases in total for domestic declarations. Filtering by start activity allows us to extract events that start with `'Declaration SUBMITTED by EMPLOYEE'` event, which is a necessary requirement for starting the process of declaration submission. It corresponds to reducing the number of cases to 10365. The next step is to filter the data on the number of activities. We assume that declarations can't consist of less than 2 events, that is, "Declaration submitted by Employee", "Declaration approved / rejected" (by any administrative instance). Also we set the minimum requirement for a total time for each activity, namely, 1 day. It brings us to a reduction in the number of cases to 10349 cases. Assuming that we are interested in only successful declarations, which end with "Payment Handled" event, filtering over end activities was performed, reducing total number of activities to 10042. Successful declarations give us the most useful information about the system. It is noteworthy that all the filtered activities last at least 25 hours. As one can see, filtering reduced the total number of cases to be analysed by 4.3%.

2. Variant analysis

Number of variants for domestic declarations is equal to 63. Basic results of variant analysis are presented in table 2. As one can see, some variants represent cases which were never approved. Overall success rate is high, the quality of the dataset is relatively good.

Table 2 - Variant analysis Domestic declarations

Property	Value	Description
Cases number	10500	Total number of cases
Handled cases	10042	Handled case as a final event
Never approved	230	Cases that were not approved
Handled ratio	0.968	Ratio of successful cases
Rejected	1301	Number of cases rejected at least once

Resubmitted	1166	Number of cases resubmitted at least once
Average duration	7 days 14 hours	Average duration of the case from submission to payment

To make performance measurements, it is necessary to define a performance indicator. First, a key strategy for an employee in a domestic declaration is to submit it, meeting all the requirements of administrative institutions in order to get the approval as fast as possible. Therefore, it is necessary to avoid resubmissions of declarations. Successful case is a case in which a declaration is submitted by an employee, finally approved by the supervisor and payment was handled.

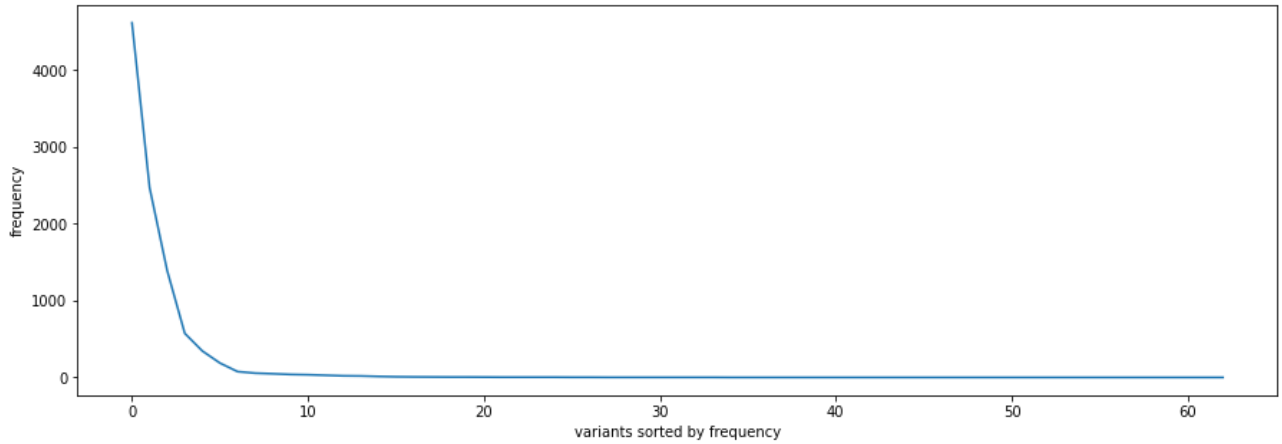


Figure 2 - Variant distribution

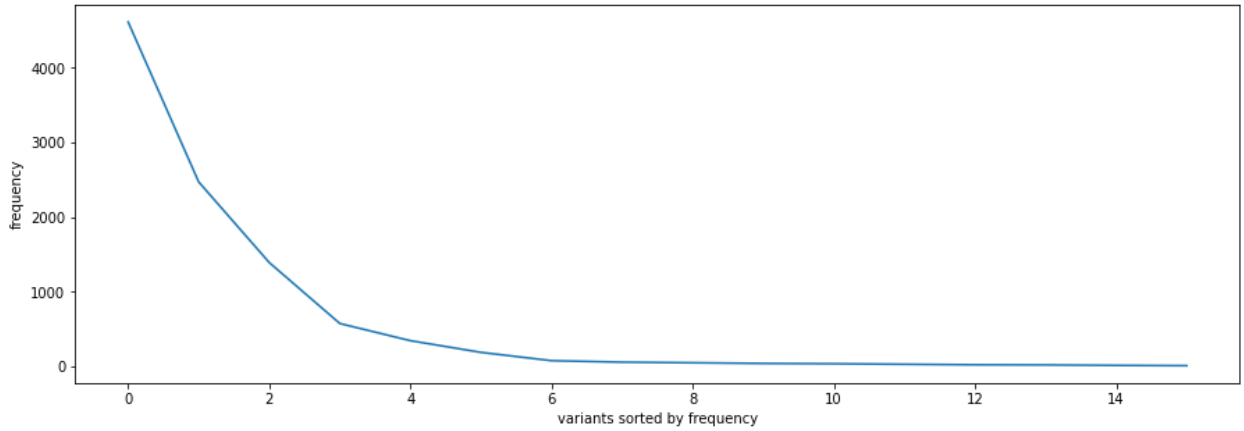


Figure 3 - Variant distribution, most frequent

More than half of the variants have frequency rate less than 10, these cases can be rarely seen. As one can see, this is an example of Pareto distribution.

Table 3 - Variant Distribution

Parameter	Value
Mean	621
Median	53
Mode	20

3. Process discovery

In order to analyse processes in the system and create the most representative data model, it is necessary to apply different process discovery techniques.

1. Alpha Miner

Alpha miner algorithm is applied taking into consideration the frequency of events.

