Optimization and Analysis of Academic Travel Processes Considering Throughput Time

Nicolas Acosta, Roshad Alipanah, Vicente Etchegaray,
Francisca Ibarra, Flavio Tarsetti, and Victor Galvez

Department of Computer Science
School of Engineering
Pontificia Universidad Catolica de Chile
{nlacosta, ralipanah, vaetchegaray,
faibarra1, flavio.tarsetti, vagalvez}@uc.cl

Abstract. For large institutions such as universities, travel-related processes are of high importance. This research makes use of different process mining techniques in order to analyse and improve a subset of processes related to the travel of the Eindhoven University of Technology academic staff, using as a main KPI the throughput time.
In particular, we found a lack of supervision regarding the international flights and a rejection flow that can end up with multiple successive rejections which adds more days to the initial throughput time of the travel declaration. Also, we found a possible subset of variables that can help reduce the time on the request for payment subprocess.
The different methodologies used during this work enable an approach from multiple points of view of the problem while ensuring the option to replicate the results in other process of multiple types, using the studied cases as an applied example that will allow us to look at specific flows and activities that are detected as having an impact in the overall throughput time of the process.

Keywords: Process Mining · Throughput Time · Rejections Flows

1 Introduction and Business Questions

The Eindhoven University of Technology has a considerable amount of staff travelling around the globe for conferences or diverse meetings regarding academic affairs [1]. In consequence, having a properly defined and optimized flow for each key process involved in these travels is crucial for the university. Based on this idea, we decided to analyse certain sections of the overall flow of our interest: the travel permits, the travel declarations from international and national flights, and the requests for payment of non-travel costs. Focusing on these processes, we will point out ways in which they could be improved and reduce their overall throughput time and rejection ratio.

We were working with data directly related to the processes described previously. To be more specific, the dataset is composed of a number of files, obtained from the logs that are generated by the Information Systems used by the University that supports the process. This allowed us to apply process mining techniques to the dataset (Case Id, Activity name and Timestamp). We then generalize the general flow of each of the processes and make a model out of it that we will work with. Each process (reimbursement, declarations and permits) has its own dataset, since there is data that is unique to each of the flows that needs to be considered, considering mainly the different activities present in each dataset, which are analyzed in the corresponding question following the subprocess of interest.

The challenge [1] describes different questions that could be answered with the analysis of the dataset, which we took into consideration when choosing the objectives to tackle in the current analysis. The focus was put on looking at the problems that, if solved, could have the biggest impact in the overall process, considering mainly throughput times. We took into consideration the overall flow of the process, and we chose specific aspects to focus on the analysis, creating business questions that, when answered, will give value to the process owner. The selected questions are:

- *How can we compare and improve the throughput time of domestic and international flights considering their differences?*
  To analyse the travel permits and the declarations part of the process we will first look at the differences between international and national flights, looking at the impact of these differences in the throughput times, gaining insight in which activities must be removed or improved in order to improve the overall throughput time.

- *How can we understand and improve the rejections flows of the national and international declarations?*
  Following the differences, we will focus on the declarations flow for both of the flight declaration types, more specifically on the rejections that occur in the process and the impact that they have in the overall throughput time.

We will analyse the different successions of activities that are present in the dataset to gain insight in the impact that they have in the overall process, while trying to improve the patterns encountered during the analysis.

– *Which variables have a significant influence on the average throughput time of the Requests For Payment?*
For this objective, we have selected a subset of variables (such as unexpected activities, case project, organizational entity, amount requested and number of rejections in a case) that we think have an impact on the overall throughput time of the flow. We will then start by obtaining the average throughput time of the process as a whole and then, we will do a brief analysis for each proposed variable in order to measure its influence on the throughput time. Finally, we will filter the variables that affected the throughput time and group them to have the optimal combination to minimize throughput time.

2 Question 1: How can we compare and improve the throughput time of domestic and international flights considering their differences?

2.1 Analysis Method

Data used In order to compare both types of flights, we used the datasets `DomesticDeclarations.csv` [1] and `InternationalDeclarations.csv` [1] and used from both tables the columns "Case-Id", "time:timestamp" and "concept:name" as the Case ID, Timestamp and Activity attributes respectively in the Celonis Intelligent Business Cloud [2].

Through an initial overview of the data, we decided to separate our analysis focusing into two different aspects: A comparison of the declaration process flow between domestic and international flights (as they share similar activities), and an analysis of the travel permit process flow (as they are activities that only happen during the international process).

Declaration process flow comparison We used a standard KPI of throughput time to determine the average time that declaration process took for each type of flight (starting from the activity "Declaration SUBMITTED by EMPLOYEE" to the activity "Payment Handled").

Then, by using the "Process Explorer" tool [3], we determined which activities were unique for each type of flight (in terms of the declaration process flow). By using our initial standard KPI, we measured their impact in the process flow considering average throughput time and by implementing the "Conformance" tool [4]. Then, based on the information provided in the challenge description [1], we created an ideal process model for international flights (Figure 1) and domestic flights (Figure 2). Finally, we determined which were the most common violations for each process and how significant was their impact in the ideal throughput time.

