Process Mining in Finance Sector: Trip Compensation Process in TU/e Business Process Intelligence Challenge 2020

Dmitrii Khodaev, Viktor Kalugin, Evgenia Korneeva, Anastasiya Balashova, Marina Savintseva

Sberbank of Russia, Moscow
{DSKhodaev, vskalugin, korneeva-ey, Balashova.An.Ni, , Savintseva.Ma.Li}@sberbank.ru

Abstract. In 2020, the BPI Challenge is held for the 10th anniversary, bringing together people from different countries searching for new tools and techniques in the field of process mining. This time, the participants were asked to analyze the process of travel reimbursement at the TU\e. All data in the provided files were anonymized or got noisy to prevent the spreading of personal data of university employees. The authors of this article aim to find the throughput capability of this process, compare the process graph obtained as a result of using process mining tools with the description that was provided, find various deviations from the planned process, as well as situations that might not have been foreseen by the authors of the process. In addition, during data analysis, preprocessing machine learning techniques were applied to find the dependency of the time spent on consideration of compensation declarations on any variables provided in datasets.

Keywords: BPIC 2020 · Process Mining · Process Discovery · Event Log · Data Analysis

1 Introduction

1.1 Introduction

Over the past few decades, humanity have made a great leap forward in the field of information technology. One of the consequences of the integration of computers into everyday life is the exponential growth of data about our world, about the operations that take place here every day. In particular, data about various processes are recorded. For example, your flight was delayed. You submit an application on the airline's website in order to receive compensation. Further, this application is processed according to the adopted strategy of the company and a decision on payment is made. However, there can be errors and deviations in this technological process. In order to find out information about the real state of affairs in the process and errors occurred, you can use the Process Mining tools. Today Process Mining is a dynamically developing method of data analysis. It gained its popularity among analysts and companies due to

its ease of use, high speed of processing logs and easy interpretation of results.

In 2020, the anniversary BPI Challenge is being held. The participants were provided with files containing information about compensation to employees of TU / e, the Netherlands, for business trips to conferences, meetings, etc.

The participants of the competition were asked to analyze the process of applying compensation declarations. In the course of the analysis, both Process Mining methods were used to find bottlenecks, as well as Data Analysis methods were used to find the dependences of the time for considering declarations on various provided features.

This article is divided into 3 main parts. The first part is an introduction and data description. The second part is data analysis. The third part is the conclusion and references.

1.2 Dataset exploration[1]

The data is split into travel permissions and several request types, particularly domestic declarations, international declarations, prepaid travel costs and requests for payment, where the latter refers to non-travel expenses (representation costs, hardware purchased for work, etc.).

The data is anonymized in such a way that no TU/e internal IDs are visible in the final dataset, i.e. all identifiers are freshly generated. Furthermore, the amounts mentioned in the data are not exact amounts. However, it is still possible to add declarations related to the same travel permission and then compare them to the original budget/estimate. Furthermore, with large enough samples of the data, the summed/averaged amounts should be roughly correct.

Staff members cannot be identified in the data. Instead, for all steps, the role of the person executed the step is recorded. The resource recorded in the data is either the SYSTEM, a STAFF MEMBER or UNKNOWN, or, on occasion, the data is MISSING.

The data is available for downloading on BPIC 2020 website.

- Requests for Payment (should not be travel related): 6,886 cases, 36,796 events: RequestForPayment.xes
- Domestic Declarations: 10,500 cases, 56,437 events: DomesticDeclarations.xes
- Prepaid Travel Cost: 2,099 cases, 18,246 events: PrepaidTravelCost.xes
- International Declarations: 6,449 cases, 72151 events: InternationalDeclarations.xes
- Travel Permits (including all related events of relevant prepaid travel cost declarations and travel declarations): 7,065 cases, 86,581 events: PermitLog.xes

1.3 The Process Flow[1]

The various declaration documents (domestic and international declarations, pre-paid

[1] This information is provided by BPIC2020 organizers

travel costs and requests for payment) all follow a similar process flow. After submission by the employee, the request is sent for approval to the travel administration. If approved, the request is then forwarded to the budget owner and after that to the supervisor. If the budget owner and supervisor are the same person, then only one of the these steps it taken. In some cases, the director also needs to approve the request.

In all cases, a rejection leads to one of two consequences. Either the employee resubmits the request, or the employee also rejects the request.

If the approval flow has a positive result, the payment is requested and made.

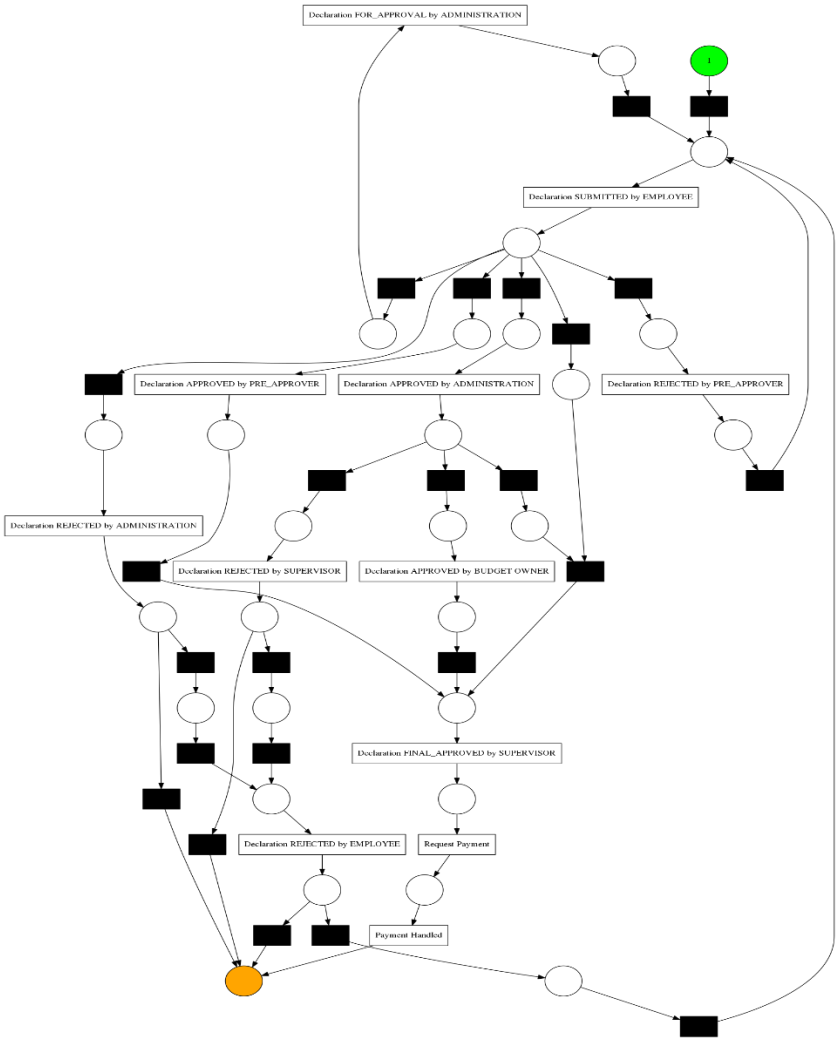
The travel permissions follow a slightly different flow as there is no payment involved. Instead, after all approval steps a trip can take place, indicated with an estimated start and end date. These dates are not exact travel dates, but rather estimated by the employee when the permit request is submitted. The actual travel dates are not recorded in the data, but should be close to the given dates in most cases.

After the end of a trip, an employee receives several reminders to submit a travel declaration.

After a travel permit is approved, but before the trip starts, employees can ask for a reimbursement of pre-paid travel costs. Several requests can be submitted independently of each other. After the trip ends, an international declaration can be submitted, although sometimes multiple declarations are seen for specific cases.

It’s important to realize that the process described above is the process for 2018. For 2017, there are some differences as this was a pilot year and the process changed slightly on several occasions.