Figure 4 - Petri net for Alpha Miner

2. Inductive Miner

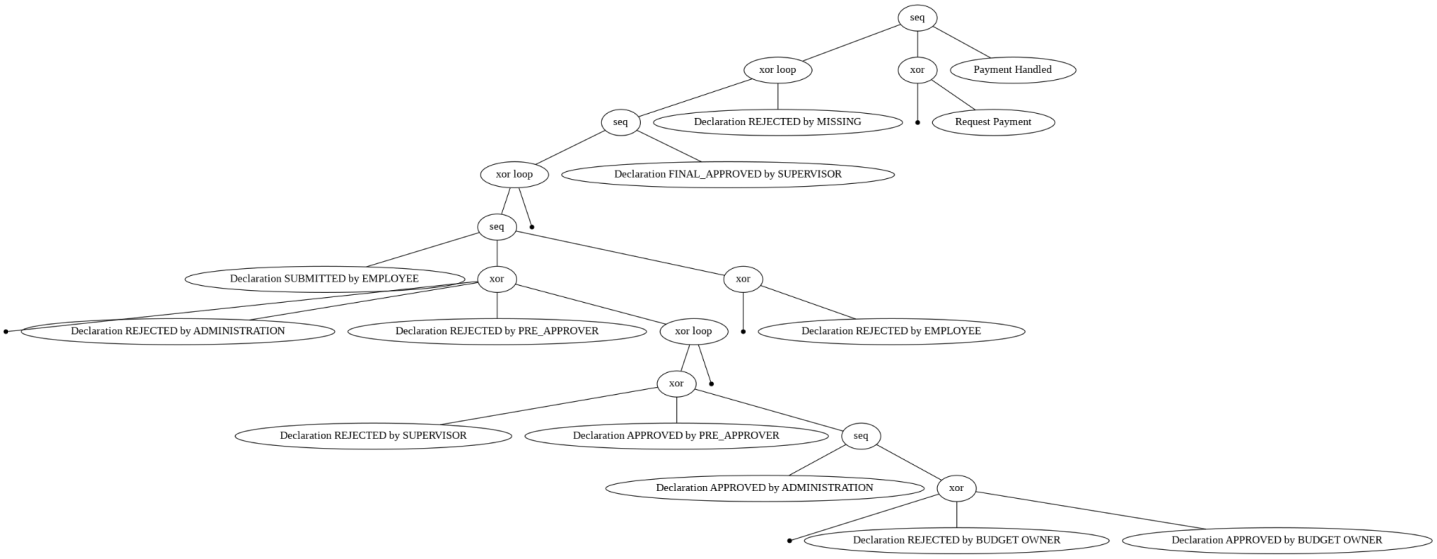


Figure 5 - Process tree Inductive Miner

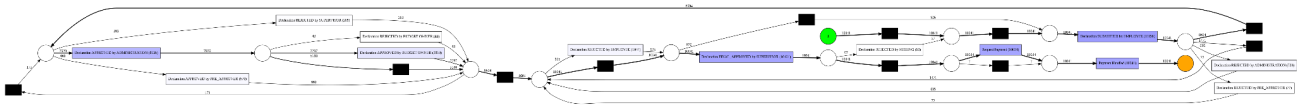


Figure 6 - Petri net Inductive Miner

3. Heuristic Miner

In order to find the most successful case, different values of dependency threshold were observed.