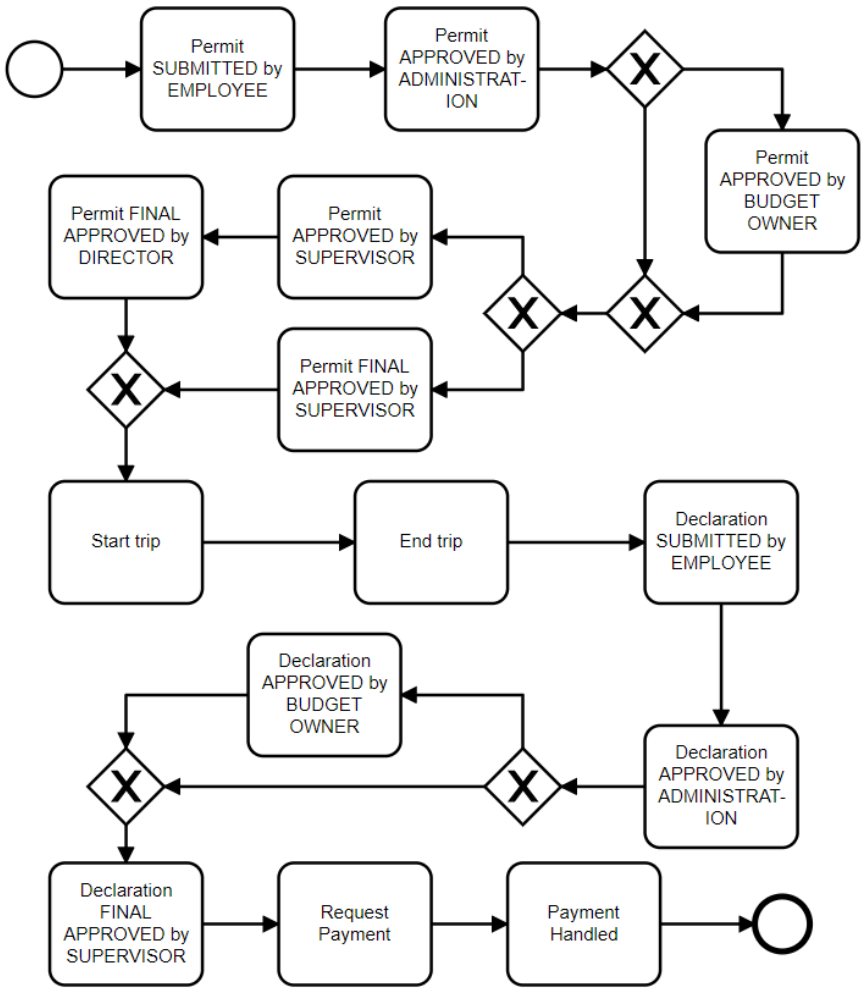

Fig. 1. BPMN diagram of ideal international process model.

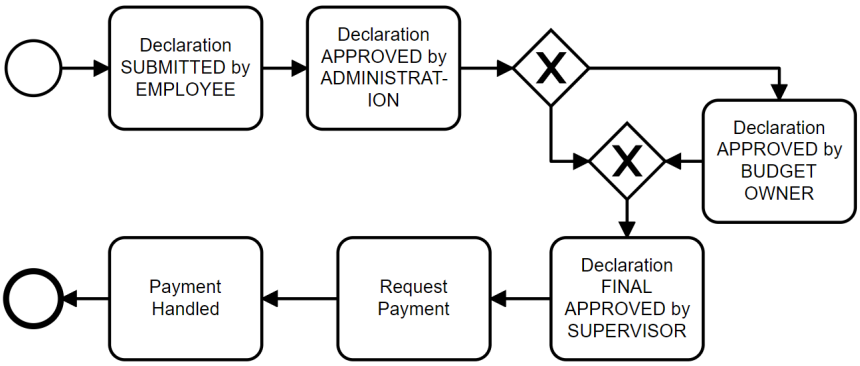

Fig. 2. BPMN diagram of ideal domestic process model.

Permit process flow analysis For this analysis we also used a standard KPI of throughput time to calculate the average permit process took, starting from the activity "Permit SUBMITTED by EMPLOYEE" to the activity "Permit FINAL APPROVED by SUPERVISOR" (or in some cases, "Permit FINAL APPROVED by SUPERVISOR"). Afterwards, using the "Process Explorer" tool [3] and the same standard KPI, we determined the impact of the different activities in the process flow in the overall throughput time.

Finally, we used the "Conformance" checking tool [4] to extract the most significant violations of the process. The first **53%** of the cases where considered, because including more cases implied including some outliers on the throughput time (cases that greatly exceeded twice the median) and those were the abnormalities that we wished to analyze. Also, with more coverage we included some model paths that were abnormal from a business logic point of view such as trips starting without permission.

2.2 Question-Driven Technical Analysis Results

Following our analysis method and limiting ourselves to only the declaration process (considering all cases), we started by determining that the average throughput time for the declaration process flow in domestic flights is **11.46 days**; as for the declaration process flow in international flights, the average throughput time is **14.46 days**.

For the latter, we discovered that it has a unique activity in its declaration process flow that's not present in its domestic counterpart: "Declaration FINAL APPROVED by DIRECTOR". Although this activity is present in only **4%** of the total cases, we can see that it extends the average throughput time to

18.24 days, around 4 more days. On the other hand, when the declaration request receives final approval by the supervisor (represented by the activity "Declaration FINAL APPROVED by SUPERVISOR", present in more than **90%** of cases in both flight types), the overall average throughput time stays the same.

A very similar behaviour occurs when we analyze the permit process flow of international flights. We determined that when the travel permit receives final approval from the supervisor (in around **83%** of the cases), the average throughput time is **3.86 days**. When the final approval is executed by the director (in around **10%** of the cases), the average throughput time increases to **6.96 days**.

Significant Violations When creating the ideal process models for both flight types (for their respective conformance checks), we discovered that in the case of domestic flights only the first two variants were enough to fit the given description of its process flow, covering **65%** of the cases (with an average throughput time of **10.5 days**, just a little shorter than the overall average of **11.46 days**). On the other hand, creating its international counterpart proved to be a more difficult task, as this type of flights contains two different processes, with only **35%** of the cases conforming to the ideal flow (with an average throughput time of **71.7 days** for the entire process).

Declaration re-submissions As far as deviations from the ideal models go, we discovered that although most of them are caused directly by rejections (which are to be discussed into detail in the second objective of this report), a very significant violation present in both flight types is when several declaration submissions by an employee take place in one single case. In both flight types, this deviation had a similar effect in the average ideal throughput time, extending it in **12 days** and in **17 days** for domestic and international flights respectively.

However, a key difference between them is that in international flights, there was a very significant increase in violation cases after the process model was perfected: from **8 cases** monthly in average in 2017 to around **100 cases** per month the next year, reaching a peak of approximately **150 cases** in several months. As for domestic flights, there was only a slight increase from the pilot year of 2017 (with around **14 cases** per month) to 2018 (which pretty much stayed consistent with **50-60 cases** monthly).