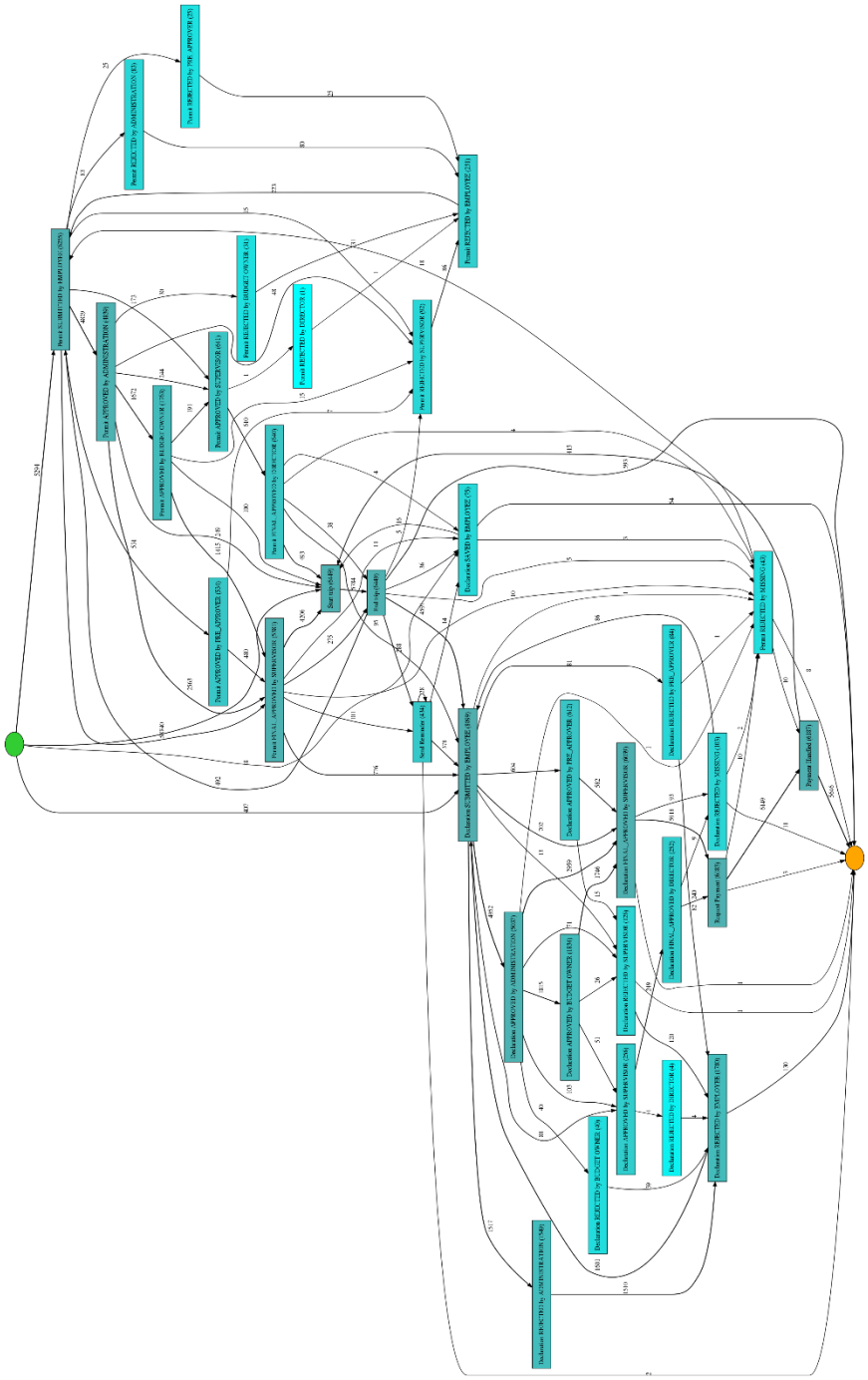
1.4 Process graph



Graph 1.4.1: Domestic Declarations Petri Net

A graph 1.4.1, that is presented above, built on the basis of the Domestic Declaration log without any filters. As you can see, this process has no critical deviations. The declaration always completes either after the “Payment Handled” status, or after one of the possible rejected-statuses.

5



Graph 1.4.2: International Declarations Heuristic Net

Graph 1.4.2. – graph without filters for International Declarations. In this case, deviations are visible, for example, there are 3 transitions from the "Request Payment" status to the "End" status. In addition to deviations, you can see that the graph consists of two parts: on the top right there are the Permits consideration statuses, in the left lower part of the graph there are the consideration statuses Declarations. Note that there is no strict division into two parts through the connecting transition through the statuses “Start trip” - “End Trip.” This means that, as described in the process description, Permit can be submitted both before and after the trip. Similar is for declarations.

2 Dataset Analysis

2.1 Travel declaration throughput

For internal declarations, the following figure was built:

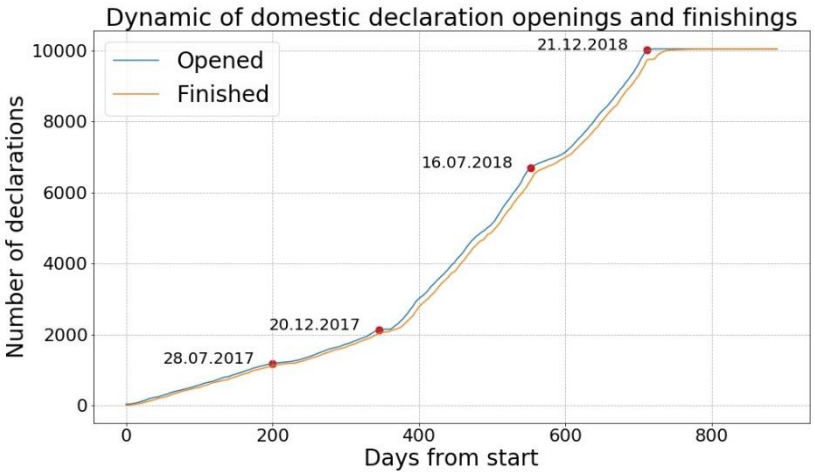


Fig. 2.1.1: Dynamics of openings and closings of internal declarations.

Features of the Figure 2.1.1:

1. There are 4 declines on the figure: approximately July 28, 2017, December 20, 2017, July 16, 2018 and December 21, 2018. December recessions are explained with the onset of the New Year and Christmas holidays, as well as, for 2018, with the end of the period under review. The July recessions are associated with a decrease in scientific activity in the second half of the summer and the vacation season at this time of the year.
2. At the border of 2017 and 2018, the growth rates of both lines significantly

changed. This is due to the fact that in 2017 the pilot of the project ended and its use began at full capacity.

3. The orange line lags slightly behind the blue line, which is clear: requests cannot be processed instantly. The mean time which the orange graph takes to reach the level of the blue graph will be called the **system response time**.
4. A final, but not insignificant feature of the figure is the line showing the number of closed orders has almost the same shape as the line of the number of opened orders. This means that the maximum system throughput has not been reached. Therefore, it can only be estimated by the maximum of the first derivative of the orange graph, which will be done later.

To find the system response time, the following technique was used: the orange graph was shifted to the left by 1 day until the standard deviation of the shifted graph from blue was minimized. The result of this method is shown in the graph below:

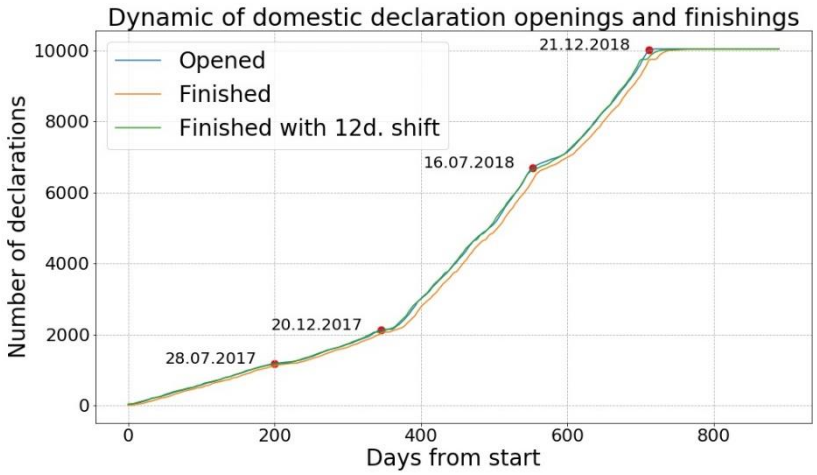


Fig. 2.1.2: Dynamics of openings and closings of internal declarations.

The green line is the closing line shifted 12 days to the left. The optimal shift was found - 12 days. Thus, compensation domestic declaration takes on average 12 days to go all the way from submission to payment.

Based on the Figure 2.1.2, the maximum throughput capability was not reached on either day. Examining the derivatives of these graphs, the achieved throughput was found - 177 declarations. This is the number of declarations received the final status "payment completed" on December 19, 2018. The maximum number of declarations was 69. This rate was reached on July 4, 2018. On average, 11.3 declarations were received per day in 2018 (weekends are also considered as days of declarations in this study). The same value is taken by the average number of declarations that have passed to the final status "payment completed".

Based on the results above, we can say that this system can cope with a large number

of declarations, the maximum throughput of the system was not reached, and on average, each declaration is processed for 12 days.