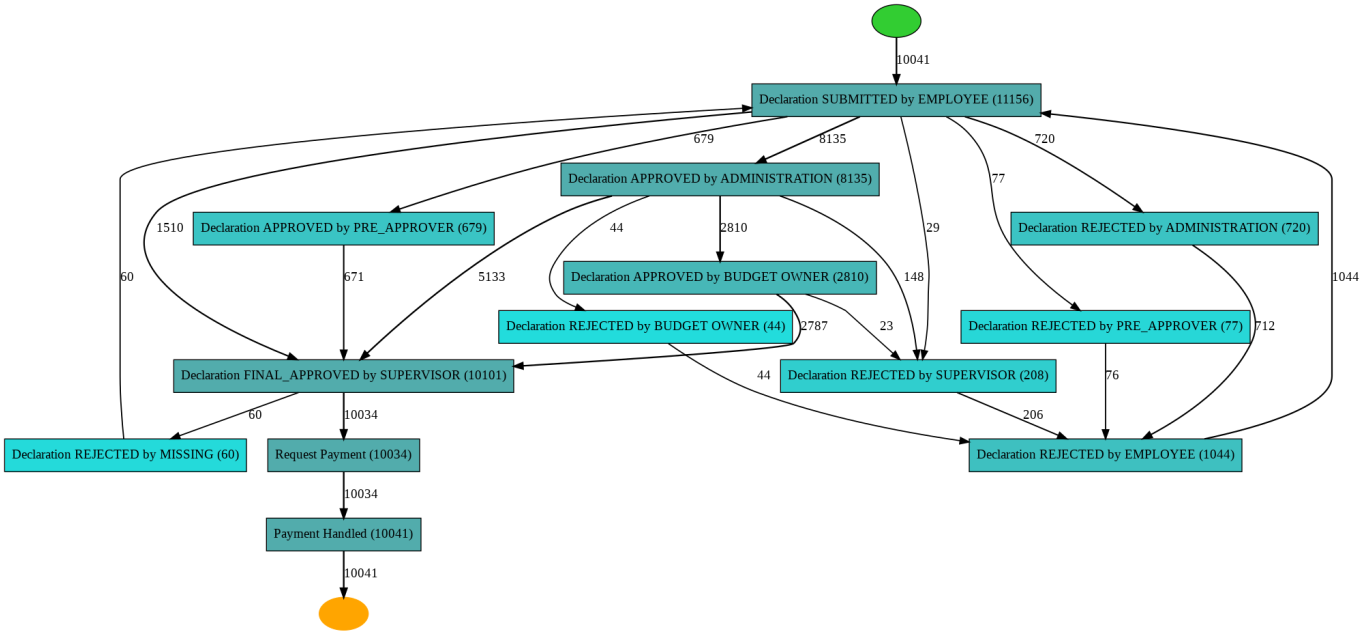


Figure 7 - Process tree Heuristic Miner, dependency threshold: 0.9

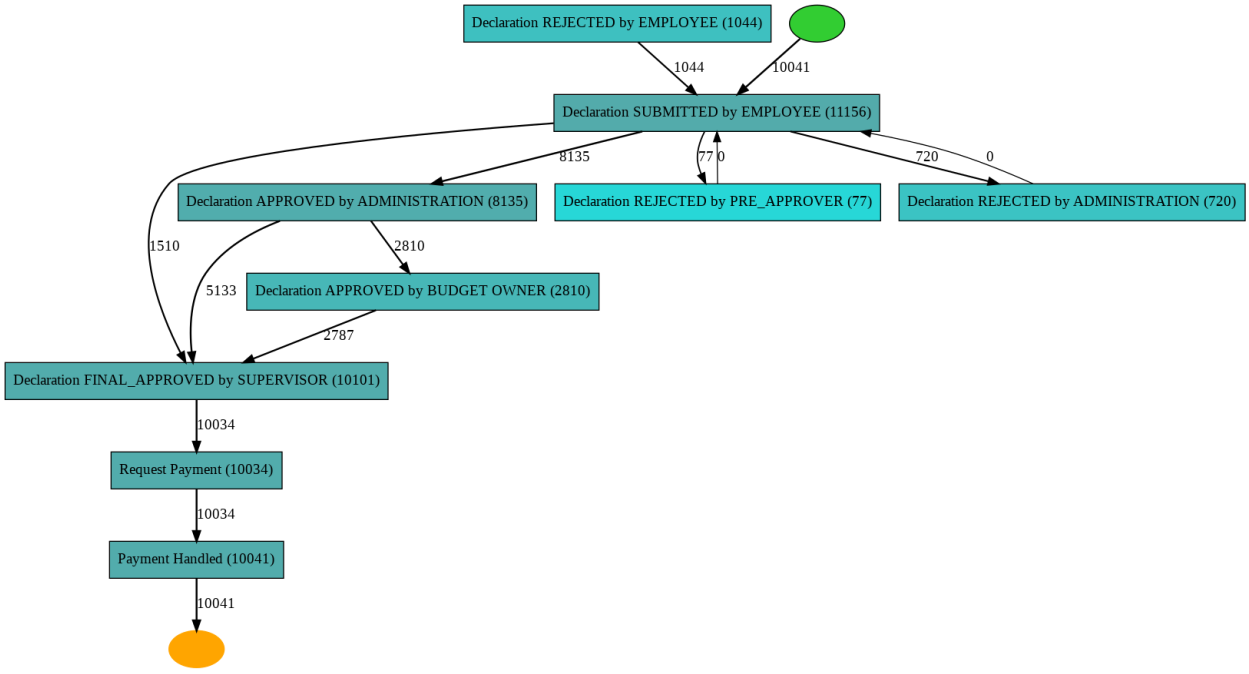


Figure 8 - Process tree Heuristic Miner, dependency threshold: 0.999

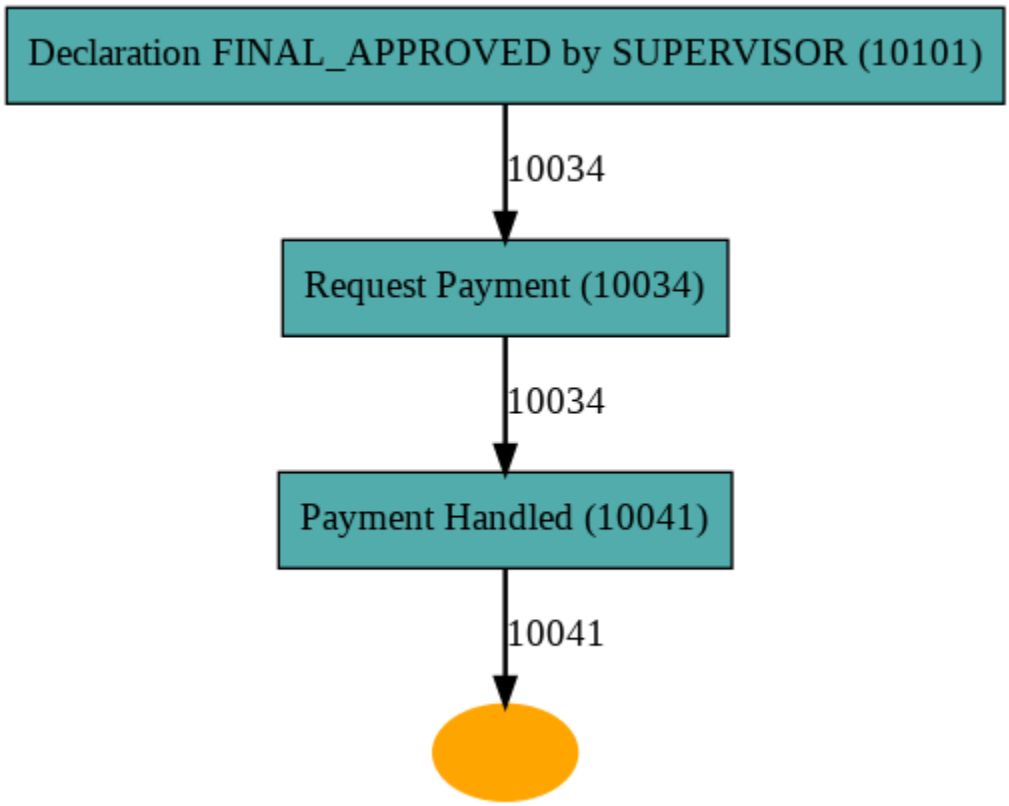


Figure 9 - Process tree Heuristic Miner, dependency threshold: 0.9999

As one can see, the majority of cases fit in the 90% threshold, which means that there are only slight deviations from the successful case.

