BPI Challenge: Process Mining Analysis for Business Travel Reimbursement

Luis González[1], Hernán Arenas[2], José Lara[3], Francisca Rebolledo[4], and Klaus Ribbeck[5]

Pontificia Universidad Católica de Chile
Santiago, Chile
{lagonza2,haarenas,jvlara,fprebolledo,kpribbeck}@uc.cl

Abstract. In this paper we analyze the process of business travel reimbursement for company employees. Data from Eindhoven University of Technology is provided by this year's BPI challenge. We make use of Process Mining tools to analyze this process in search of bottlenecks and inefficiencies. We uploaded the data to Celonis and began to analyze it using tools like the Process Explorer and carrying out conformance checking analyzes. This works shows that in international travel declarations employees first need to go through a permit sub-process, where the trip needs to be approved by a supervisor. This is identified as a bottleneck and averages 41 days to complete. We also found that the budget declared by the employees are having a strong influence in whether the declaration is approved or not. In many cases these rejected declarations are re issued by the corresponding employee, entering a loop of Declaring-Rejection until it gets finally approved or dismissed. Giving special attention to these cases could significantly increase the speed of the process, avoiding the extra work that these loops require.

Keywords: Process Mining · Celonis · Conformance checking.

1 Introduction and Business Questions

The following document aims to provide a Process Mining driven analysis to this year's BPI challenge at https://icpmconference.org/2020/bpi-challenge/. The data to be analyzed consists of business travel declarations with an associated budget submitted by employees from the Eindhoven University of Technology (TU/e). Each declaration must be accepted or rejected by the company to be reimbursed. Trips are divided in two categories, local and international. Declarations from these groups go through slightly different processes, which we will analyze and compare to try to reach conclusions on how these can be improved.

For this study we will work with the Celonis platform. According to its manufacturer, Celonis is the number 1 software for Process Mining (https://www.celonis.com/company), it provides tools to extract, aggregate, and visualize data in simple and useful ways. Allowing us to analyze the main process flow and all its deviations, and to aggregate data in tables, graphs, and other

visual tools. Advanced statistical analysis and machine learning algorithms are integrated natively into Celonis, its supports scripting language allowing user to run advanced prediction techniques directly in Celonis [1]

Techniques used in process mining are useful to extract insights about the performance of a process from a set of event records, or collections of records also known as event logs [2]. We propose three different business questions regarding distinct aspects of the available process logs. In the rest of this report we will attempt to answer these in a way that gives us useful insights of the workflow, so we can provide possible solutions or improvements. These are the following:

- **Question 1:** *Does the differences between national and international declarations affects the throughput time in the overall process time?*
  Identifying the differences between the processes is the first step in order to improve deviations from the main process flow. Additionally, it allows us to discover inefficiencies in activities, which can generate bottlenecks, among other problems.
- **Question 2:** *When comparing cases where a permit was first declared before the trip was taken and cases where it was declared after, what differences/similarities exist, in terms of throughput time, declaration's budgets and percentage of rejection vs declared budget?*
  Understanding the effects of permit declarations that are made after the respective trip was already made can help us identify activities and subprocesses that are slowing down the overall process flow. Analyzing throughput time, budgets and rejections will give us more insight on how the process is behaving, allowing us to better understand the problem and propose possible improvements.
- **Question 3:** *How well do the proposed model adjusts to the model described in the event logs? What workflow is really following the process?*
  A conformance checking analysis will allow us to see where does the model deviates from reality and where are these irregularities occurring. This information will allow us to detect why these deviations happen and provide insights on how to correct them.

2 Question 1: Does the differences between domestic and international declarations affects the throughput time in the overall process time?

2.1 Analysis Method

One advantages of process mining is that it starts with the data that is already there, there is no need to first set up a data collection framework [3]. To answer this and the following questions, we had to preprocess the data so that it was structured in a way so that we could analyze it. We needed to define an activity table to apply Process Mining techniques to our data. First, we define the case-id as our CASE_KEY, concept:name as ACTIVITY_EN, and time:timestamp as

EVENTTIME. Now that we have the minimum to build our activity_table, we add the following columns: Sorting was kept as SORTING, org:role as ROLE, org:resource as RESOURCE, Case-Amount as AMOUNT, Case-Permit OrganizationalEntity as ORG_ENTITY, and Case-Permit RequestedBudget as REQUESTED_BUDGET. Having done this, our data was now ready to process and analyze through the Celonis platform and other tools used in this study.