A similar analysis was made for international declarations. The graph of the dynamics of opening and closing of declarations for compensation is presented below. This graph only shows declarations that were ultimately approved and compensated as a result of their processing.

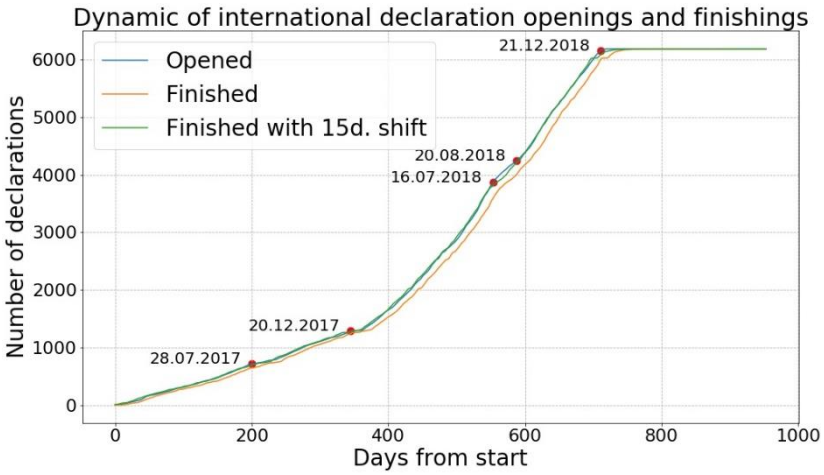


Fig. 2.1.3: Dynamics of openings and closings of international declarations.

Green line - the closing line shifted 15 days to the left. Here we can see similar behavior features as in the graphs of the dynamics of internal declarations. However, there is a significant difference - the optimal shift in the closing schedule is 15 days, not 12. We also note that compensations for international trips were paid 4000 less than for domestic ones (40% less).

Based on the Figure 2.1.3, the maximum throughput may not have been reached on either day. Examining the derivatives of these graphs, the achieved throughput of 117 declarations was found. This number of declarations received the final "Payment Handled" status on July 11, 2018. The maximum number of submitted declarations was 49. This figure was reached on June 24, 2018. On average, 6.5 declarations were received per day in 2018 (weekends are also considered declaration days in this study). The same value is taken by the average number of declarations received the final status "Payment Handled".

2.2 Domestic and International Throughput differences

With relation to a maximum throughput, it is impossible to talk about differences for international and domestic declarations, since it is not clear whether they have been achieved. However, in terms of achieved throughput, domestic declarations are processed in larger numbers (maximum and average achieved rate) and faster (3 days) than

international declarations.

Here in after, we call a chain of two consecutive events with one declaration as a **transition**.

2.3 Clusters of declarations differences

Clustering by projects, departments, requested amount, etc. In order to find out whether there is some division of claims in time depending on the input features[2], machine learning clustering methods can be used. However, first of all, it is better to visualize provided data. To do so, we have built pairwise dependencies of the features available in the data and the time spent on each declaration[3] (pair plots 2.3.1). In addition, a matrix (Matrix 2.3.2) of pairwise correlations of features and consideration time was calculated.

Considering all possible pairwise dependencies and the matrix of pairwise correlation coefficients (Matrix 2.3.2.), we can conclude that there is no simple, monotonic dependence between any two considered features, or a direct dependence of the time of declaration consideration on any either a sign. Pairwise correlation coefficients do not exceed 0.1.

To visualize provided data, it was decided to transform the feature space into a two-dimensional feature space (Figure 2.3.3.) using the T-SNE algorithm. As you can see from the Figure 2.3.3, all declarations in the feature space of a reduced dimension represent several dozen clusters. We assumed that several of these clusters would be grouped by declaration consideration time. At the same time, outliers can indicate problem declarations that were considered for several hundred hours or tens of days. However, in various clusters there were declarations with a short period of consideration, and several times longer. In the scatterplot below, all declarations have been divided into two classes: the declaration was processed in less than 30 days and more than 30 days (green and red, respectively). As can be seen from the graph, almost every cluster contains declarations that have been considered for more than 30 days.

[2] The following attributes were considered: organizational entity, requested amount, project number, start month, start weekday

[3] Declarations were selected by the status “Declaration Submitted by Employee” and the status “Payment Handled” among International Declarations for 2018. The time spent on the declaration was calculated as the difference in the times of these statuses in hours.

Pair plots 2.3.1: Pairwise dependences of signs and spent time. Left to right and top to bottom: org. entity, req. amount, project, start month, start weekday, spent time. The main diagonal contains histograms of the distribution

Matrix 2.3.2: Matrix of pairwise correlation coefficients of features. Left to right and top to bottom: org. entity, req. amount, project, start month, start weekday, spent hours.

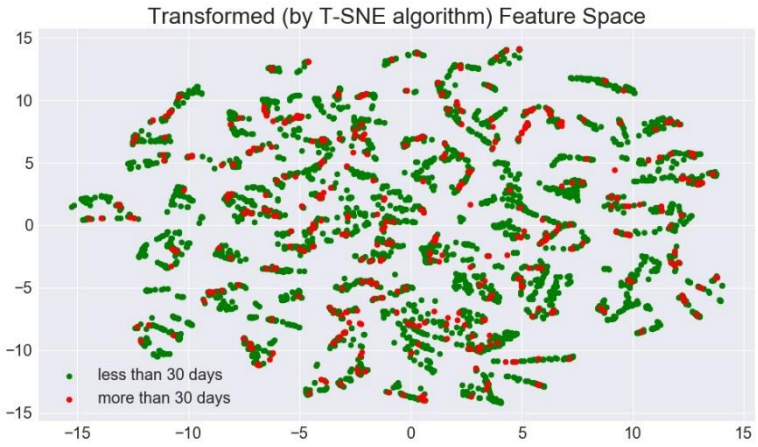


Fig. 2.3.3: Reduced dimensionality feature space. The red dots mark the declarations with more than 30-days consideration time. Green - less than 30 days.

Thus, none of the features under consideration, as well as a simple combination of these features, affect the declaration consideration time.

2.4 Process step throughput

At the beginning of this study, the throughput capability of the entire technological process was considered in terms of the mean transit time through the statuses. A similar technique cannot be applied to the statuses themselves, since the log does not contain information about the start time and end time of each status. In addition, it is not always known, for example, how many Budget owners or Supervisors are involved in this process. Therefore, it was decided to collect some statistics on the process, which are presented below separately for Domestic and International Declarations.

Table 2.4.1. Status statistics, International Declarations, 2018.

Status	Active Days	Total	Median	Mean	Max	Number of Max	Max Date
Declaration SUBMITTED by EMPLOYEE	354	6397	17	18,07	67	1	02.07.2018
Start trip	357	4952	11	13,87	293	1	10.04.2018
End trip	364	4952	10	13,6	317	1	13.04.2018
Declaration APPROVED by ADMINISTRATION	363	4875	12	13,43	48	1	02.07.2018
Permit SUBMITTED by EMPLOYEE	334	4858	14	14,54	41	2	22.05.2018
Permit APPROVED by ADMINISTRATION	335	4777	14	14,26	44	1	09.04.2018
Payment Handled	120	4741	41	39,51	117	1	12.07.2018
Request Payment	322	4737	12	14,71	62	1	18.10.2018
Declaration FINAL_APPROVED by SUPERVISOR	309	4596	13	14,87	69	1	09.07.2018
Permit FINAL_APPROVED by SUPERVISOR	274	4240	14	15,47	54	1	01.05.2018
Declaration APPROVED by BUDGET OWNER	255	1792	6	7,03	36	1	31.10.2018
Permit APPROVED by BUDGET OWNER	231	1739	6	7,53	27	3	09.04.2018
Declaration REJECTED by EMPLOYEE	233	1630	5	7	37	1	09.07.2018
Declaration REJECTED by ADMINISTRATION	299	1517	4	5,07	20	1	18.06.2018
Permit APPROVED by SUPERVISOR	166	447	2	2,69	15	1	01.02.2018
Permit FINAL_APPROVED by DIRECTOR	174	446	2	2,56	13	1	02.02.2018

Send Reminder	12	393	36	32,75	61	1	01.08.2018
Permit REJECTED by EMPLOYEE	107	169	1	1,58	4	4	08.05.2018
Declaration APPROVED by SUPERVISOR	101	146	1	1,45	3	6	05.09.2018
Declaration FINAL_APPROVED by DIRECTOR	104	145	1	1,39	4	3	18.12.2018
Declaration REJECTED by SUPERVISOR	70	94	1	1,34	4	2	14.08.2018
Permit REJECTED by ADMINISTRATION	67	79	1	1,18	3	2	22.06.2018
Permit REJECTED by SUPERVISOR	49	63	1	1,29	3	2	24.04.2018
Declaration SAVED by EMPLOYEE	47	58	1	1,23	4	1	06.11.2018
Declaration REJECTED by BUDGET OWNER	34	39	1	1,15	2	5	17.12.2018
Permit REJECTED by BUDGET OWNER	23	27	1	1,17	2	4	27.03.2018
Permit REJECTED by MISSING	1	1	1	1	1	1	21.02.2018
Declaration REJECTED by DIRECTOR	1	1	1	1	1	1	28.08.2018
Permit REJECTED by DIRECTOR	1	1	1	1	1	1	05.09.2018

Table 2.4.2. Status statistics, Domestic Declarations, 2018.