Immediate final approval Another major violation shared by both types of flights, was when a declaration submission (made by an employee) was directly followed by the final approval from the supervisor (present in **20%** and **19%** of the cases in domestic and international processes respectively), skipping required approvals from other roles. Although the time this specific succession of activities took around the same time (**1 day**) in both types of flight, there is a very substantial difference when it comes to the impact it had on the average ideal throughput time.

In domestic flights, this deviation had basically no impact at all in the ideal process in terms of throughput time. On the other hand, in international flights,

this was one of the most severe violations as it added an additional **42 days** to the entire process. When analyzing the different international flight cases that presented this issue, we were able to discover that the attribute "CASE-PERMIT ACTIVITY NUMBER" with a value of "UNKNOWN" in the original `InternationalDeclarations.csv` [1] dataset, was the main root cause for this violation (having a strong correlation with almost every single case). This means that there might be a problem in the travel permit process regarding unregistered or unspecified information that could be causing this issue.

Overall Conformance Check Aside from those violations, the domestic travel process, since it doesn't require a lot of steps in terms of authorizations (as it is only conformed of one specific process and there is no involvement from the director), is much more fault tolerant: most of the other deviations only occur in **1%** or less of the cases.

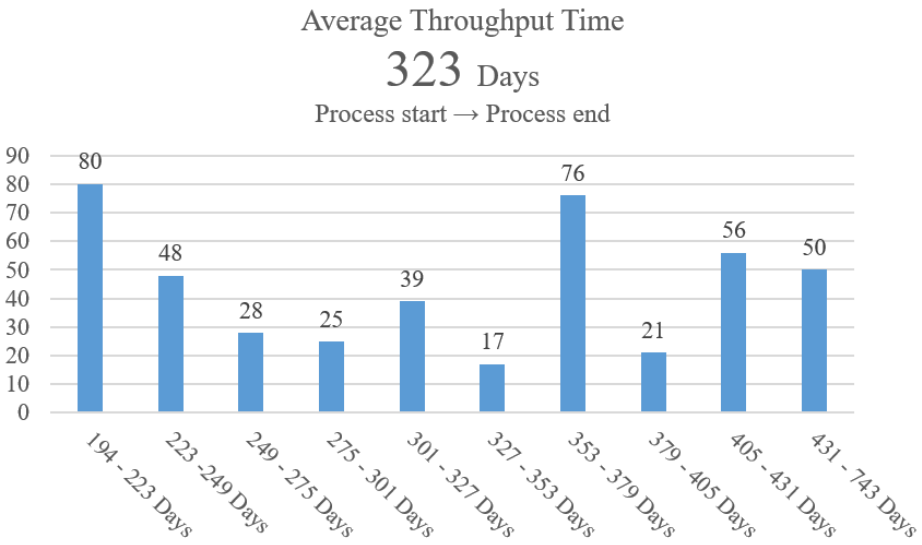

Fig. 3. Cases exceeding 194 days.

In comparison, the International travel process is much more prone to violations than the domestic travel process. When checking the permits process, we detected that in the international flights process, **12%** of the trips are started without authorization, even when it's required for all travels. Also, when choosing the 9 variants with a higher throughput time, there are cases when the

employee has a rejection of the permit but executes the trip anyways. After the trip, it sends another request for permit, but this only occurs after an average of **194 days**. That amount of days is not healthy for the process, because the median throughput time for all of the process is **42** days, leading to a loss of control over those specific cases due to their length. Even more, those processes that exceed **194 days** have an average of **323 days**, almost a year of handling those cases.

2.3 Business Owner Conclusions

In accordance with our analysis and comparison between both types of flights, we can conclude that overall, the domestic process flow is much less susceptible to deviations, as it is conformed by a very simple and specific sequence of activities (as opposed to its international counterpart). International processes require a significant higher amount of approvals and authorizations, and tend to involve the participation of a key role in the approval an rejection of permits and declarations: the organization's director (which is probably due to the intricate nature and multiple branches of the different processes for this type of flight) which, in consequence, impacts the overall throughput time.

Also, based on our conformance checking analysis, we can conclude that the lack of reappraisal and correction of unregistered or undefined data can have major repercussions in the throughput time of these activities, specially if they are part of the more complex international process flow.

Therefore, our main suggestions to improve the overall throughput time is to specify which activities should each role be in charge of, and to better define the sequence of activities that are necessary for the authorization of business trips and the approval of travel permits and declarations. Also, we recommend to improve and make data registration and more rigorous supervision, in order to avoid possible disruptions in the process flow, specially in the international travel process which tends to be more susceptible to deviations.

In this objective, we see further analysis opportunities in checking those cases where there are activities that occur between the start and end of the trip. Those cases should not be possible, but they exist in our analysis. Also there is an opportunity to check internally why some cases take so long to be finished, because those cases could later be lost on the regular process, so the implementation of a process of reminders should be evaluated.

3 Question 2: How can we review the performance of the main rejection flow of the national and international declarations considering the different variations?

3.1 Analysis Method

To understand the rejection flows of the process of declarations, we must first process the available data, defining the datasets that will be used and the methodology to select the information that is important to solve the question.

Data used We have limited the problem to the declarations only, so we used the domestic and international declarations datasets. As told in the previous question, the differences between international and national declarations are minimal, so we considered the union of both datasets, cropping the process of international declarations to only consider the activities that occur between "DECLARATION Submitted by EMPLOYEE" and the end of the process (following the model shown in Figure 2 of the previous question).

We considered the column "Case-Id" as the Case ID parameter, "timestamp" as the Timestamp parameter, "Event-Id" as the Id parameter of each activity and "concept:name" as the activity name, which includes the role of the performer of the activity, considering the same datasets as for the previous question: `DomesticDeclarations.csv` [1] and `InternationalDeclarations.csv` [1].

Data transformation Due to technical limitations of the tool used (a maximum of 100,000 rows per activity table, we had to limit the number of cases of the combined dataset), for this we randomly removed cases from the data analysed, going from 13,580 cases to 12,688, which allowed us to comply with the technical limits while conserving the overall distribution of the cases. The selection of the deleted cases was done using the random module provided by the Python Standard Library, selecting randomly different Case-Id from those present in the dataset and removing all the rows corresponding to this specific case.