Using Celonis we analyze the main differences between domestic and international declaration processes, focusing on the permission sub-process and identifying the presence of bottlenecks. In Celonis we created 2 analysis, each one with a different data model (domestic and international), then using the Process Explorer tool we compared both process flows and activities.

2.2 Question-Driven Technical Analysis Results

To analyze how affects the permission sub-process, first we compare all activities that both process, domestic and international, have in common. In the case of international declarations, these correspond to the activities that come just after finishing the trip. As can be seen in figures 1 and 2, the time needed to carry out the 5 activities they have in common is 10 days.

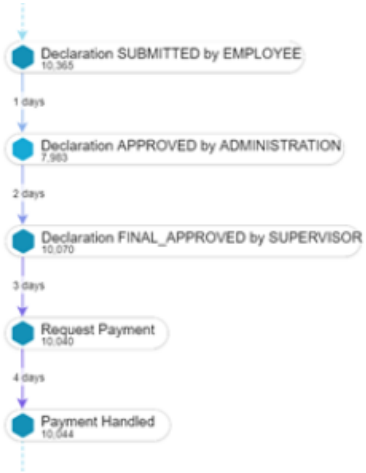

Fig. 1: Domestic declarations.

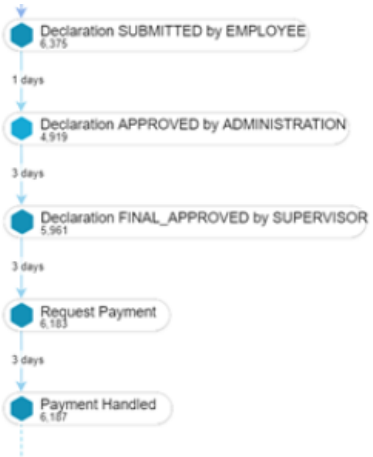

Fig. 2: International declarations.

Now taking into account the time it takes to complete the entire permission sub-process, that is, all the activities that are carried out before the international trip begins (added in the fewest number of variants to maintain consistency in this analysis), it takes 41 days (see figure 3). In other words, this permission process means a 410% increase in time compared to domestic trips (this is without considering the time of the trip itself), so it does have a huge impact on the entire process.

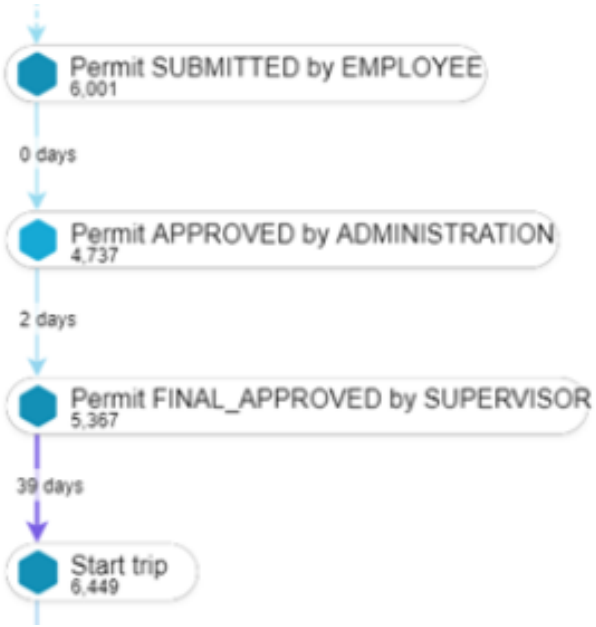

Fig. 3: International declarations on the permission sub-process (with minimal number of case variants).

After all this analysis, a bottleneck was identified, which occurs in the permission sub-process, specifically in the activities of "Permit FINAL_APPROVED by SUPERVISOR" and "Permit FINAL_APPROVED by DIRECTOR" which takes 39 and 55 days respectively (see figure 4). Considering that the average throughput time of the complete process is 86 days, these activities represent 45.3% and 63.9% respectively of the total process time.

2.3 Business Owner Conclusions

In conclusion, the main difference between domestic and international declarations is the permission sub-process. The presence of this sub-process represent a 410% increased time in comparison with the base process flow (the common flow for both domestic and international declarations). Additionally we found a bottleneck in the activities of "Permit FINAL_APPROVED by SUPERVISOR" and "Permit FINAL_APPROVED by DIRECTOR", which represent 45.3% and 63.9% respectively of the total process time. This could be explained because it is only 2 people who must give the final permission, assuming that there is only 1 supervisor and 1 director. A recommendation for reducing this bottleneck is to increase the number of supervisors, or give the authority to other roles to give the final approval.