Status	Active Days	Total	Median	Mean	Max	Number of Max	Max Date
Declaration SUBMITTED by EMPLOYEE	355	8843	25	24,91	74	1	09.07.2018
Declaration APPROVED by ADMINISTRATION	369	8120	21	22,01	68	1	12.07.2018
Declaration FINAL_APPROVED by SUPERVISOR	322	7903	22	24,54	107	1	15.10.2018
Payment Handled	116	7903	70	68,13	177	1	20.12.2018
Request Payment	335	7896	17	23,57	110	1	18.12.2018
Declaration APPROVED by BUDGET OWNER	258	2807	10	10,88	44	1	17.12.2018
Declaration REJECTED by EMPLOYEE	209	928	4	4,44	53	1	10.07.2018
Declaration REJECTED by ADMINISTRATION	246	717	2	2,91	22	1	09.07.2018
Declaration REJECTED by SUPERVISOR	106	172	1	1,62	8	1	10.12.2018

Declaration REJECTED by BUDGET OWNER	34	45	1	1,32	8	1	08.02.2018
Declaration FOR_APPROVAL by ADMINISTRATION	1	1	1	1	1	1	01.05.2018

In tables 2.4.1, 2.4.2 the following parameters are indicated:

- Active days - the number of dates in 2018 - early 2019 on which the status was found
- Total - the total number of statuses in the log
- Median - the median number of statuses per day[4]
- Mean - the mean number of statuses per day
- Max - the maximum number of statuses per day
- Number of Max - how many times the maximum number of statuses per day has been reached
- Max Date - the date (one of the dates, if there are many of them) when the Number of Max (one of the Maxes) was reached

Based on the throughput of the technological process itself, as well as on the indicators in tables 2.4.1, 2.4.2, we can say that the system can handle the load with the current parameters. Similar to the overall throughput, our team believes the maximum throughput capability of the system remains unknown. In this case, we can characterize the throughput capability of each stage of the process as the median value in 2018.

2.5 Process bottlenecks

Domestic Declarations. To answer this question, all declarations from the Domestic Declarations log for 2018 were considered. The mean and median time for the transition of declarations between different statuses was determined:

Table 2.5.1. Mean and median transition times between statuses. Domestic Declarations, 2018.

Start status	End Status	Median time	Mean time	Number of transitions	Total time
Request Payment	Payment Handled	3,21	3,45	7896	27241,20
Declaration FINAL_APPROVED by SUPERVISOR	Request Payment	1,04	2,84	7896	22424,64
Declaration APPROVED by ADMINISTRATION	Declaration FINAL_APPROVED by SUPERVISOR	0,83	1,9	5119	9726,10
Declaration SUBMITTED by EMPLOYEE	Declaration APPROVED by ADMINISTRATION	0	1,1	8120	8932,00

[4] Day with 0 statuses are excluded from the study

Declaration APPROVED by BUDGET OWNER	Declaration FINAL_APPROVED by SUPERVISOR	1,88	2,86	2784	7962,24
Declaration APPROVED by ADMINISTRATION	Declaration APPROVED by BUDGET OWNER	0,83	1,9	2807	5333,30
Declaration REJECTED by EMPLOYEE	Declaration SUBMITTED by EMPLOYEE	0,17	4,43	928	4111,04
Declaration SUBMITTED by EMPLOYEE	Declaration REJECTED by ADMINISTRATION	0	4,22	717	3025,74
Declaration REJECTED by ADMINISTRATION	Declaration REJECTED by EMPLOYEE	0,92	1,98	709	1403,82
Declaration FINAL_APPROVED by SUPERVISOR	Payment Handled	48,33	81,58	7	571,06
Declaration REJECTED by SUPERVISOR	Declaration REJECTED by EMPLOYEE	1	2,83	170	481,10
Declaration APPROVED by ADMINISTRATION	Declaration REJECTED by SUPERVISOR	0,88	2,24	149	333,76
Declaration APPROVED by BUDGET OWNER	Declaration REJECTED by SUPERVISOR	4,04	4,24	23	97,52
Declaration REJECTED by BUDGET OWNER	Declaration REJECTED by EMPLOYEE	1,04	2,18	44	95,92
Declaration APPROVED by ADMINISTRATION	Declaration REJECTED by BUDGET OWNER	0,54	1,55	45	69,75
Declaration REJECTED by ADMINISTRATION	Declaration SUBMITTED by EMPLOYEE	0,88	0,83	8	6,64
Declaration REJECTED by BUDGET OWNER	Declaration SUBMITTED by EMPLOYEE	6,08	6,09	1	6,09
Declaration REJECTED by SUPERVISOR	Declaration SUBMITTED by EMPLOYEE	1,9	1,91	2	3,82
Declaration FOR_APPROVAL by ADMINISTRATION	Declaration SUBMITTED by EMPLOYEE	1	1,02	1	1,02
Declaration SUBMITTED by EMPLOYEE	Declaration FOR_APPROVAL by ADMINISTRATION	0	0,01	1	0,01
Declaration SUBMITTED by EMPLOYEE	Declaration REJECTED by EMPLOYEE	0	0	5	0,00

In total, 20 different types of transitions were found in the Domestic Declarations log for 2018. This table is sorted by the total time spent on all transitions (the rightmost column). The table shows that the most time is spent on making a payment during submitting domestic declarations. However, most likely it is impossible to reduce the payment time, since it is associated with banking procedures (not related to the business owner).

The next bottleneck in terms of total spent time is the "Declaration Final Approved by Supervisor" - "Request Payment" transition. In total, 7903 such transitions were recorded in 2018. 50% of these transitions were completed in less than 27 hours, which is a good indicator. However, as our study shows, more than 22% (1701) of such transitions are completed in more than 96 hours. In this regard, we decided to find out what the problem might be.

Distribution by day of the week. The first thing that was decided to investigate was the dependence of the median and mean time depending on the day of the week of the "Declaration Final Approved by Supervisor" event. Below is presented a group of the resulting graphs (Figure 2.5.2).

Note that for all such transitions, the median and mean times have a peak of 87 and 109 hours, respectively. On Friday - an increase of 170% and 27%, respectively, compared to Thursday, which is explained by the coming weekend. Confirmation of this fact is the decrease in the median time by about 22 hours per day on Saturday and Sunday and an overall decrease in the mean transition time by 40 hours over the weekend. At the same time, there is no noticeable peak on Friday for problem declarations. Growth on Friday compared to Thursday is 19 hours or 15% for the median time. Most problem declarations were submitted on Thursday - 40% of the total number of problem declarations.

Based on the dynamics of problem figures, the following conclusion can be drawn: the majority of declarations go into the "Request Payment" status on Tuesday. This fact is confirmed by Figure 2.5.3. Almost 40% (650) of problem declarations went into "Request Payment" status that day. At the same time, most of all transitions to this status for all declarations were made on Tuesday and Thursday (4837 or 61%). To solve this problem, it is necessary to more evenly distribute the transition to the "Request Payment" status on business days, since, for example, problem declarations confirmed by the supervisor on Monday will, on average, wait another 8 days before the transition to the "Request Payment" status.