Another important consideration in the analysis method is that there's a change that occurred during the process described by the logs in relation to the process of approval/rejection of the declarations. In 2018, the ADMINISTRATION role takes a bigger part in the process, approving the majority of the declarations that are present in the dataset. In fact, after 2018, 96% and 97% of the total cases of the domestic and international declarations respectively were approved by the ADMINISTRATION, which was nearly absent in the years prior, when the approval of the supervisor was sufficient, with only a few cases of declarations that were approved by PRE_APPROVER. This is a big change in the overall process, that also impacts in the rejections activities, which are the ones that will be analysed for this objective.

Considering this change, we will only consider the cases that occur after 2018, to be able to obtain meaningful insights, considering the different number of activities required to approve a declaration and the impact that could have in the rejection numbers.

Data analysis For the analysis, the main variables considered are case count and average throughput time as the KPI. The average throughput time is rounded to the closest day number, considering only work days in the computation. The process was modeled as shown in Figures 1 and 2 and as a Petri Net using the the Conformance Checking capabilities of Celonis. We obtained similar results as those shown in the previous question, which allowed us to get an overview of the rejections flows, and pointed out the particularities present in the dataset

that are then analyzed in detail. The data mining techniques used are those present in Celonis Intelligent Business Cloud [2], which gave us all the variables considered, while giving us the possibility to filter on the data to look at patterns in the process which are explained in the analysis.

3.2 Question-Driven Technical Analysis Results

Following the explained method, we identified the following rejection activities, considering the performer of the activity as part of the key difference: Declaration REJECTED by BUDGET OWNER, Declaration REJECTED by ADMINISTRATION, Declaration REJECTED by SUPERVISOR, Declaration REJECTED by EMPLOYEE and Declaration REJECTED by DIRECTOR. We then obtained the case count for each activity (see Table 1):

Table 1. Different rejection activities and case counts by declaration type.

Rejection executor	Domestic count	International count	Total count
BUDGET OWNER	57	35	92
ADMINISTRATION	817	1,205	2,022
SUPERVISOR	207	88	295
EMPLOYEE	1,031	1,286	2,317
DIRECTOR	0	1	1

Rejection flows With the overview of the rejections activities in mind, we proceeded to identify and analyze the different patterns of successive rejections activities present in the dataset. In total, there are **2,331** cases that contain a rejection activity, corresponding to **18%** of the total case number.

Main rejection flow The most common rejection flow correspond to the succession of "Declaration REJECTED by ADMINISTRATION" with "Declaration REJECTED by EMPLOYEE", present in **1,990** cases (**16%** of the total case count and **85%** of the cases that contains a rejection activity). This flow can lead to two different outcomes: the re-submission of the declaration by the employee or the end of the declaration. The first case occurs in **1,699** cases, leading to another flow of submission of the declaration, which can be approved or rejected again.

The average throughput time between the 2 rejections activities is **3 days**, which must be added to the time that takes to resubmit the declaration in the cases when the declaration is submitted again (average time between "Declaration REJECTED by EMPLOYEE" and "Declaration SUBMITTED by EMPLOYEE"), which has an average throughput time of **4 days**, leading to a total of **7 days** of delay in average for the most common rejected declarations.

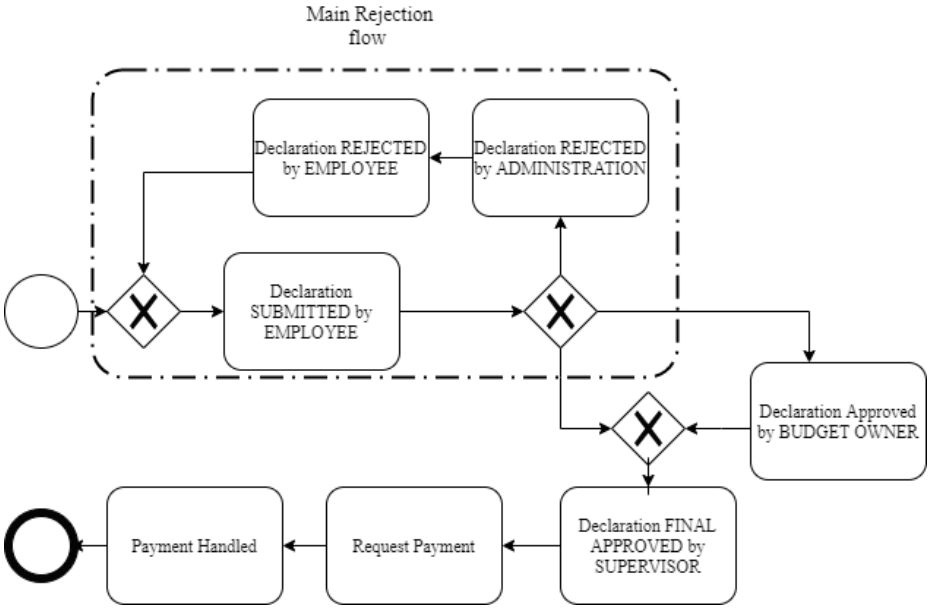

Fig. 4. BPMN diagram of the declarations subprocess, with the most common rejection flow (which can be repeated multiple times).

Main rejection flow variations The main variations of this flow consists of subsequent rejections of the same declarations that have already been rejected, with the majority of these cases containing 2 or more consecutive rejections that follow the same pattern ("Declaration REJECTED by ADMINISTRATION" followed by "Declaration REJECTED by EMPLOYEE" activity), as shown in Figure 4. The case count for these variations is summarized in the Table 2, which also specifies the number of cases where the declaration is resubmitted and where it is canceled after the last rejection. The number of rejections corresponds to the number of rejections that occur in the same case containing the main rejection flow.

Table 2. Different variations of the main rejection flow cases count

Number of rejections	Declarations re-submitted	Declarations canceled	Total case count
1	1.421	277	1.699
2	231	22	253
3	24	6	30
4	7	1	8

We conclude that the majority of rejections involve a re-submission of the declaration, with an average of **85%** of declarations that are resubmitted in the cases considered. We also conclude that the number of canceled declarations doesn't correlate to the number of prior rejections that the declarations have in the logs, which shows the importance of analyzing the throughput time of these rejections flows, given that they can occur multiple times for the same declaration and are nearly as important in number as the canceled declarations.