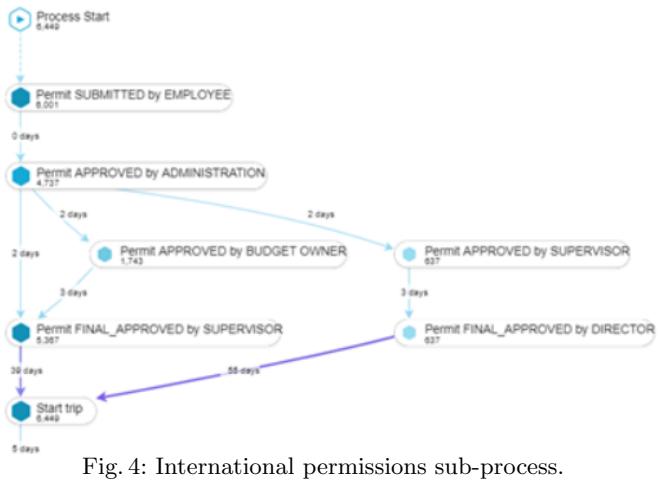

Fig. 4: International permissions sub-process.

3 Question 2: When comparing cases where a permit was first declared before the trip was taken and cases where it was declared after, what differences/similarities exist, in terms of throughput time, declaration's budgets and percentage of rejection vs declared budget?

3.1 Analysis Method

We first clean up the data by extracting only the columns that are of our interest. These include case_key, concept:name, timestamp and amount, assuming the latter to be the declared budget once the trip has ended. When ready, the data is uploaded to the Celonis platform to begin the analysis. Celonis provides tools to visualize the process, apply filters and aggregate data to help us make calculations and reach conclusions. Throughput time is reviewed with graphs made by Celonis. For the budget, we create a set of KPI's (calculated from the database) that will allow us to see if there exists a significant difference between them. Next, we will calculate de average of the "amount" column in the declaration table, then proceed by filtering cases where a rejection has occurred and make a conformance analysis to see how does this affect the process and if there is a connection with this activity and the 'Send Reminder' activity.

3.2 Question-Driven Technical Analysis Results

From now on we will call the cases where the permit is first declared after the start of the trip as Group 1 and the rest of the cases as Group 2.

Regarding throughput time, figures 5 and 6 show us that Group 1 has an average time of 97 days, while Group 2 has an average of 76 days. Furthermore,

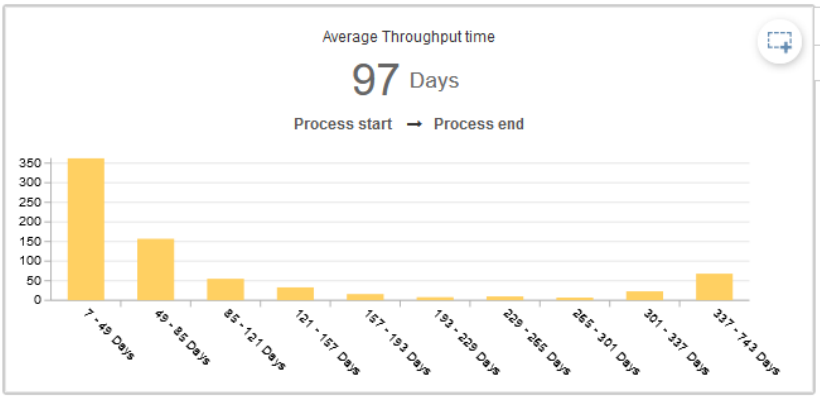

Fig. 5: Group 1 Average throughput time.