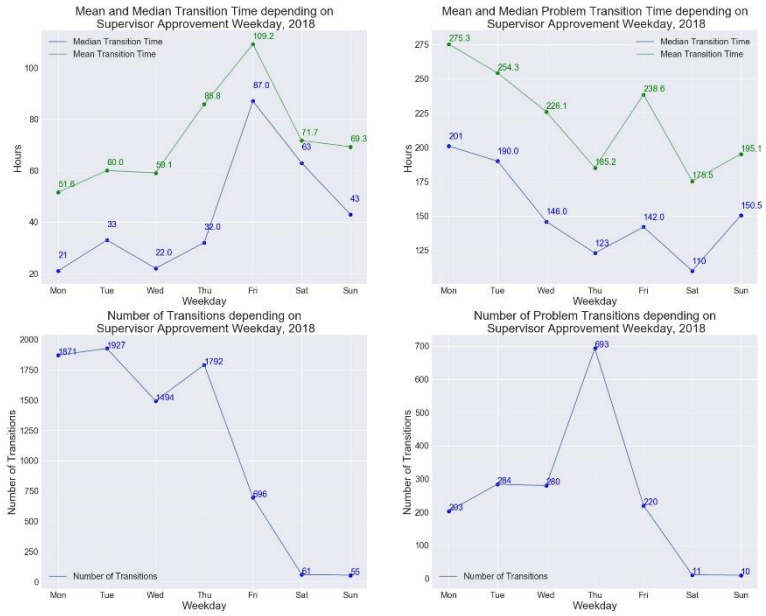


Fig. 2.5.2: Mean and median transition times for "Declaration Final Approved by Supervisor" - "Request Payment" depending on the day of the week for all declarations (left) and for declarations with a transition time of more than 96 hours (right), 2018

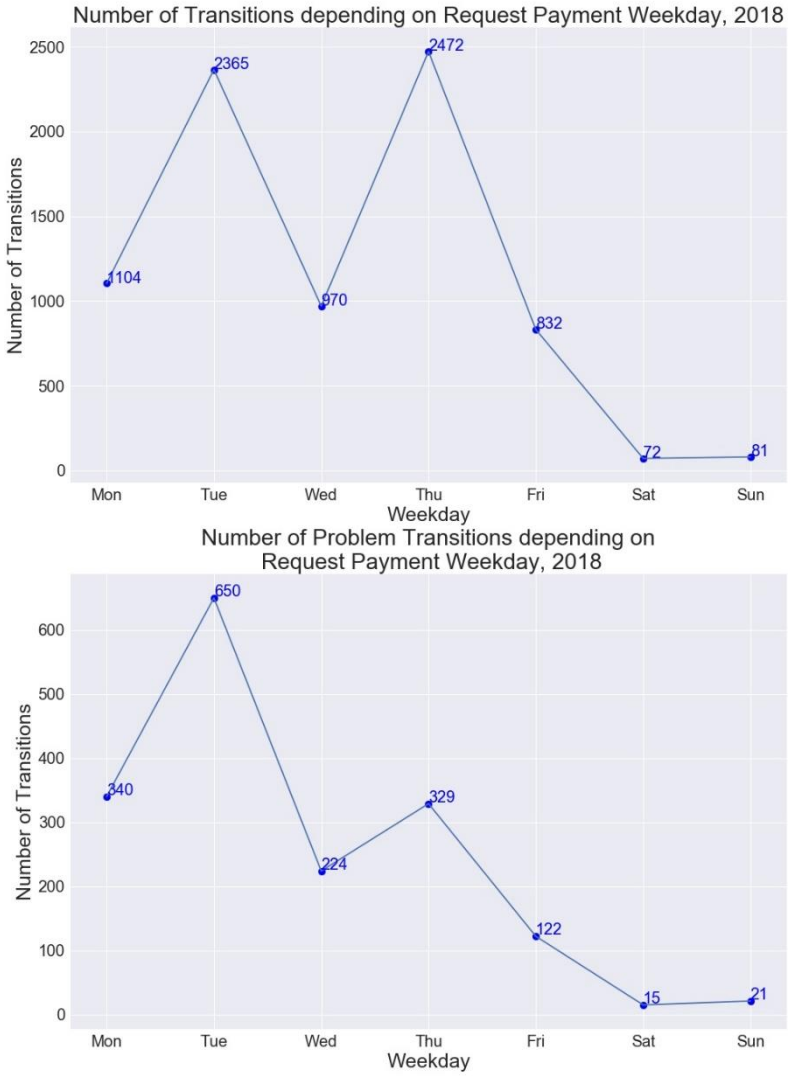


Fig. 2.5.3: The number of declarations that have passed to the "Request Payment" status by days of the week for all declarations (above) and for declarations that have a transition time of more than 96 hours (below), 2018

A similar analysis was also carried out in terms of distribution by month. The group of analysis graphs is presented below.

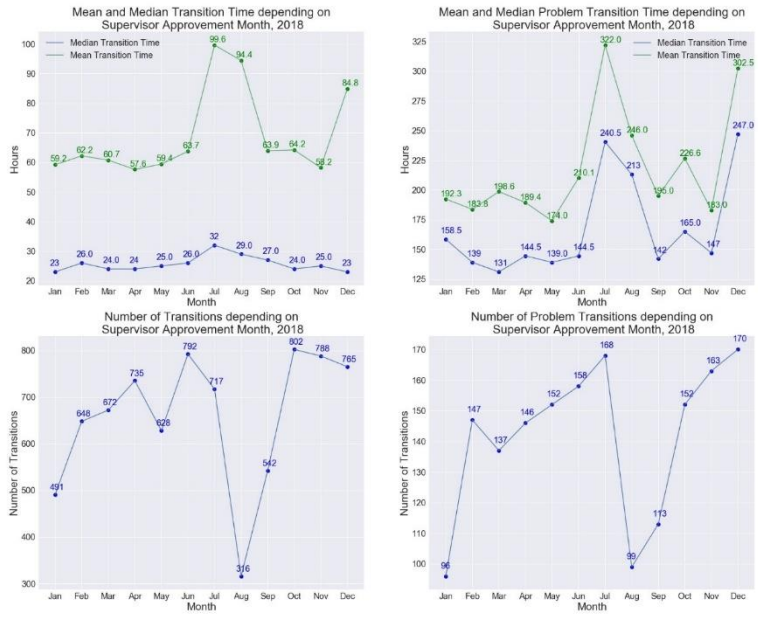


Fig. 2.5.4: Mean and median transition times of "Declaration Final Approved by Supervisor" - "Request Payment" depending on the month for all declarations (left) and for declarations with a transition time of more than 96 hours (right), 2018

As can be seen from Figure 2.5.4, the most problematic months in terms of median time both for problem declarations and for all, is July. This can be easily explained by the summer vacation period. In addition, the median time of all declarations and problem ones grows noticeably in December. Similar to July, this increase is most likely due to the New Year and Christmas holidays.

Dependence of the transition "Declaration Final Approved by Supervisor" - "Request Payment" on the requested amount.

In order to determine the presence or absence of this kind of dependence, the cost distribution functions were constructed for all declarations and for problem ones.

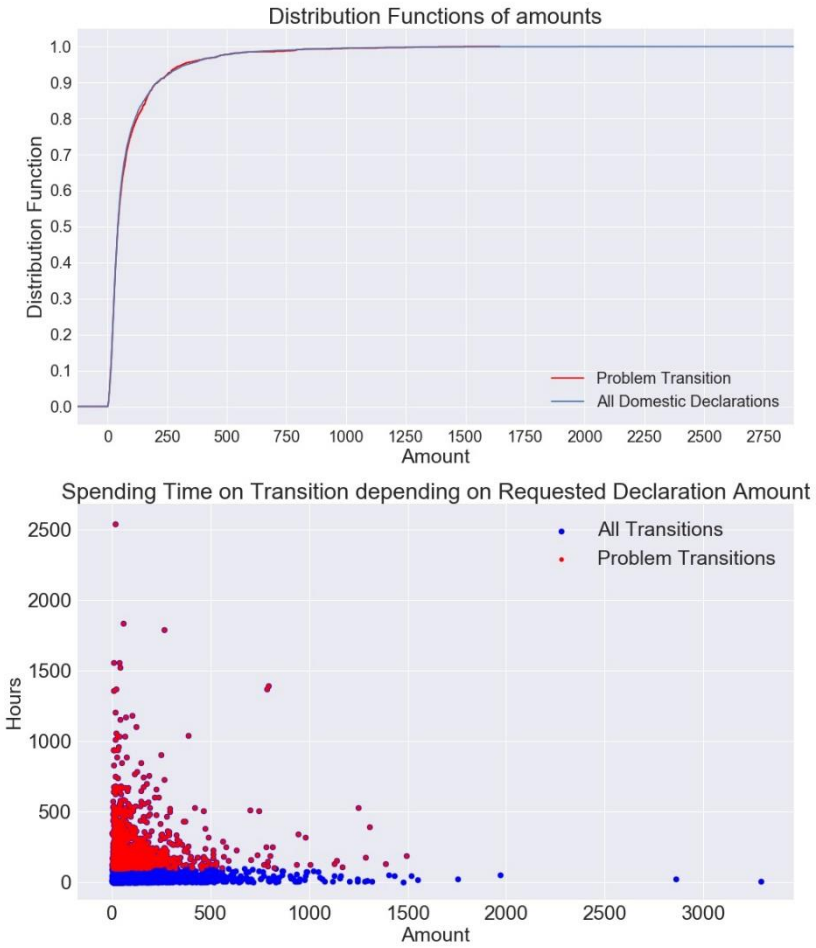


Fig. 2.5.5: Cost distribution functions for all declarations and for problem declarations (up). Dependence of the transition time "Declaration Final Approved by Supervisor" - "Request Payment" on the requested payment (lower). The red set of points is a subset of the blue set.