Throughput time analysis To measure the impact of these rejections flows in the overall process, we analyzed their throughput time, considering the distribution of the throughput time, which corresponds to the total time that occurs between the first rejection by the administration and the last declaration submission by the employee, considering only the cases where the declaration is re-submitted and is eventually approved. This analysis was done for all the rejections numbers shown in the last table, with the aim of displaying the impact that the subsequent rejections of the declaration can have in the throughput time of the process. The results of this analysis are presented in Figure 5.

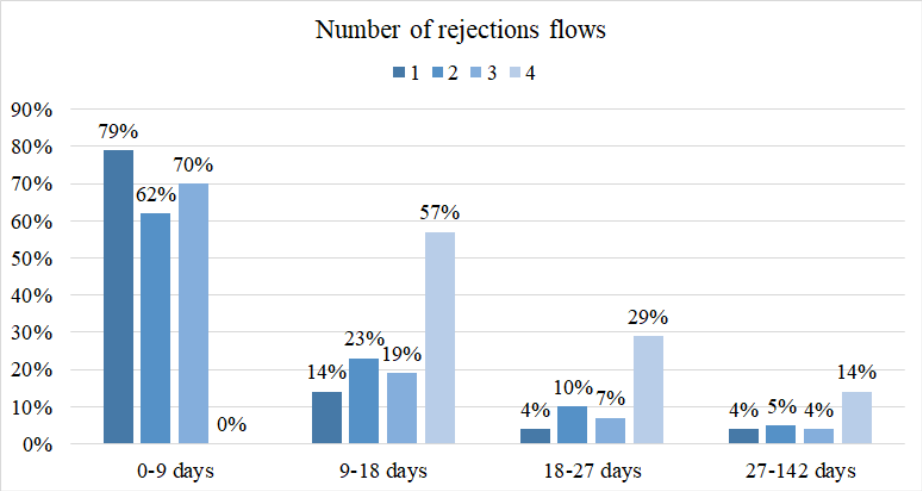

Fig. 5. Distribution of throughput times of rejections flows depending on the number of consecutive rejections.

The average throughput time of rejections flows are presented in Table 3. The data shows that the throughput time of the total rejection flow grows depending on the number of rejections, following a nearly lineal growth, except for the declarations including 4 rejections, which present an erratic behaviour, due probably to the small sample size (7 cases are considered).

The distribution also gives an insight about the possible presence of outliers, which correspond to cases that have an abnormally high throughput time, specially considering the differences existing in the median and average throughput time (as shown in Table 3). These cases have a bigger impact in the average time when considering rejections flows that include 3 or more rejections, due to their much inferior number of cases.

Table 3. Average throughput times of rejections flows

Number of rejections	Average rejection throughput time	Median rejection throughput time
1	7 days	3 days
2	12 days	8 days
3	24 days	18 days
4	16 days	12 days

The analysis that can be done is then limited, due to the difficulties of working with samples of drastically different sizes, pointing us out the suggestion of the importance of taking special attention to the growth of the number of rejections that represents outliers in the average rejection throughput which can be crucial to understand the overall performance of the rejection process.

3.3 Business Owner Conclusions

The technical analysis done in response to the question was centered around the rejection flow of the declarations, which can be related to any process which has possible rejection flows. In this case, the emphasis was placed in the main rejection flow and the variations that exists, covering the majority of the rejections cases present in the dataset.

The most important aspect considered was the number of subsequent rejection flows that occurs in the same case, and the impact that it has in the performance of the rejection flow throughput, considering average throughput times as the main KPI. As expected, the average throughput time increases nearly linearly when increasing the number of subsequent flows. The distribution of these times was also took into consideration in order to try to understand the results and gain insight on how to improve the process, with mixed results: the number of cases of some variations of the main rejection flows varies tremendously, making it difficult to draw definitive conclusions, even if the presence of outlier cases can be hypothesized as the main cause.

The main recommendation is to always consider an analysis of the rejections flows (if they exist in the process) when reviewing its overall performance, existing multiple particularities in each process that can be recognized in order to gain insight in the improving opportunities that it provides, which in the case analyzed are a rework of the subsequent rejection process, which must be tack-

led in order to decrease the overall throughput times for the cases where the declaration is rejected multiple times.

4 Question 3: Which variables have a significant influence on the average throughput time of the Requests For Payment?

4.1 Analysis Method

Data used The dataset used was the `RequestForPayment.csv` [1] data, which included the steps each case took in order to request a reimbursement of the non-travel related costs of the trip. In particular, we considered the column "Case-Id" as the Case ID parameter, "time:timestamp" as the Timestamp parameter, "concept:name" as the Activity name and "sorting" as an auxiliary sorting key for the activities inside each case. There was no need for any other kind of previous pre-processing or change on the dataset besides the manual selection of the previous columns.

Data transformation Since this objective has a focus on lowering the throughput time of a request, we started removing from the data all the cases that were not ultimately successful (did not have the "Payment Handled" activity on their flow), since these rejected cases could have a negative impact on the throughput time. Based on a preliminary analysis of the information found on the remaining raw data, we selected the following five variables as candidates for individual analysis: Unexpected process activities and connections, Case project, Organizational Entity, Requested amount and Number of Rejections in a case.

In order to have a baseline regarding the throughput time and main process flow, we used the "Process Explorer" tool [3] which built the average process flow the data followed and then calculated the average throughput time of the whole process, measured in days, of the complete dataset. This way, we could know if the performance of our ideal combination would be higher or lower than the average.

As a further measure to isolate the influence of each variable we built on the platform an analysis sheet for each of the variables. All of the analysis themselves followed a similar pattern, where we first built an OLAP Table [5] that used the currently studied variable as a dimension and then the throughput time as a KPI, alongside other KPI's that were chosen based on the picked variable. Then, using that table, we built either a line chart or a marker chart that measured the Throughput Time against the studied variable. The selection of the type of chart was based on whether the variable was numerical or categorical. Once this process was performed each variable would be judged as relevant or not for the objective of lowering the overall throughput time and the ideal range of values was proposed.