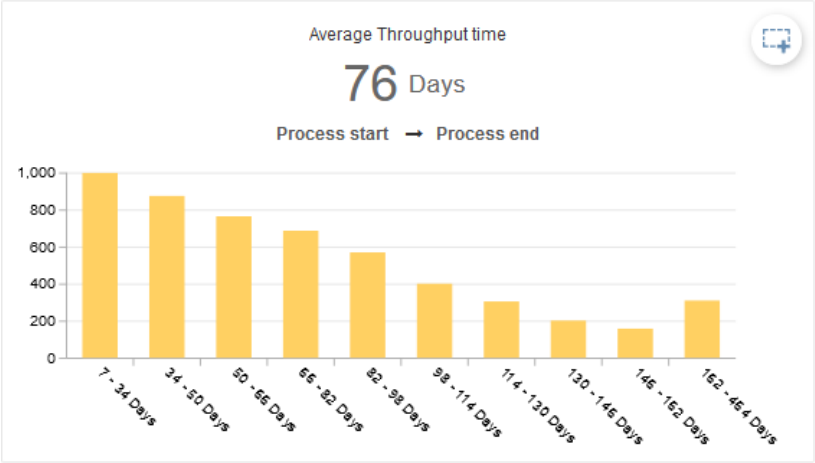

Fig. 6: Group 2 Average throughput time.

cases that last more than 121 days represent 23% of cases in Group 1, but only 16% in Group 2.

From figures 5 and 6 we can see that the distribution of the cases looks quite different for each group. Group 2 looks like a decreasing line equation, where the amount of cases steadily decreases with time. In contrast, Group 1 shows a distribution similar to a U shape, where most cases a grouped in the first two time intervals, but many cases are scattered along the rest of the distribution and accumulate again towards the end, showing large throughput times. We can see that, in average, cases from Group 1 last 21 more days than cases from Group 1.

Analyzing budgets, we found that employees from Group 1 declare permits for an average amount of $664.66 USD, while employees from Group 2 declare an average amount of $882.57 USD. Given this, we conclude that permits declared after the trip has ended tend to be for lower values, compared to those declared with anticipation.

Moving on to percentage of rejected declarations per employee, we found that 21% of cases from Group 1 are rejected, delaying the process for an average of 2 days. On the other side, 24% of cases from Group 2 are rejected, delaying the process for an average of 14 days, which is 7 times as much as those from Group 1. We went a bit further to check if the amount declared also affected the result of the rejections and found that cases rejected from both groups show a higher average value for the declared amount. In case of Group 1, the average budget declared for rejected permits is $836.13 USD ($171.47 USD increase) and for Group 2, the amount went up to $946.71 USD ($64.14 USD increase). Therefore, we see that there is a tendency in both groups to reject declarations with higher budget values. It is also worth noting that 15% of rejected cases from Group 1 and 14% of rejected cases from Group 2 show a declared budget of 0.

Finally, we found that the 'Send Reminder' activity occurs in 14% of cases from Group 1, delaying the process for an average of 24 days, while it occurs only in 6% of cases from Group 2, but delaying the process for an average of 42 days.

3.3 Business Owner Conclusions

In conclusion, cases from Group 1 declare smaller budgets than those from Group 2 and have an average throughput time that is 21 days larger, given this, we advise to take a closer look to business trips that begin before a permit has been declared. Both groups show a similar percentage of rejected declarations, where many of these are either with a higher declared budget or with budget equal to 0. Also, rejected declaration that belong to Group 2 are delayed 14 days in average, which is a lot more than the rejected cases from the opposite group and is the most recurrent deviation in the main process. This makes it an incredibly significant point to try to avoid or to optimize, since it slows down the process and has a major chance of happening.

4 Question 3: How well do the proposed model adjusts to the model described in the event logs? What workflow is really following the process?

4.1 Analysis Method

We will be doing conformance checking analysis to check how our representation of the model differentiates from the actual process described by the event logs. To begin this, we started by creating a BPMN model of the processes described by the authors of the challenge. One model was created for the international trips and another for the domestic trips. A Business Process Modeling Notation (BPMN) model is a tool for visually representing a process. This model gives information of the activities/events inside a process and the sequence flows they follow, with information of the responsible parties and other relevant external components inside a process. Actually, BPMN model is widely used languages to model business processes [2]. BPMN is avalaible in tools and has been standardized by the OMG [4], our model will be created with the Bizagi tool. BPMN is a standard model used for modeling business process, in a graphical notation for specifying business processes based on a flowcharting technique. The main objective of a BPMN model is to support business process management, useful for technical users and business users, for that provides an intuitive notation for both user groups [5].

To see how the process is really behaving we will use the Celonis platform. We will be working with the process explorer component, which allows us to visualize the activities and most common workflows and its deviations. Then, we will compare the results shown by Celonis with the BPMN model we already created. In addition, we will use the Python library Pandas to do statistical calculations and to further process our data to reach important conclusions from the process.