Figure 10 - Heuristic Miner, Petri Net

4. Directly-followed graph that shows good successful case representation contains 90% of all the cases.

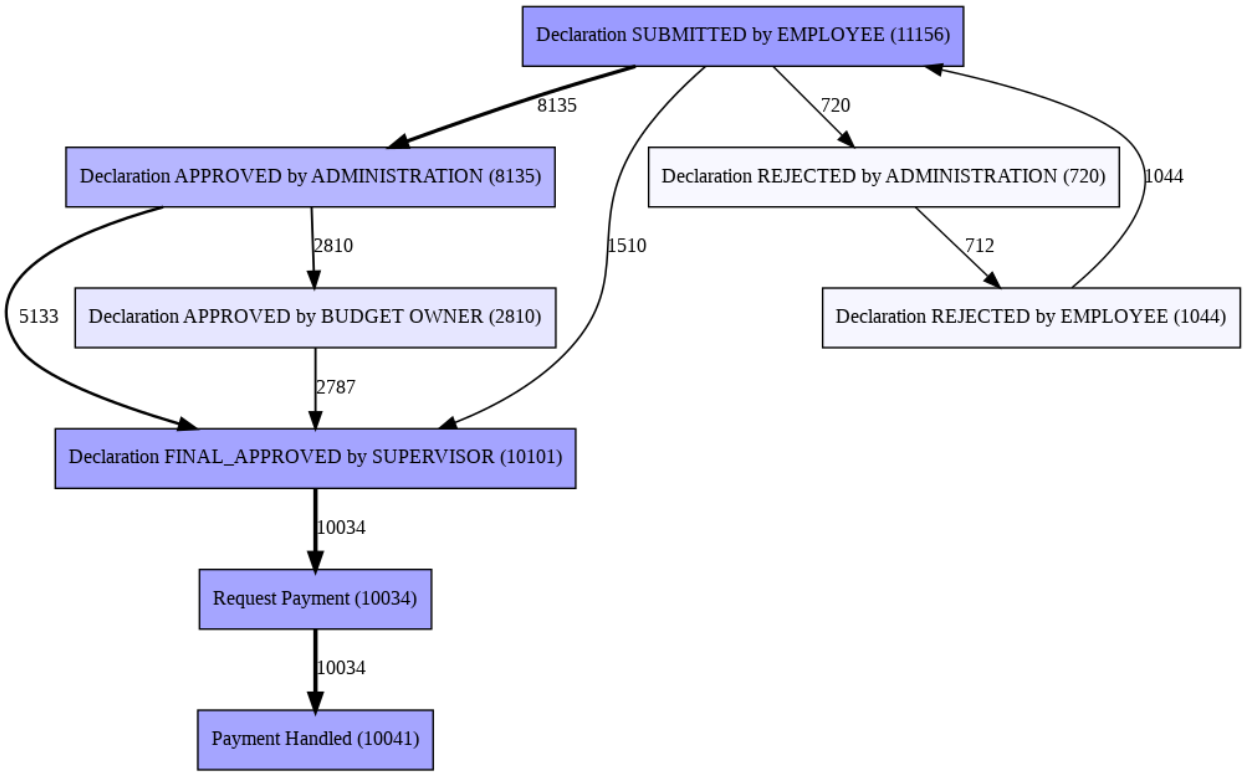


Figure 11 - Directly-followed graph

4. Conformance Analysis

In order to understand the accuracy of developed models, it is necessary to provide a conformance analysis - validation of predictive models.

Table 4 -- Conformance Analysis

Algorithm	Fitness	Precision	Generalisation	Simplicity
Alpha miner	0.96	0.95	0.97	0.81
Inductive miner	1.0	0.62	0.94	0.62
Heuristic miner	0.96	0.95	0.97	0.82

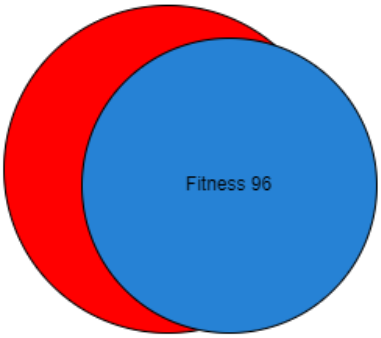


Figure - Epistemic diagram heuristic miner

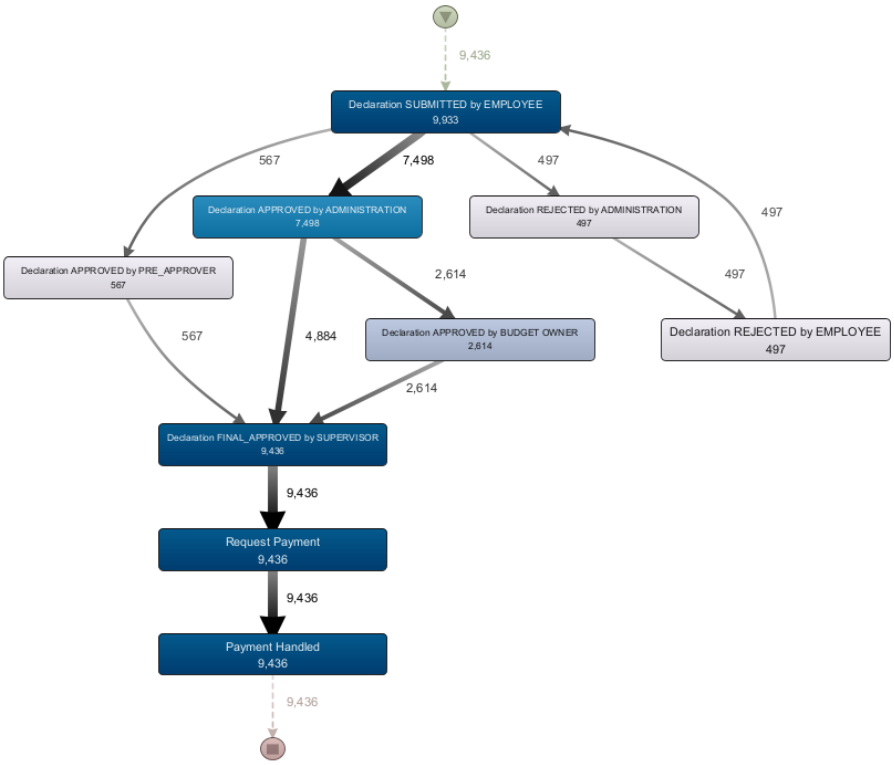


Figure 12 - Domestic Declarations Disco

The throughput of a travel declaration from submission to paying handled is equal to 7.3 days on average for domestic declarations. On average it takes 2.3 days between submission of the declaration and approval by any of the organisational entities. Intermediate approvals as well as rejections and resubmissions take around 3 days on average. It takes an additional 1.5 days from the final approval by the supervisor to handle the payment. Among all the cases, only 567 were approved by Pre Approver. The other way - approval by administration and budget owner. 7498 cases follow this path before the final approval by the supervisor. 2614 cases were approved by the budget owner. When an employee is late with pre approval, declarations are sent to administration. This bottleneck could be avoided by submitting declarations on time. Only around 2.1% of all the declarations were rejected. Rejected by the administration: 1%, rejected by employees: 1%. But the last case scenario is used for resubmission of the declaration. 230 declarations were never approved. Total number of corrections made in declarations, that is equal to a number of resubmissions, is for 5.2% of declarations. 4884 domestic declarations are not approved by budget holders in time (7 days) and eventually they are rerouted to supervisors.

II. Analysis of International declarations

International declarations have different structure in comparison to domestic ones. In international declarations it is possible to specify start and end dates of trip and deal with all the payments later on. It brings an additional number of variants, additional parallelisation and loops.

1. Filtering

Since it is not necessary to start the activity by submitting the declaration or permit for a trip and huge variability in multiple cases, it would be incorrect to filter all the activities by starting events. The only property of a successful case is supposed to be a handled payment, assuming that all the trips can not be for free. Total number of activities was reduced from 6449 to 6187.