As you can see from the distribution functions, there is no explicit threshold, from which the declaration has a chance to become problematic. The requested amount distribution functions for all requests and for problem ones behave in the same way, except for small deviations. A similar conclusion can be made based on the dependence of the time spent on the transaction on the requested amount. Thus, there is no reason

to believe that an increase in the cost of an declaration leads to its long transition.
Other transitions Domestic Declarations. The rest of the transitions with long mean times (more than 3 days) can be divided into three classes:

1. The time of transition from one status to another depending on the employee.
2. The median transition time differs from the mean transition time by no more than 2.5 times. As well, the median time does not exceed 24 hours.
3. The number of transitions does not exceed 50, which is a small number compared to the number of declarations submitted (0.6% of the total).

Therefore, they were not considered in the same detail as the transition above.

Deviations in the workflow when requesting a payment. Domestic Declarations.

It was found that 7 requests do not have the "Request Payment" status, while the status is "Payment Handled". All information on these declarations is presented below. The most likely reason for the missing status is a write error. Note that for all declarations the time difference between the last two statuses "Declaration Final Approved by Supervisor" and "Payment Handled" is from 37 days to 9.5 months.

Table 2.5.6. Full log of declarations that received the "Payment Handled" status without passing the "Request Payment" status.

SN	id	action	datetime	Amount
0	115669	Declaration SUBMITTED by EMPLOYEE	2018-04-24 15:45:50	49.460878
1	115669	Declaration APPROVED by ADMINISTRATION	2018-04-24 15:46:05	49.460878
2	115669	Declaration FINAL_APPROVED by SUPERVISOR	2018-04-26 11:55:46	49.460878
3	115669	Payment Handled	2019-02-04 17:31:16	49.460878
4	124535	Declaration SUBMITTED by EMPLOYEE	2018-05-07 10:22:08	66.534149
5	124535	Declaration APPROVED by ADMINISTRATION	2018-05-07 10:30:58	66.534149
6	124535	Declaration REJECTED by SUPERVISOR	2018-05-07 11:08:51	66.534149
7	124535	Declaration SUBMITTED by EMPLOYEE	2018-05-08 13:01:19	66.534149
8	124535	Declaration APPROVED by ADMINISTRATION	2018-05-17 12:40:37	66.534149
9	124535	Declaration FINAL_APPROVED by SUPERVISOR	2018-05-17 14:00:46	66.534149
10	124535	Payment Handled	2018-07-12 17:31:18	66.534149
11	136996	Declaration SUBMITTED by EMPLOYEE	2018-12-12 13:02:00	5.593556
12	136996	Declaration APPROVED by ADMINISTRATION	2018-12-12 13:02:15	5.593556
13	136996	Declaration APPROVED by BUDGET OWNER	2018-12-17 10:17:59	5.593556
14	136996	Declaration FINAL_APPROVED by SUPERVISOR	2018-12-18 08:09:53	5.593556
15	136996	Payment Handled	2019-02-04 17:31:16	5.593556
16	138147	Declaration SUBMITTED by EMPLOYEE	2018-12-10 10:13:26	54.350829
17	138147	Declaration APPROVED by ADMINISTRATION	2018-12-10 10:31:55	54.350829
18	138147	Declaration APPROVED by BUDGET OWNER	2018-12-14 14:01:07	54.350829

19	138147	Declaration FINAL_APPROVED by SUPERVISOR	2018-12-18 08:44:05	54.350829
20	138147	Payment Handled	2019-02-04 17:31:16	54.350829
21	138710	Declaration SUBMITTED by EMPLOYEE	2018-12-17 11:44:51	241.706610
22	138710	Declaration APPROVED by ADMINISTRATION	2018-12-17 11:44:54	241.706610
23	138710	Declaration FINAL_APPROVED by SUPERVISOR	2018-12-18 11:32:18	241.706610
24	138710	Payment Handled	2019-02-04 17:31:16	241.706610
25	141310	Declaration SUBMITTED by EMPLOYEE	2018-12-18 09:25:39	44.514530
26	141310	Declaration APPROVED by ADMINISTRATION	2018-12-18 09:25:42	44.514530
27	141310	Declaration FINAL_APPROVED by SUPERVISOR	2018-12-18 10:10:24	44.514530
28	141310	Payment Handled	2019-02-04 17:31:16	44.514530
29	142992	Declaration SUBMITTED by EMPLOYEE	2018-12-17 14:21:55	14.010742
30	142992	Declaration APPROVED by ADMINISTRATION	2018-12-17 14:23:30	14.010742
31	142992	Declaration FINAL_APPROVED by SUPERVISOR	2018-12-18 08:42:13	14.010742
32	142992	Payment Handled	2019-01-24 17:31:39	14.010742

International Declarations

Table 2.5.7. Mean and median transition times between statuses. International Declarations, 2018.

Status before	Status after	Median time	Mean time	Number of Transitions	Total time
Request Payment	Payment Handled	3,21	3,46	6149	21275,54
Declaration FINAL_APPROVED by SUPERVISOR	Request Payment	1,17	3,21	5918	18996,78
Permit FINAL_APPROVED by SUPERVISOR	Declaration SUBMITTED by EMPLOYEE	6,17	16,56	776	12850,56
Declaration APPROVED by ADMINISTRATION	Declaration FINAL_APPROVED by SUPERVISOR	1,08	2,88	2959	8521,92
Permit FINAL_APPROVED by DIRECTOR	Declaration SUBMITTED by EMPLOYEE	26,29	76,04	95	7223,80
Declaration SUBMITTED by EMPLOYEE	Declaration APPROVED by ADMINISTRATION	0	1,4	4952	6932,80
Declaration APPROVED by ADMINISTRATION	Declaration APPROVED by BUDGET OWNER	1,04	3,12	1815	5662,80
Declaration APPROVED by	Declaration FINAL_APPROVED by	2	3,04	1746	5307,84

BUDGET OWNER	SUPERVISOR				
Declaration REJECTED by ADMINISTRATION	Declaration REJECTED by EMPLOYEE	1,79	3,43	1510	5179,30
Permit APPROVED by ADMINISTRATION	Permit FINAL_APPROVED by SUPERVISOR	0,88	1,92	2563	4920,96
Permit APPROVED by BUDGET OWNER	Permit FINAL_APPROVED by SUPERVISOR	2,04	3,1	1415	4386,50
Declaration REJECTED by EMPLOYEE	Declaration SUBMITTED by EMPLOYEE	0,08	2,58	1601	4130,58
Declaration SUBMITTED by EMPLOYEE	Declaration REJECTED by ADMINISTRATION	0	2,1	1517	3185,70
Permit APPROVED by ADMINISTRATION	Permit APPROVED by BUDGET OWNER	0,88	1,75	1672	2926,00
Permit SUBMITTED by EMPLOYEE	Permit APPROVED by ADMINISTRATION	0	0,39	4829	1883,31
Permit APPROVED by SUPERVISOR	Permit FINAL_APPROVED by DIRECTOR	0,92	2,61	610	1592,10
Declaration APPROVED by PRE_APPROVER	Declaration FINAL_APPROVED by SUPERVISOR	0,92	1,9	582	1105,80
Permit REJECTED by EMPLOYEE	Permit SUBMITTED by EMPLOYEE	0,46	4,21	223	938,83
Permit APPROVED by PRE_APPROVER	Permit FINAL_APPROVED by SUPERVISOR	0,92	1,91	480	916,80
Declaration SUBMITTED by EMPLOYEE	Declaration FINAL_APPROVED by SUPERVISOR	0	1,18	702	828,36
Declaration APPROVED by SUPERVISOR	Declaration FINAL_APPROVED by DIRECTOR	1,12	3,22	249	801,78
Declaration FINAL_APPROVED by DIRECTOR	Request Payment	1	2,81	240	674,40
Declaration SUBMITTED by EMPLOYEE	Declaration APPROVED by PRE_APPROVER	0	1,09	604	658,36
Declaration REJECTED by SUPERVISOR	Declaration REJECTED by EMPLOYEE	1,94	5,09	120	610,80
Permit APPROVED by BUDGET OWNER	Permit APPROVED by SUPERVISOR	2,38	3,19	191	609,29
Declaration REJECTED by MISSING	Declaration SUBMITTED by EMPLOYEE	0,79	6,67	86	573,62
Permit APPROVED by ADMINISTRATION	Permit APPROVED by SUPERVISOR	1	1,96	244	478,24