4.2 Question-Driven Technical Analysis Results

Before starting with the analysis, we point out two important characteristics of the studied processes, which will help us define useful metrics:

- Each time the permit/declaration 'SUBMITTED' event is present in a case, we say there was an attempt generated by the employee. Every attempt present in a case means that the employee had to re issue his permit/declaration.
- The time it takes to approve these attempts is the difference between the first 'SUBMITTED' activity and the 'APPROVED' activity. If there is no 'APPROVED' activity, it means the employee's declaration was rejected and he/she gave up.

International trips analysis (6450 cases):

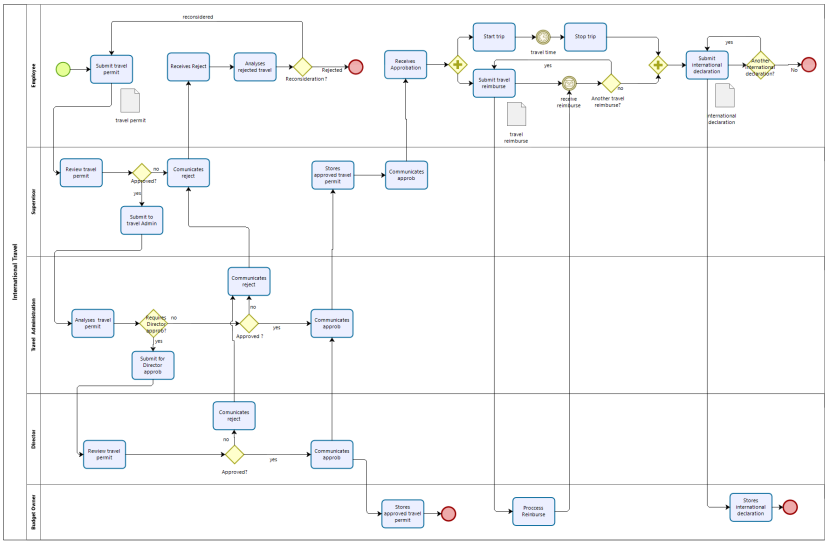

Fig. 7: International BPMN model.

To analyze the international travel process, we built a diagram of BPMN business models and processes (figure 7 International BPMN model), which allows us to identify the logic of the information flows for the international travel process.

The main differences that we found between the BPMN model performed with what Celonis gave us for the first 7 variants (figure 8), was the starting activity.

There are 3 possibilities (one will be omitted since it represents less than 1% of the cases), being the following:

- Start trip (11% of total cases): We can see in figure 9 how employees send a permit after they have finished travelling or during their trip.
- Declaration SUBMITTED by EMPLOYEE (6% of total cases): In this case, employees declared their trip before requesting permission or travelling. This may mean that these entries are misclassified and should be in a different dataset, or that cases where the trip was already planned by the company in question and the employees only have to present a declaration, because the permits were assumed to be granted.

Other differences were found, but because they were use cases with less than 5 occurrences, so these were not analyzed in depth. Besides, the differences in these cases were about the number of times an employee was rejected or by whom the declaration was passing through (director, budget owner, etc.).

To be able to better understand these processes, the dataset entries were grouped using their case_id. Then, we analyzed the loops in which the employees try to approve their requests, which were described in the previous section. We found the following:

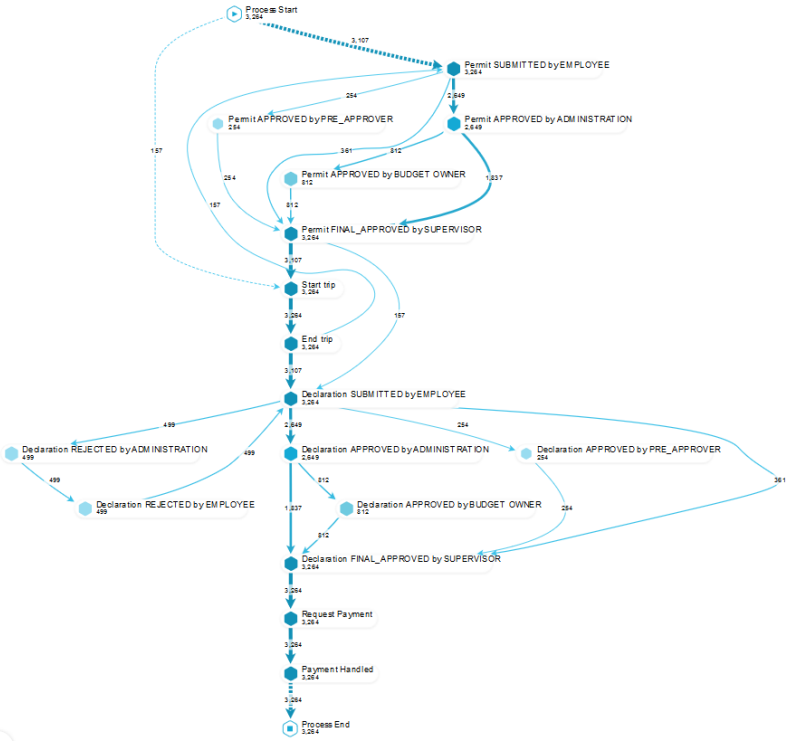

Fig. 8: International 7 most common variants.

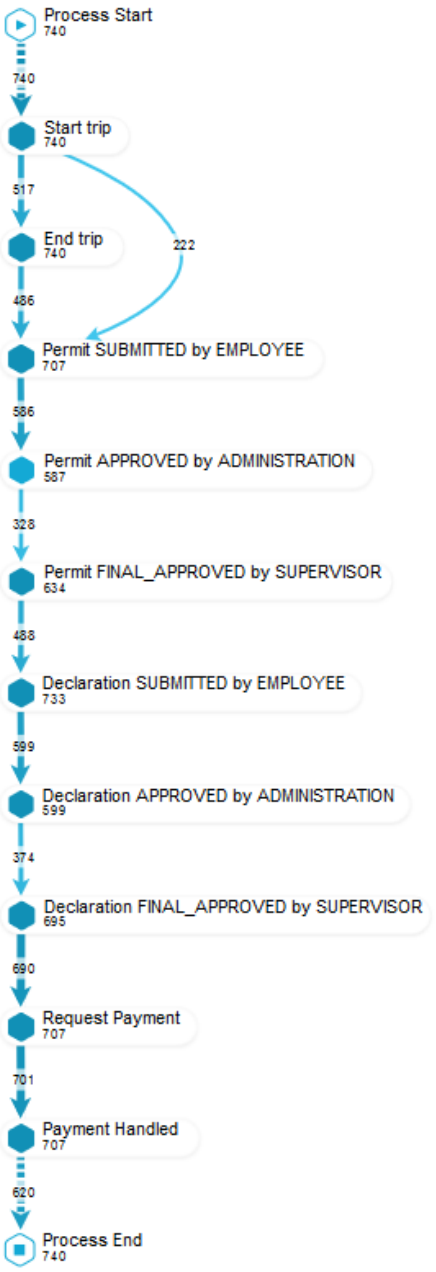

Fig. 9: Start trip.

- On average, an employee tries to get his permit accepted 1.04 times, which is quite good (leaving out cases where permission was skipped).
- On average, an employee tries to get his permit accepted 1.247 times (leaving out cases where permission was skipped).

As for the permit attempts, these cases were grouped by the number of attempts made in each. In addition, we calculated the number of attempts that each organization made. Regarding permits we have the following (figures 12 and 13).

- Increasing the case budget permit increases the number of attempts (except for attempts 3 and 4, since they only have 18 entries so we can consider them as outliers).
- Organization 65454 skipped permits 448 times and was the only one to carry out this practice. This may indicate previously planned trips, where they themselves were responsible for carrying them out, since the declarations were made.
- Being rejected the first time increases the wait time by an average of 7.2 days to be approved on the second attempt.

Proceeding to carry out this same analysis for the declaration cases, we have the following (figures 14 and 15, declaration table):

- There is a tendency for the employee to be rejected if the amount that he wants to reimburse is larger. This greatly affects the number of days he ends up waiting to be approved.
- Being rejected on the first attempt increases an average of 6.8 more days in being approved the second time.

Domestic trips analysis (10500 cases):

For this case, to analyze the domestic trip, we also built a BPMN diagram (figure 10)

In this case, when we compared our BPMN model to the representation made by Celonis that includes the 7 most common variants (figure 11), we found no differences. As with international travels, loops are also present in this model. These consist of employee's declaration, so we will proceed to analyze these statistically to be able to propose possible improvements.

- For an employee, it takes an average of 1.11 attempts for a declaration to be approved (leaving out cases where the 'ACCEPTED' activity did not happen).
- 174 cases were rejected.