2. Variant Analysis

Table 5 - Variant analysis International declarations

Property	Value	Description
Cases number	6449	Total number of cases
Handled cases	6187	Handled case as a final event
Never approved	20	Cases that were not approved
Handled ratio	0.959	Ratio of successful cases
Rejected	1755	Number of cases rejected at least once
Resubmitted	7915	Number of cases resubmitted at least once
Final approval by supervisor	5961	Cases in which employee had a supervisor's approve
Permit submitted by Employee	5761	Permit submitted by Employee at least once
Average duration	56 days 15 hours	Average duration of the case from submission to payment

As far as we can see from basic variant analysis, it is more likely that an employee resubmits the declaration. To put it simply, it is more difficult to prepare the declaration that will satisfy the requirements for an international trip in comparison to a domestic one. Total number of variants is 645. Number of variants with frequency more than 10 is 71, dropping 9 times.

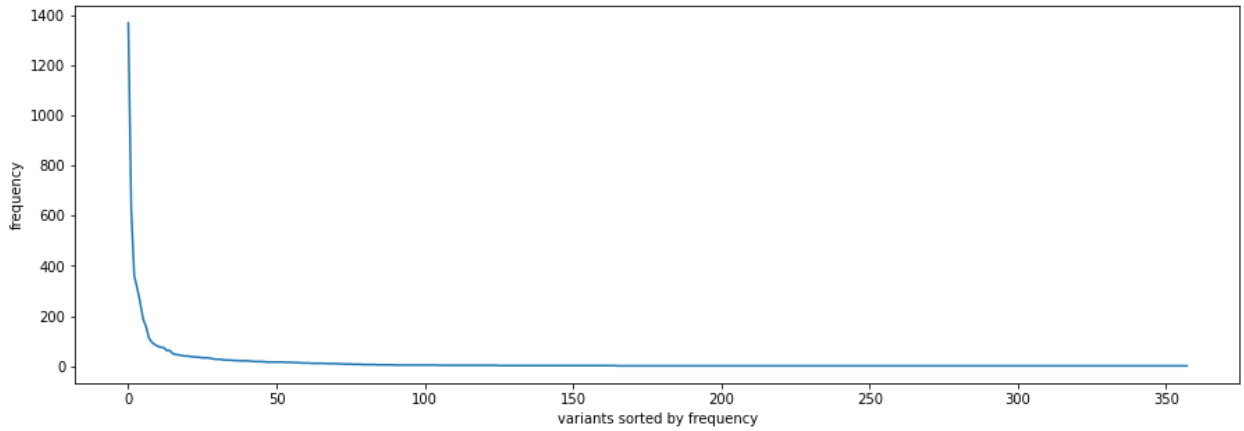


Figure 12 - Distribution of variants

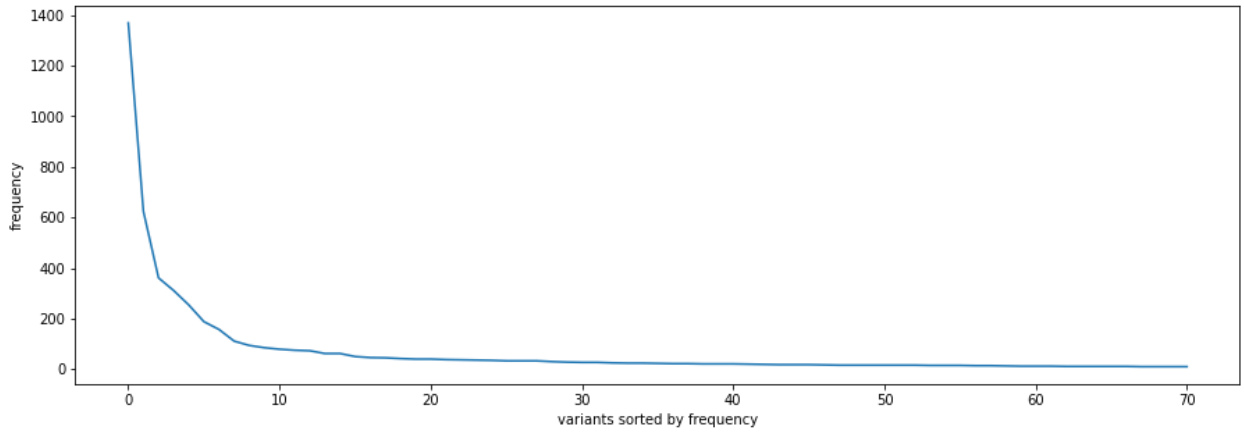


Figure 13 - Distribution of variants, most frequent cases

Table 6 - Variant distribution

Parameter	Value
Mean	72.54
Median	23
Mode	16

3. Process Discovery

1. Alpha Miner

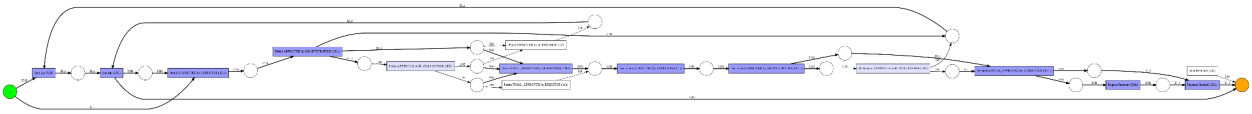


Figure 14 - Alpha Miner Petri net

Even though the number of variants was reduced up to 117 and all the rejections in variants were eliminated, the graphical representation of the Petri net for Alpha miner algorithm is quite complicated. The reason is that there are still many variants because of the presence of shuffles between different events that can be assumed as successful.