Declaration FINAL_APPROVED by SUPERVISOR	Declaration REJECTED by MISSING	2,75	4,73	93	439,89
Declaration APPROVED by ADMINISTRATION	Declaration REJECTED by SUPERVISOR	2,75	4,35	71	308,85
Declaration APPROVED by ADMINISTRATION	Declaration APPROVED by SUPERVISOR	1,08	2,95	103	303,85
Permit REJECTED by SUPERVISOR	Permit REJECTED by EMPLOYEE	1,23	2,93	86	251,98
Permit SUBMITTED by EMPLOYEE	Permit FINAL_APPROVED by SUPERVISOR	0	0,34	581	197,54
Declaration SUBMITTED by EMPLOYEE	Declaration APPROVED by SUPERVISOR	0	1,84	88	161,92
Permit SUBMITTED by EMPLOYEE	Permit APPROVED by SUPERVISOR	0	0,9	173	155,70
Declaration APPROVED by BUDGET OWNER	Declaration APPROVED by SUPERVISOR	1,96	2,8	51	142,80
Permit REJECTED by ADMINISTRATION	Permit REJECTED by EMPLOYEE	0,73	1,6	80	128,00
Declaration REJECTED by PRE_APPROVER	Declaration REJECTED by EMPLOYEE	0,77	1,24	82	101,68
Declaration SUBMITTED by EMPLOYEE	Declaration REJECTED by PRE_APPROVER	0	1,07	81	86,67
Permit SUBMITTED by EMPLOYEE	Permit APPROVED by PRE_APPROVER	0	0,06	531	31,86
Permit SUBMITTED by EMPLOYEE	Permit REJECTED by ADMINISTRATION	0	0,02	83	1,66

The table above shows a part of transitions (transitions have been removed, one of the events of which was “Start trip”, “End Trip” or “Send Reminder”).

Note that the top-2 transitions of International Declarations are absolutely identical to the top-2 transitions in total time for Domestic Declarations. Comments on the “Request Payment” - “Payment Handled” transition are similar to those on Domestic Declarations.

For the transition “Declaration Final Approved by Supervisor” - “Request Payment”, a similar analysis was carried out. Results are presented below.

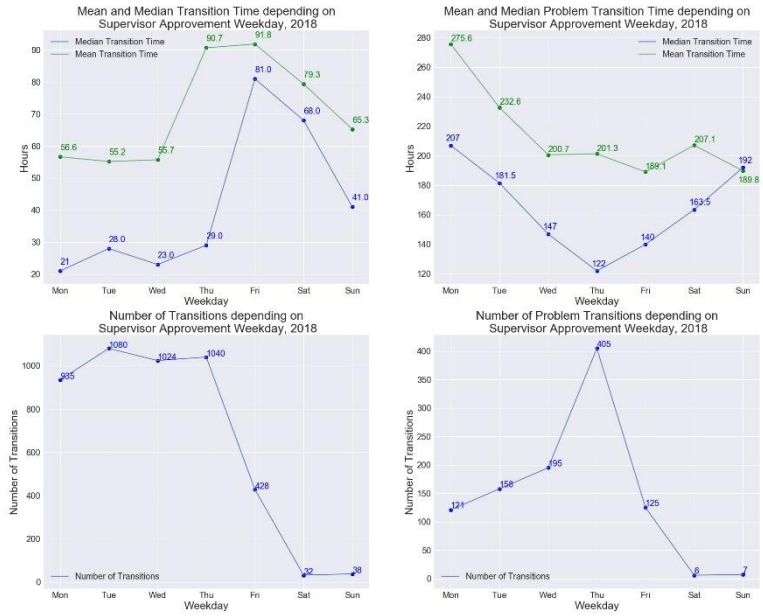


Fig. 2.5.8: Mean and median transition times for "Declaration Final Approved by Supervisor" - "Request Payment" depending on the day of the week for all declarations (left) and for declarations with a transition time of more than 96 hours (right), 2018

For International Declarations, as well as for Domestic Declarations, there is a distinctive peak in the number of problem declarations on Thursday. At the same time, from Thursday to Sunday, there is a monotonous increase in the median transition time, which was not typical for Domestic Declarations transitions.

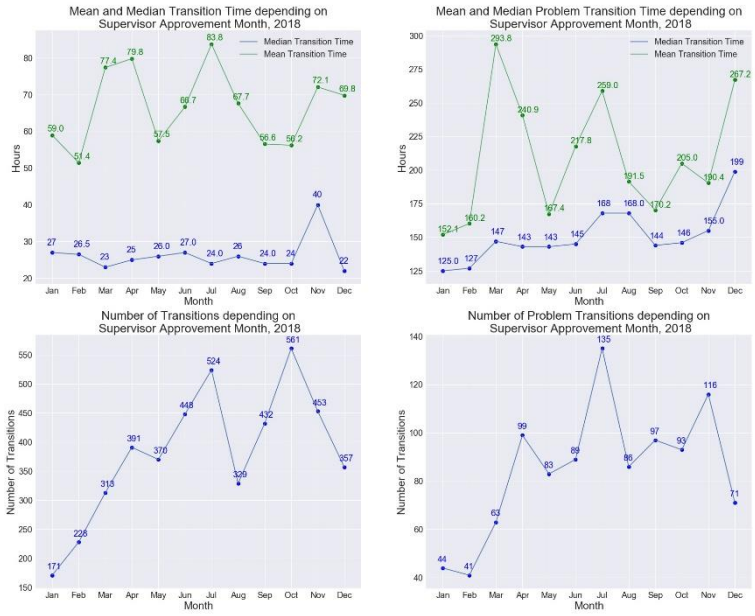


Fig. 2.5.9: Mean and median transition times for "Declaration Final Approved by "Supervisor" - "Request Payment" depending on the month for all declarations (left) and for declarations with a transition time of more than 96 hours (right), 2018

Comparing the distribution of the median time by month, a distinctive feature is revealed for International Declarations: the peak of the median time in November (an increase of 67% compared to October). Otherwise, the median transition time for all declarations is about 1 day. This was not the case with Domestic Declarations.

Considering problem declarations, it is clear that the behavior of the median transition time for International Declarations and Domestic Declarations is the same. Particularly, its growth is observed in the second half of summer and in December 2018. In addition, there is a strong increase in March (84% compared to February) in the mean time of problem transitions. Since the median time remains at about the same level in March, we can say that in March there was a serious failure for some declarations.

All transitions in March were investigated, the time of which was more than 96 hours. During the examination, it turned out that the long transition time had nothing to do with the project, but three organizational units had the most problems: 17 out of 63 (27%) transitions from 'organizational unit 65460', 11 out of 63 (17%) transitions from 'organizational unit 65454' and 9 out of 63 (14%) are on 'organizational unit 65458'. Thus, the 3 organizational unit accounts for over 58% of problem requests. The remaining 42% of problem transitions are distributed among other 9 organizational units.

The median request amount of problem declarations in March was 410 currency

units, the median request amount of all declarations was 515 currency units. As can be seen from these statistics, the delay in processing in March is hardly related to the large amounts requested. In fact, for International Declarations, the same dependence of the transition time on the request amount is observed, as for Domestic Declarations: the larger the request amount is, the faster the transition will be made (Figures 2.5.5 and 2.5.10).

Figure 2.5.10. (see below) shows functions of distribution of the requested amount of problem requests and all requests and the dependence of the transition time “Declaration Final Approved by Supervisor” - “Request Payment”. As can be seen from the upper figure (Fig. 2.5.), the distribution functions do not differ radically. In particular, the distribution function of problem declarations does not have hard jumps, which could be interpreted as a certain threshold that is responsible for increasing the processing time of declarations. Thus, as in the case of Domestic Declarations, it cannot be claimed that the requested amount negatively affects the time for declaration consideration.

In general, all other transitions, not related to Employee, have good statistics: 50% of all transitions are completed in 36-48 hours (Table 2.5.7.).

With regard to the Permit, we have not found any problem areas, since, as mentioned above, almost all other transitions were performed along the median in less than 2 days.