When grouping by the amount of declaration attempts and calculating aggregated values, we get the following (figure 14):

- Increasing the amount requested by the employee increases the number of attempts the employee must make. In other words, higher amounts increase the chance of the declaration being rejected.

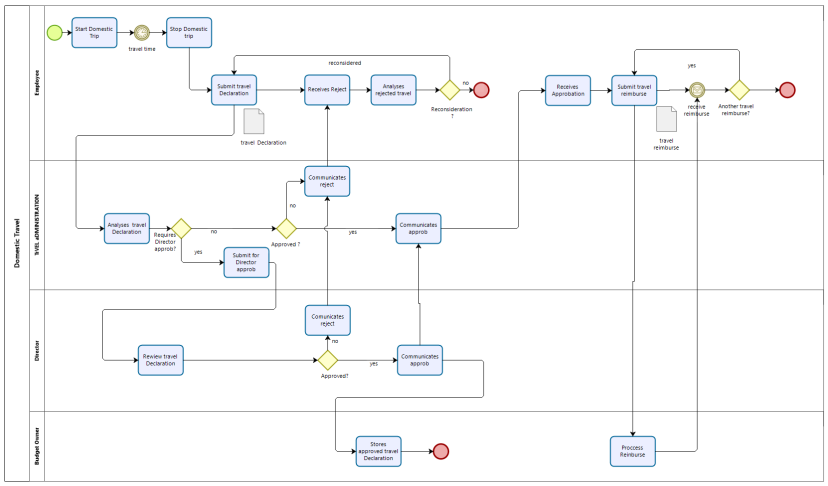

Fig. 10: National BPMN model.

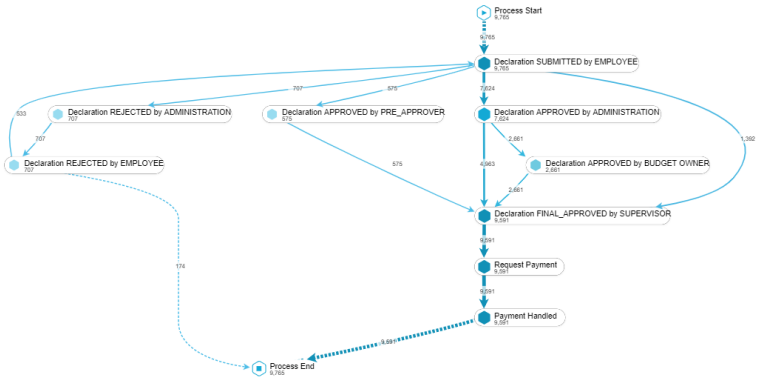

Fig. 11: National 7 most common variants.

– The amount of time it takes to process a declaration, once it was rejected for the first time is a lot higher that the time it takes on the first attempt (10-day difference). As more attempts occur, even larger amounts of time elapse before it is processed.

4.3 Business Owner Conclusions

An important conclusion obtained from the analyzes is that the amount or case budget is a key factor in deciding whether it is approved or not. This applies for both permits and declarations. That is why we suggest paying special attention to these cases, since a large percentage is approved in the same way, but with a significant expense of time for the employee. A possible way to improve these, would be to create a specialized entity assigned to review only requests with a budget above a certain amount. This way, these cases could be reviewed in detail in a much faster manner, optimizing the process and wasting less time for everyone.

5 Conclusions

Many companies have the need to pay for business travels for their employees and need to go through a standardized process. In this study we analyzed the process proposed in the BPI challenge corresponding to TU/e. We compared international travel declarations with national travel declarations. We found that the main difference between national and international declaration processes is the permit sub—process. The final approval of these permits is the bottleneck of this process and would be good to optimize this activity. Putting more people in charge of the final approval would be a good way to accelerate the process.

After analyzing both national and international declarations, we can conclude that one of the main reasons for delays in the process is the rejection of permits. Declared budgets seem to be a major influence in whether the permit is approved or not, though many times these cases end up being approved anyway, meaning the employee had to emit a permit many times, and these needed to be processed many times also. This is a clear example of extra work and a delay for the process. Finally, we advise to assign a specialized entity to be in charge of declarations over certain budgets, to be able to give special care to, so that these can be resolved avoiding all the extra work.