2. Inductive Miner

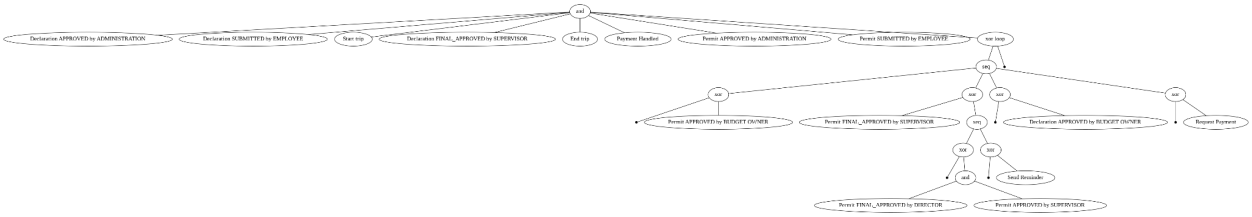


Figure 15 - Inductive Miner Tree

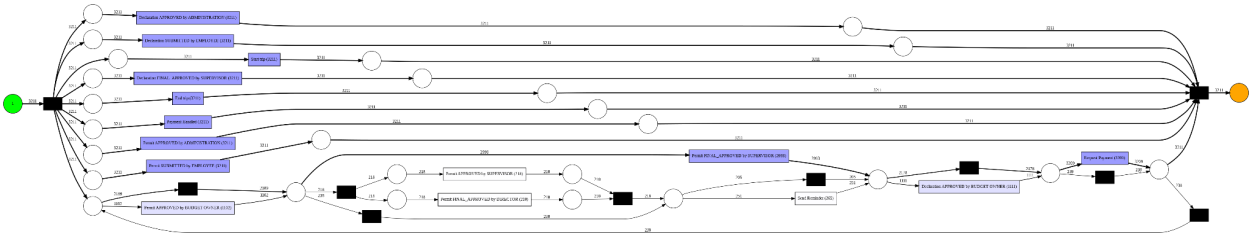


Figure 16 - Inductive Miner Petri net

3. Heuristic Miner

Heuristic miner represented in graph shows big variability in cases, but Petri net for it has less noise.