Finally, after each variable was judged as relevant or not, we proceeded to group the relevant variables together in order to obtain the optimal combination

for minimizing the throughput time. Using again the tools that Celonis provides, we applied a filter to the complete dataset in order to see if there were cases that used the optimal combination in their variables, and then check if our hypothesis was correct once compared with the original average time values.

4.2 Question-Driven Technical Analysis Results

Complete dataset flow and average throughput time Using all the retrieved cases from the dataset and filtering only the successful ones, we found that the average case flow matched with the challenge description [1]: the employee submitted the Request for Payment, the administration and supervisor approved the Request, the employee requested the payment and the payment was handled appropriately. This means that the average flow ends with a successful request, with no rejections or need for approval of any other organisms in between. Another baseline KPI of interest for this objective is the average throughput time of all of the cases, which is **12.38 days**.

Unexpected activities and connections analysis Comparing the dataset to the challenge's proposed flow, we found that there was an extra activity on the data, which was "Request For Payment REJECTED by MISSING". This activity occurs **21** times on the dataset and corresponds to a rejection whose actor was not properly logged and as such is an expected deviation from the norm. Further analysis shows that this activity always follows the "Request For Payment FINAL_APPROVED by SUPERVISOR" activities and occurs on average once per case.

Another anomaly is the occurrence of the connections of "Request For Payment REJECTED by EMPLOYEE" - "Request For Payment SUBMITTED by EMPLOYEE". This means that after the Employee was rejected his request, they resubmitted it once again to restart the process. The connection occurs in **9%** of the cases, taking in average **7** days, and directly contradicts the proposed flow [1], where it is pointed out that if the employee rejects their own request, the case should end.

In order to see if the inclusion of these two variants affected the total time, we calculated the average throughput time while including and excluding them. There was **1** case on the data that included both anomalies with a throughput time of **13** days, **521** cases that only included the "Request For Payment REJECTED by MISSING" activity with an average time of **33.76** days, **542** cases that only included the "Request For Payment REJECTED by EMPLOYEE" - "Request For Payment SUBMITTED by EMPLOYEE" connection with an average time of **24.99** days and **6,307** cases that included none of the anomalies with an average time of 10.77 days.

Considering these values, we conclude that the inclusion or exclusion of these activities is relevant to the overall average throughput time of this process, being the optimal condition that both of these anomalies are excluded.

Amount requested analysis In average, the requested amount of each case is **268.700** money units. All of the cases studied, **81%** requested under the average amount, and **19%** requested over or the average amount. The group that requested under the average had an average throughput time of **11.93** days which falls slightly under the original average time, while the other group had an average throughput time of **14.45** days, two days over the original time.

In Figure 6, the correlation between the amount requested and the throughput time is mapped onto a line chart in order to find a potential regularity. The figure does not show a distinct slope as the requested amount increases. Instead, the throughput time is spread rather randomly over the chart. Taking into account this figure, alongside the previous KPI's of the groups over and under the average requested amount, we can conclude that the requested amount is not a relevant variable regarding to the throughput time of a case.

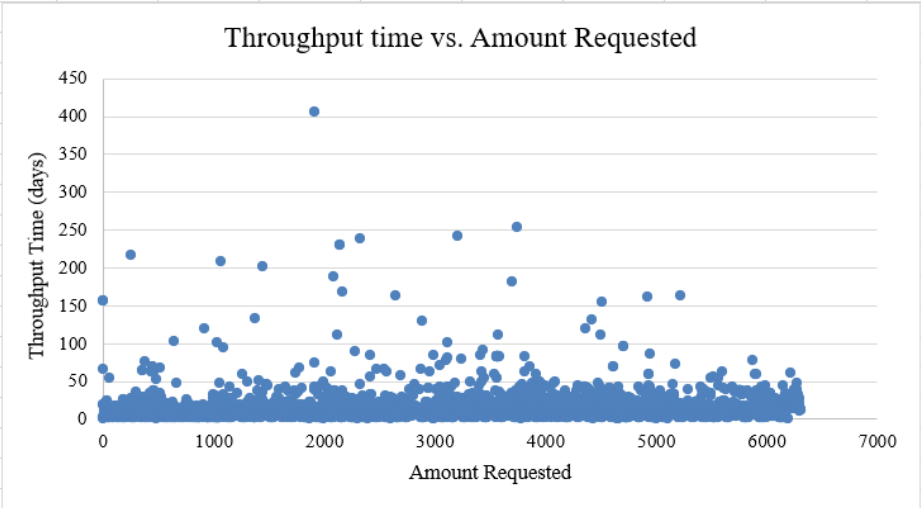

Fig. 6. Marker chart of Throughput time according to the requested amount on each case.

Organizational entity analysis In order to find the relevance of "Organizational Entity" on the throughput time, and possibly calculate the optimal entity, we used a table with that variable as the dimension, and as a KPI the average throughput time and the case count. There are **36** different organizational units. Once again, we filtered out the units that have less than **1%** of the total cases (**63** cases) which leaves **15** distinct organizational units. The unit with the longest average throughput time is "organizational unit 65482" with **253** cases and an average of **27.36** days, while the unit with the shortest average throughput time was "organizational unit 65463", with **902** cases and an average throughput time of **9.33** days.

Figure 7 shows the relation between the throughput time and the organizational unit. Since organizational units can have a significant difference on average throughput time (up to seven days, not including extreme outliers), we will also consider this variable as relevant. In specific, we will consider as optimal of the entities "organizational unit 65458", "organizational unit 65463" and "organizational unit 65469", because they have an average throughput time of under the data average.