References

1. Veit, Fabian Geyer-Klingeberg, Jerome Madrzak, Julian Haug, Manuel Thomson *The Proactive Insights Engine: Process Mining meets Machine Learning and Artificial Intelligence* Springer International Publishing AG 2017
2. Wil van der Aalst *Process Mining Data Science in Action Second Edition.* Springer Heidelberg New York Dordrecht London.
3. Fluxixom *Process Mining Book* https://fluxicon.com/book/read/

4. *Business Process Model and Notation (BPMN).* OMG Object Management Group, 2010.
5. Jorge Muñoz-Gama *Conformance Checking and Diagnosis in Process Mining* Springer International Publishing AG 2016

6 Attachments

Permit attempts	Average delay in days	Quantity	Average Budget-permit
0	0.000000	448	923.081315
1	3.675394	5767	1254.384879
2	10.907407	216	1546.166289
3	22.625000	16	1180.084201
4	63.500000	2	601.000157

Fig. 12: Grouped permit attempts.

Organization	0 Attempts	1 Attempts	2 Attempts	3 Attempts	4 Attempts	Mean of attemps
organizational unit 65458	0	1364	62	5	1	1.05
organizational unit 65455	0	1044	36	1	0	1.04
organizational unit 65456	0	857	32	0	0	1.04
organizational unit 65459	0	485	8	0	0	1.02
organizational unit 65454	448	477	31	1	0	0.57
organizational unit 65460	0	472	8	0	0	1.02
organizational unit 65464	0	300	5	0	0	1.02
organizational unit 65457	0	299	5	0	0	1.02
organizational unit 65466	0	188	25	6	1	1.18
organizational unit 65461	0	75	2	0	0	1.03
organizational unit 65469	0	50	0	0	0	1
organizational unit 65470	0	40	0	0	0	1
organizational unit 65467	0	18	0	0	0	1
organizational unit 65472	0	18	0	1	0	1.11
organizational unit 65475	0	16	0	0	0	1
organizational unit 65468	0	13	0	0	0	1
organizational unit 65473	0	11	1	0	0	1.08
organizational unit 65465	0	8	0	0	0	1
organizational unit 65471	0	7	0	0	0	1
organizational unit 65480	0	7	0	2	0	1.44
organizational unit 65482	0	6	1	0	0	1.14
organizational unit 65477	0	4	0	0	0	1
organizational unit 65484	0	3	0	0	0	1
organizational unit 65486	0	2	0	0	0	1
organizational unit 65462	0	1	0	0	0	1
organizational unit 65478	0	1	0	0	0	1
organizational unit 65488	0	1	0	0	0	1

Fig. 13: Permit attempts by organization.

Declaration attempts	Average delay in days	Quantity	Average amount
0	0.000000	74	0.000000
1	5.763863	4815	730.029969
2	12.557232	1127	999.582381
3	20.779279	222	1137.194885
4	29.344828	29	1138.470090
5	32.636364	11	1562.833164

Fig. 14: Grouped declaration attempts.

Organization	0 Attempts	1 Attempts	2 Attempts	3 Attempts	4 Attempts	Mean of attemps
organizational unit 65458	15	1149	215	46	4	1.22
organizational unit 65455	7	952	104	16	2	1.12
organizational unit 65454	15	716	185	36	5	1.27
organizational unit 65456	9	632	193	44	7	1.35
organizational unit 65460	9	370	87	12	1	1.23
organizational unit 65459	12	336	109	35	1	1.34
organizational unit 65457	2	236	63	1	2	1.23
organizational unit 65464	1	217	71	11	5	1.35
organizational unit 65466	0	150	50	16	1	1.44
organizational unit 65461	0	64	10	2	1	1.22
organizational unit 65469	0	36	14	0	0	1.28
organizational unit 65470	1	26	9	4	0	1.4
organizational unit 65467	0	13	4	1	0	1.33
organizational unit 65475	1	12	3	0	0	1.13
organizational unit 65472	0	10	7	1	1	1.63
organizational unit 65473	0	10	2	0	0	1.17
organizational unit 65468	2	10	1	0	0	0.92
organizational unit 65480	0	6	3	0	0	1.33
organizational unit 65465	0	5	3	0	0	1.38
organizational unit 65482	0	5	1	1	0	1.43
organizational unit 65471	0	4	3	0	0	1.43
organizational unit 65477	0	3	1	0	0	1.25
organizational unit 65484	0	3	0	0	0	1
organizational unit 65486	0	2	0	0	0	1
organizational unit 65462	0	1	0	0	0	1
organizational unit 65478	0	1	0	0	0	1
organizational unit 65488	0	1	0	0	0	1

Fig. 15: Declaration attempts by organization.