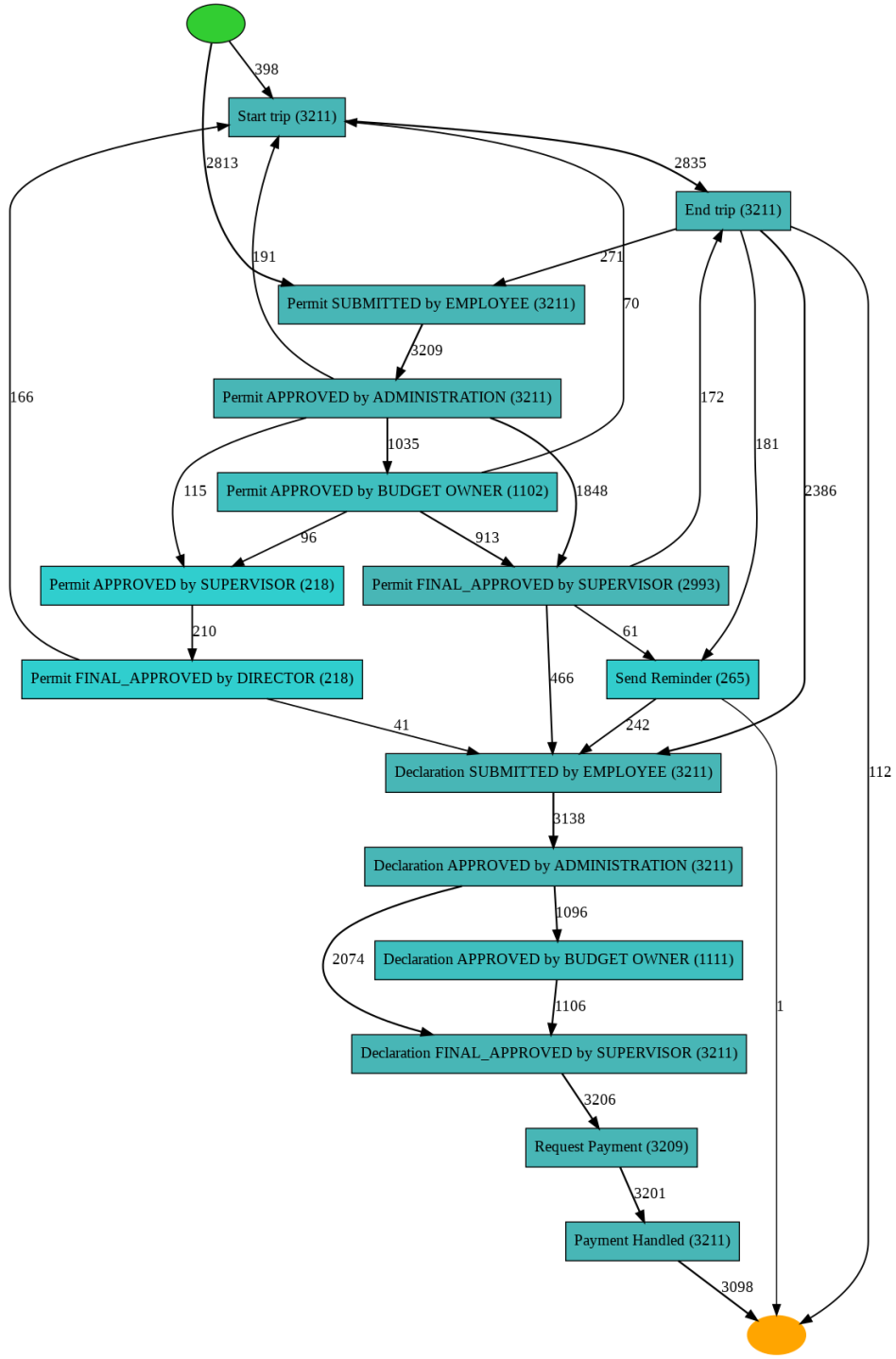


Figure 17 - Heuristic miner Graph

Figure 18 - Heuristic miner Petri net

4. Directly-followed graph

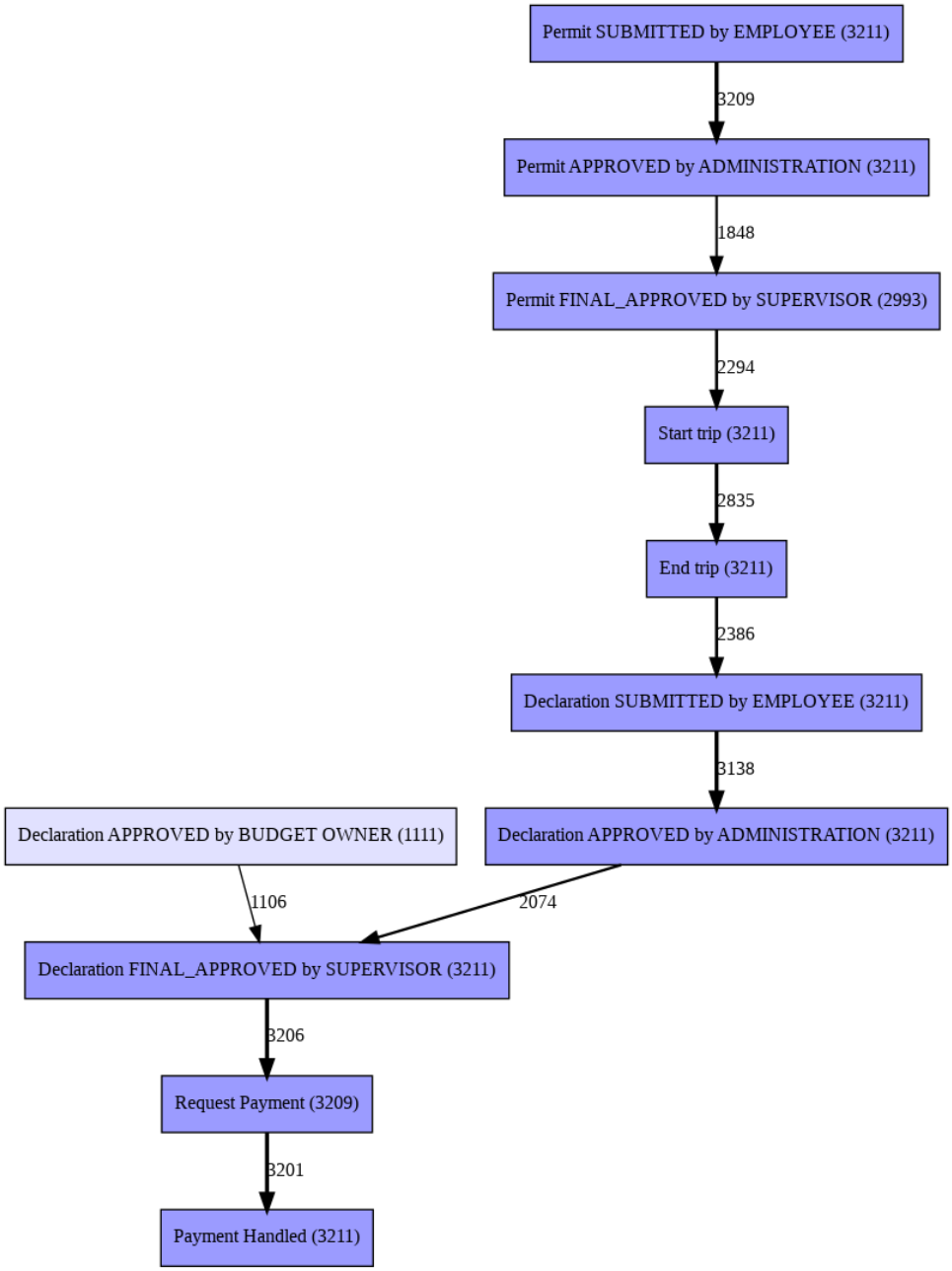


Figure 19 - Directly-followed graph

4. Conformance Analysis

Table 7 -- Conformance Analysis

Algorithm	Fitness	Precision	Generalisation	Simplicity
Alpha miner	0.75	0.78	0.97	0.63
Alpha miner 30% cases	0.93	0.95	0.96	0.86
Inductive miner	1.0	0.73	0.96	0.64
Heuristic miner	0.94	0.97	0.88	1.0

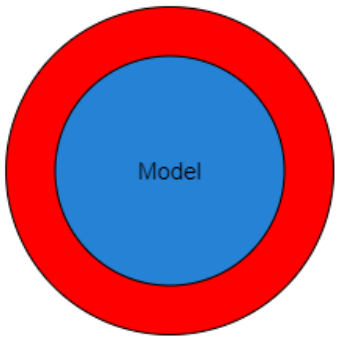


Figure - Epistemic diagram Inductive miner

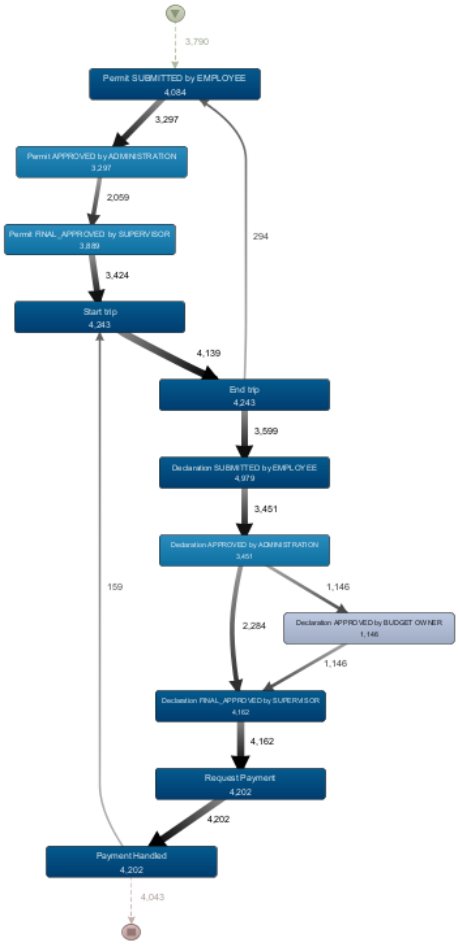
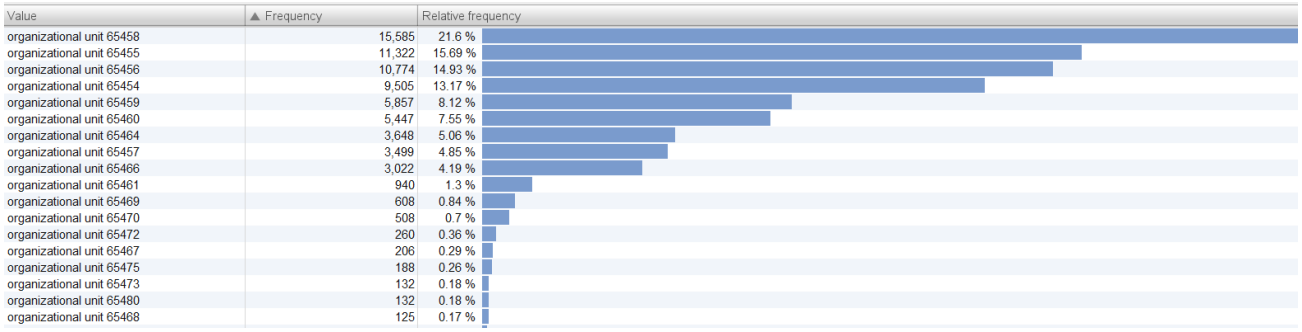


Figure 20 - Disco visualisation International declarations

The throughput of a travel declaration from submission to paying is 56 days for international declarations. International declarations can be divided into clusters by task number, projects, departments etc. 44537 cases, 61% of all declarations are processed by Task 427. 37% of them are processed by Unknown tasks, making less than 2% of all declarations being processed by other tasks. Distribution of tasks by permit project number is the following: 37% for unknown project and 4.93% for project 426. Other projects take less than 1% of all the declarations each. Distribution of tasks by permit organisational entity is presented on the following table.