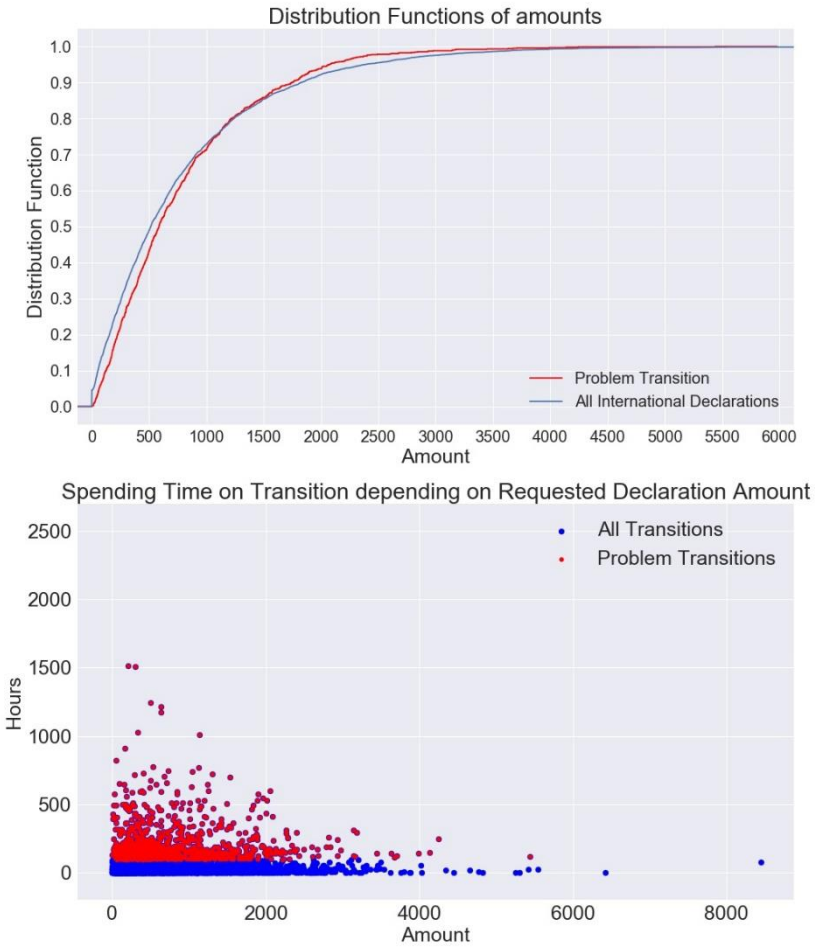


Fig. 2.5.10: Cost distribution functions for all declarations and for problem declarations (up). Dependence of the transition time "Declaration Final Approved by Supervisor" - "Request Payment" on the requested payment (lower). The red set of points is a subset of the blue set.

2.6 Rejected Declarations

Domestic Declarations. In total, 10500 unique Domestic Declarations were created in

2017-2018 (2240 in 2017 and 8260 in 2018). A total of 1301 (12.39%) Domestic Declarations were rejected at various stages. Payment was not eventually received in 456 cases. Compared to the pilot, the number of Domestic Declarations that were rejected at various stages increased by 2.76 percentage points in 2018. The "Declaration Rejected by Employee" status depends on other rejected statuses. Most often, Domestic Declarations both in 2017 and 2018 are rejected by administration. Below are summary tables for the entire period, and separately for the pilot and for 2018.

Table 2.6.1. Summary table of rejected statuses, Domestic Declarations 2017-2018.

Status	Cases	Unique ID's	Percentage (UniqueID's/number of all ID's)
Declaration REJECTED by EMPLOYEE	1365	1212	11,54
Declaration REJECTED by ADMINISTRATION	952	846	8,06
Declaration REJECTED by SUPERVISOR	293	281	2,68
Declaration REJECTED by MISSING	91	87	0,83
Declaration REJECTED by PRE_APPROVER	86	81	0,77
Declaration REJECTED by BUDGET OWNER	59	58	0,55
Total	**2846**	**1301**	**12,39**
With no payment	**456**	**456**	**4,34**

Table 2.6.2. Summary table of rejected statuses, Domestic Declarations 2017, pilot.

Status	Cases	Unique ID's	Percentage (UniqueID's/number of all ID's)
Declaration REJECTED by EMPLOYEE	161	149	6,65
Declaration REJECTED by MISSING	91	87	3,88
Declaration REJECTED by PRE_APPROVER	86	81	3,62
Declaration REJECTED by SUPERVISOR	71	64	2,86
Declaration REJECTED by ADMINISTRATION	4	4	0,18
Total	**413**	**229**	**10,22**
With no payment	**99**	**99**	**4,42**

Table 2.6.3. Summary table of rejected statuses, Domestic Declarations 2018.

Status	Cases	Unique ID's	Percentage (UniqueID's/number of all ID's)
Declaration REJECTED by EMPLOYEE	1204	1063	12,87
Declaration REJECTED by ADMINISTRATION	948	842	10,19
Declaration REJECTED by SUPERVISOR	222	217	2,63

Declaration REJECTED by BUDGET OWNER	59	58	0,7
Total	**2433**	**1072**	**12,98**
With no payment	**357**	**357**	**4,32**

International Declarations. A total of 6449 unique International Declarations were created for 2017-2018 (1341 in 2017 and 5108 in 2018). A total of 1576 (24.44%) International Declarations were rejected at various stages. Payment was not eventually received in 262 (4.06%) cases. Compared to 2018, the number of International Declarations rejected by Administration has grown sharply both in percentage and in absolute terms. In 2017, there was only 1 declaration rejected by administration (0.07%). In 2018, there were 1286 such declarations (over 25%). Below are summary tables for rejected statuses for the entire period and separately for the pilot and 2018.

Table 2.6.4. Summary table of rejected statuses, International Declarations 2017-2018.

Status	Cases	Unique ID's	Percentage (UniqueID's/number of all ID's)
Declaration REJECTED by EMPLOYEE	1780	1483	23
Declaration REJECTED by ADMINISTRATION	1549	1287	19,96
Declaration REJECTED by SUPERVISOR	126	122	1,89
Declaration REJECTED by MISSING	103	98	1,52
Declaration REJECTED by PRE_APPROVER	84	82	1,27
Declaration REJECTED by BUDGET OWNER	40	40	0,62
Declaration REJECTED by DIRECTOR	4	4	0,06
Total	**3686**	**1576**	**24,44**
With no payment	**262**	**262**	**4,06**

Table 2.6.5. Summary table of rejected statuses, International Declarations 2017, pilot.

Status	Cases	Unique ID's	Percentage (UniqueID's/number of all ID's)
Declaration REJECTED by EMPLOYEE	115	109	8,13
Declaration REJECTED by MISSING	103	98	7,31
Declaration REJECTED by PRE_APPROVER	84	82	6,11
Declaration REJECTED by SUPERVISOR	29	28	2,09
Declaration REJECTED by DIRECTOR	2	2	0,15
Declaration REJECTED by ADMINISTRATION	1	1	0,07
Total	**334**	**195**	**14,54**
With no payment	**44**	**44**	**3,28**

Table 2.6.6. Summary table of rejected statuses, International Declarations 2018.

Status	Cases	Unique ID's	Percentage (UniqueID's/number of all ID's)
Declaration REJECTED by EMPLOYEE	1665	1374	26,9
Declaration REJECTED by ADMINISTRATION	1548	1286	25,18
Declaration REJECTED by SUPERVISOR	97	94	1,84
Declaration REJECTED by BUDGET OWNER	40	40	0,78
Declaration REJECTED by DIRECTOR	2	2	0,04
Total	3352	1381	27,04
With no payment	218	218	4,27

3 Conclusion and References

3.1 Conclusion

Process Mining is one of the best process analysis tools. It can detect bottlenecks, deviations and useful statistics about the process, including the transition times between statuses and their number. ML preprocessing tools can be used for detecting a dependency after process analysis with Process Mining.

The travel compensation process was analyzed using the pm4py python library, and preprocessing tools from the sklearn python library were used. The most problematic areas of the process were payment and the transition from "Declaration Final Approved by Supervisor" to "Request Payment" status. An attention must be paid to these two lengthy operations during process optimization.

The mean time for Domestic Declarations is 12 days, for International Declarations - 15 days. During the study, it was found that there is no simple correlation between the processing time of the declarations and the features presented in the log's. In addition, no complex dependence on the features proposed by the authors of the article was found. Clusters obtained as a result of decreasing the dimension of the feature space are indistinguishable in time.

In general, the constructed graphs of the filing process for Domestic Declarations and International Declarations correspond to the declared model of the process behavior. There are several deviations for International Declarations, however, it is possible that the system may have crashed. More data is required to find the problem.

3.2 References

1. W. van der Aalst. Process Mining – Data Science in Action. Springer, 2016.
2. BPI 2020 website https://icpmconference.org/2020/bpi-challenge/.