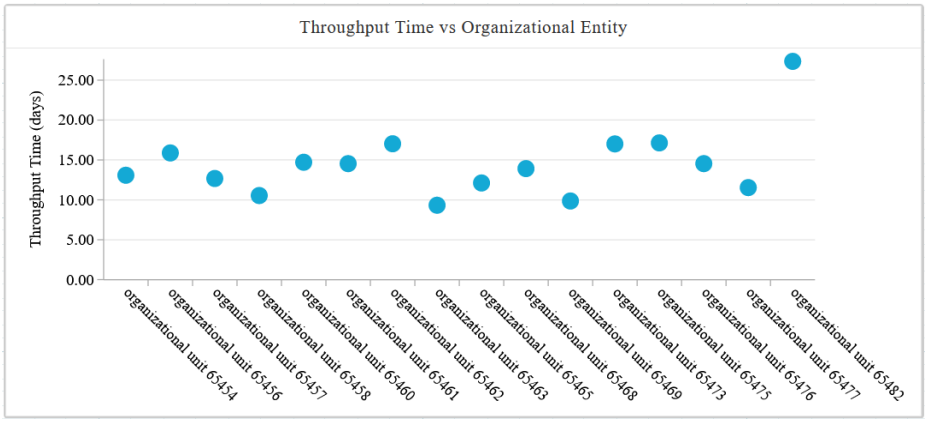

Fig. 7. Marker chart of Throughput time according to the Organizational Entity.

Case project analysis There are currently **78** different case projects on the data, which makes it an interesting variable to test. An initial analysis shows that "Project 147546" is the case project with the greater amount of cases at **1,036** cases and an average throughput time of **12.58** days. The second case project with most cases is "Project 503" with **982** cases and an average throughput time of **16.02** days. If we rank the case projects by throughput time, we get that both the case project with higher and lower throughput time ("project 163434" with **46** days and "project 163222" with **2** days) have only one case count. Since case projects with lower case count are not truly representative of the full sample, a filter is applied to the table to exclude case projects that have case counts lower than approximately **1%** of the total cases (63 cases).

Figure 8 and 9 show the distribution of the case counts according to the case project variable and the average throughput time of each case project. We can see that the case project variable has indeed a significant influence in the average throughput time (of 3 or 4 days in some cases). We define the optimal case projects as the ones with an average throughput time lower than **12.38 days** (the average throughput time of all cases on the dataset) and a case count higher than

1% of the total cases (63 cases). Using both figures as references, we can point out that optimal cases have a case project belonging to one of the following seven project cases: “project 147531”, “project 147572”, “project 147843”, “project 147908”, “project 148052”, “project 148681” or “project 152803”.

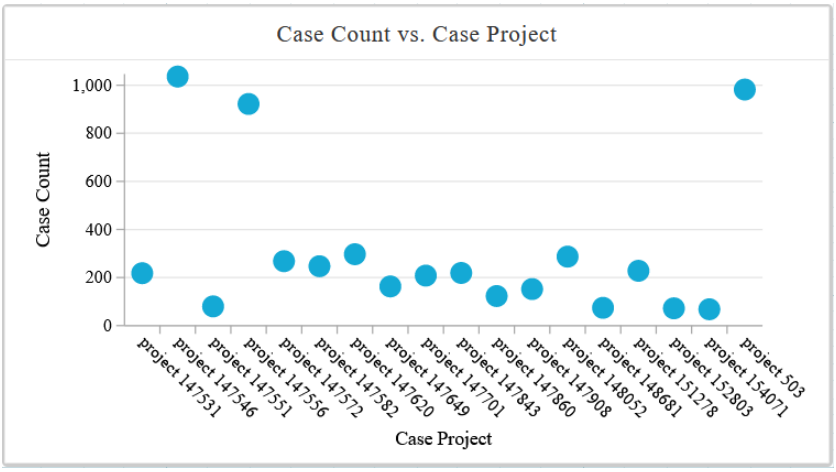

Fig. 8. Marker chart of the Case count according to the Case project.

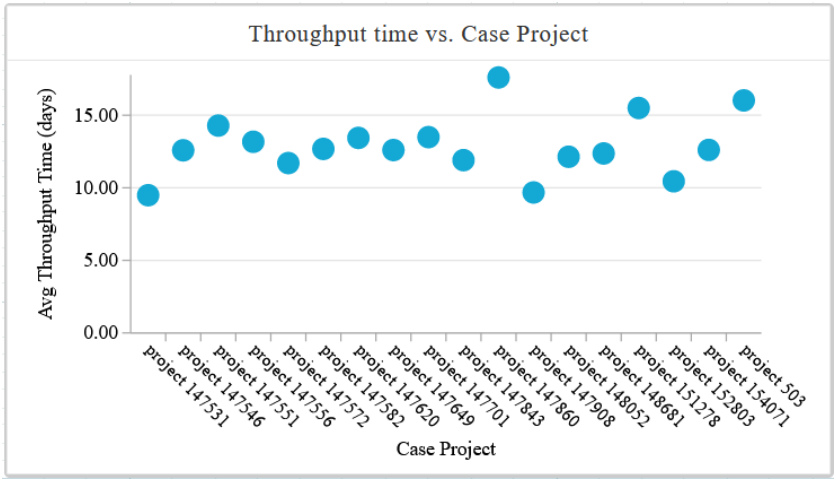

Fig. 9. Marker chart of the total Throughput time according to the Case project.

Number of rejections analysis We built a table with"CASE-ID" as the dimension and as KPI's the columns "Times Rejected", a count of each time the case flowed through an activity that was of the type "REJECTED" and throughput time of the case. An initial analysis shows that the minimum amount of rejections was **0**, which covers **5.737** cases or the **83%** of the total. The maximum amount of rejections was of **10** times, and occurs in only one case ("CASE-ID" 5526). The average amount of rejections per case was of **0.197** times.

We built a line chart in order to see the correlation between the rejections and the average time, pictured in Figure 10. Although the chart does not present a continuous positive slope, the throughput time tends to increase as the amount of rejection. Therefore, we conclude that the amount of rejections is a relevant variable regarding the throughput time of a case, with an optimal value of 0 rejections per case.