Table 8 - Tasks by permit organisational entity



Value	Frequency	Relative frequency
organizational unit 65458	15,585	21.6 %
organizational unit 65455	11,322	15.69 %
organizational unit 65456	10,774	14.93 %
organizational unit 65454	9,505	13.17 %
organizational unit 65459	5,857	8.12 %
organizational unit 65460	5,447	7.55 %
organizational unit 65464	3,648	5.06 %
organizational unit 65457	3,499	4.85 %
organizational unit 65466	3,022	4.19 %
organizational unit 65461	940	1.3 %
organizational unit 65469	608	0.84 %
organizational unit 65470	508	0.7 %
organizational unit 65472	260	0.36 %
organizational unit 65467	206	0.29 %
organizational unit 65475	188	0.26 %
organizational unit 65473	132	0.18 %
organizational unit 65480	132	0.18 %
organizational unit 65468	125	0.17 %

We can see that organisational units 65458, 65455, 65456, 65454 together compose the majority of cases.
According to figure 20, 5037 declarations were approved by the administrator, 1834 were approved by the budget owner. The direct approval from administration to final approval by the supervisor covers 2959 cases. Approval by the budget owner is necessary in 1672 cases after approval by administration. But permit can be finally approved by a supervisor directly (2563 cases). It means that the average delay of 66 hours for approval by the budget owner can be a bottleneck of the process.
Some of the travel declarations were rejected in various processing steps.
Table 9 - Rejections

Organisational entity	Number of rejected declarations, %
Employee	2.47
Administration	2.15
Supervisor	0.17
Missing	0.14
Pre-approver	0.12

20 cases were never approved in total.

63% percent of travel declarations are booked on projects. It is remarkable that 71% of international declarations were resubmitted (corrections were made). It means that the success rate to submit the right declaration in time for international trips is significantly less than for domestic declarations. But there were no double payments in the process. In 194 of all the cases employees went on a trip without permit. 2.43% of all the declarations were not preceded properly by an approved travel permit.

311 declarations, 4.82% of all the cases, are first rejected because they were submitted more than 2 months after the end of a trip. These declarations were resubmitted again. All these declarations were rejected by the same organisational unit 65469.

About 2959 travel declarations are not approved by the budget holder in a 7 days time frame and automatically rerouted to the supervisor.

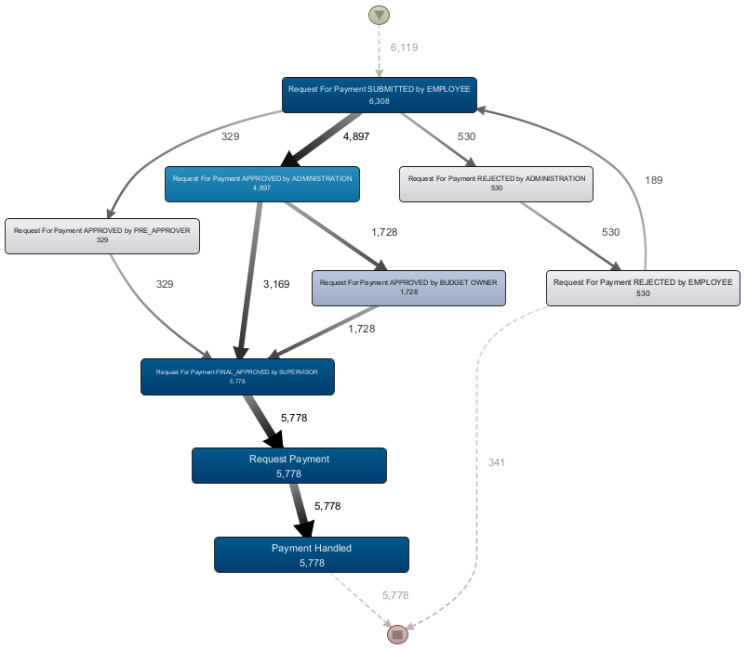


Figure 21 - Requests for payment

In figure 21 we can see that 341 cases, 5.57% - submitted by employees, were rejected by administration automatically / very quickly within 2 hours. Probably these are requests made by mistake and not processed by the system.

5. Conclusion

The business process observed was decomposed and analysed using various methods of process mining. Variant analysis and process discovery allowed us to answer the questions of the challenge.

International declarations had more variety in activities. Many variants were presented in only one case. Though domestic declarations had less variability, the successful submission rate of it was higher than in international declarations.

Mean duration of International declarations take significantly longer than for domestic trips, making the process from submission to payment longer in terms of activity steps.
Process discovery step was held by creating data models. Conformance analysis showed that the best models for domestic declarations were alpha and heuristic miners with high dependency threshold allowing to have 0.96 fitness. That high dependency allows us to observe the most successful trace.

The best model for international declarations was heuristic miner with 0.94 fitness and 0.97 precision. Though accurate filtering was produced using pm4py methods, leaving only successful traces, alpha miner showed low performance of 0.75 fitness. Alpha miner with decreased number of traces showed 93% fitness. Variability of the event log was reduced and the number of different cases was also reduced up to 30% only.

Inductive miner Petri net for international declarations has less simplicity and more generalisation than for domestic declarations due to a high variability of cases.

There are a lot of different rejections in international cases, held by multiple organisational entities, projects and other classes of declarations. Declarations, that were submitted more than 2 months after the end of a trip, were rejected only by one organisational unit 65469, making 52% of total performance of this unit.

The most important bottleneck is related to approvals made by different organisational entities. Budget owner approvals take additional time for both international and domestic trips, making about half and a quarter of all approvals correspondingly take more time than direct approval by administration and then final approval by supervisor. Probably it might be possible to reduce approval time at this point, or number of declarations sent for approval in the budget owner unit.

6. References

1. BPI challenge 2020
2. Pm4Py process mining library.
3. Wil M.P. Van der Aalst - Process mining: discovery, conformance and enhancement - Springer.
4. Source code of the project