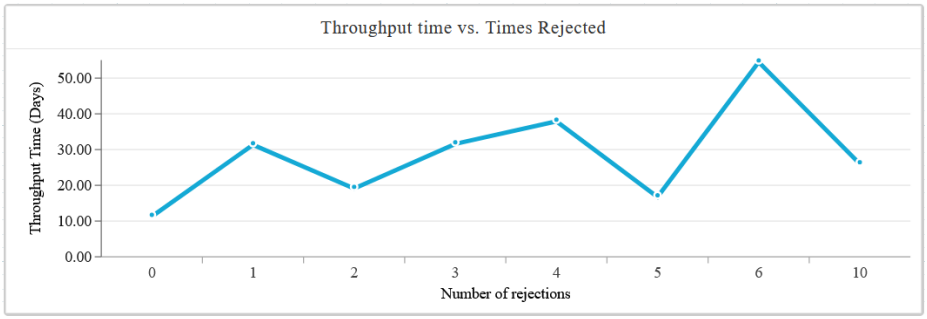

Fig. 10. Line chart of Throughput time according to the amount of rejections on each case.

Optimal combination analysis Once all the variables were analyzed, we took into account the relevant ones in order to produce an optimal combination that could help us reduce the average throughput time.

Regarding unexpected activities and connections on the process flow, we have chosen to exclude them altogether. The amount requested did not proved to be irrelevant to the total time, so there are no filters for that variable. We have chosen to only include the organizational entities "organizational unit 65458", "organizational unit 65463" and "organizational unit 65469" and case projects "project 147531", "project 147572", "project 147843","project 147908", "project 148052", "project 148681", "project 152803". Finally, our last filter are the cases with 0 previous rejections.

This set of filters fits to **315** of the original cases, which corresponds to a **4.99%** of the total successful requests. Finally, the average throughput time of

this set is of **8.10** days, which is lower than the original average by at least 4 days, a significant amount of time. As such, we can consider the analysis of variables a success.

4.3 Business Owner Conclusions

The conclusions of most relevance for this analysis are the validation of the variables "Unexpected activities and connections", "Organizational Entity", "Case Project" and "Number of rejections" as influential for the overall throughput time of a successful Request for Payment while the variable "Amount Requested" doesn't seem to have an impact on this time. With this set of restrictions, we have managed to lower the average throughput time from **12.38** days to **8.10** days, which is a significant time period when it comes to such a relevant process, the reimbursement of non travel-related costs of a trip for Academic reasons.

We sense an opportunity for further analysis regarding the classified variables as important. For example, there is a need for an internal investigation regarding the appearance of a connection between activities that contradicts the proposed ideal flow. Or in the case of "Organizational Entity", we propose an analysis on how the organizational units selected as optimal manage the travel processes of their staff, in comparison to the units with longest throughput time. The same can be applied to "Case project", in order to discover differences that could affect the throughput time between different projects.

Finally, we propose a further investigation on other variables that remained uninvestigated for this exact process flow, or take the opposing case and make an analysis of the throughput time of the ultimately unsuccessful cases of requests for payment, in order to compare the average time of a failed and a successful request flow.

5 Conclusions

After working on the analysis that could answer the three initial objectives, we obtained significant insights about the specific problems that are presented by the organization in relation to the travel declarations and permits for their academic employees, allowing us to develop multiple methodologies that helped us to better understand the process and obtain valuable insight that could be used to improve the performance of the process, considering the throughput time as the main KPI to consider during the analysis.

Comparison between domestic and international flights In this analysis, we identified the main activities and subsections of both process flows that had a significant impact on the overall throughput time. We discovered, that one of the key differences between domestic and international flights is the involvement of the director in the latter, extending the average time in both permit and declaration approval processes. Also, because the international travel process requires more control than the domestic flights, it's more prone to deviations and errors of

the applicant. There are even cases when there is a rejection (**10** cases) and the trip is executed anyways or when the trip is executed without any authorization (**748** cases). That means that even when those controls exist, there is not enough supervision on the correct execution of the steps of the process.

Rejections of declarations analysis The analysis of the rejections flows for both national and international declarations allowed us to better understand the different variants that occur in the process, while giving us insight on why some declarations have higher than average throughput times, considering the multiple successive rejections that can happen to the same declaration. The rejections flows encountered vary from **0** consecutive rejections to **4**, existing a nearly lineal growth in the overall throughput time of the entire rejection process, pointing out the importance of checking the different variants of the rejection flows in order to understand where the process can be improved.

Average Throughput Time of Requests for Payment Analysis For the last analysis, we tested five possible variables that seemed to have a significant influence on the average throughput time of a successful Request For Payment, in order to verify their actual impact on the throughput time and then, build an optimal combination that could help minimize the duration of the process. We concluded that the variables that did have a relevance on the duration were "Unexpected activities and connections", "Case Project", "Number of rejections", while the variable "Amount Requested" did not have a visible influence on the total case time. With these results, we proposed a combination of variables that would optimize the throughput time. This combination could be found on **4.99%** of the successful cases, and had in average a throughput time of **8.10** days, a value that indeed was lower than the average of all the successful cases.

Repeatability of the analysis The same analysis can be done for different types of processes and datasets, considering that the particularities of the process were only taken into consideration when obtaining the results of each analysis and building the conclusions based on the context of each objective. The different methodologies were defined at the beginning of each of the objectives, so they could be easily altered or extended depending on the dataset and process considered.

Acknowledgements

We thank Celonis Academic Alliance, specially to Jerome Geyer-Klingeberg, for giving us access to Celonis software and helping us with technical issues when was needed.

References

1. BPI Challenge Homepage, https://icpmconference.org/2020/bpi-challenge/. Last accessed 1 July 2020

2. SAP Process Mining by Celonis 4.2 Documentation (EN), https://support.celonis.de/pages/viewpage.action?pageId=5341382. Last accessed 1 July 2020
3. Process (Explorer) - SAP Process Mining by Celonis 4.2 Documentation (EN), https://support.celonis.de/pages/viewpage.action?pageId=5341302. Last accessed 1 July 2020
4. PI Conformance - SAP Process Mining by Celonis 4.2 Documentation (EN), https://support.celonis.de/display/C4E/PI+Conformance. Last accessed 1 July 2020.
5. OLAP Table - SAP Process Mining by Celonis 4.2 Documentation (EN), https://support.celonis.de/display/C4E/OLAP+Table. Last accessed 1 July 2020.
6. James L. Peterson.: Petri Nets. ACM Comput. Surv. 9, 3 (Sept. 1977), 223–252. DOI:https://doi.org/10.1145/356698.356702