BPI Challenge 2020 Report: Analyzing International and Domestic Travel Processes

Nikolayuk Alexander, Sdvizhkova Evgeniya, Khabarova Yulia, Shtokolova Anna
Tarasov Yuri

[1] Central Chernozem Bank PJSC Sberbank
[2] January 9th Street, Building 28, Voronezh, 394000, Russia
Nikolayuk.A.A@sberbank.ru

Abstract. Business Process Intelligence Challenge (BPI Challenge) allows to test the mining skills based on process data from real business operations.
In 2020, the competition is aimed at analyzing the processes of departure and registration of documents for the payment of employees of the company of the Eindhoven University of Technology (TU / e). Employees travel to conferences or other universities to meet with their colleagues. The unloading of traffic in the company processes of domestic and international travel were proposed for the analysis. Domestic trips did not require travel approval, unlike international ones. Data analysis was carried out in the Disco software. For the analysis, the constructed process graphs in Disco were used. The analysis identified 3 main processes:

1. submitting an internal declaration
2. submitting an international declaration
    a. obtaining a travel permission
    b. formation of an international declaration
3. payment for the trip
    a. requesting and receiving of prepayment
    b. request for payment for completed trips

Graph analysis showed the presence of bottlenecks in the above processes. The bottlenecks were found to be identical on both domestic and international travel. In addition, the facts of double payment of declarations were established.

Keywords: Process Mining, Disco.

1 Introduction

The 2020 Business Process Intelligence (BPI) Challenge organizers are asking all entrants to analyze data on TU / e's ride reimbursement process. Using the provided dataset in the form of logs: Domestic Declarations, International Declarations, Permit Log, Prepaid Travel Cost, Request For Payment, it is necessary to answer a number of questions, as defined in the problem description [1]. Answering questions will allow the owner to improve these processes. The direction of this research is the analysis of the constructed graphs, which can give a detailed idea of the inconsistencies that arise.

BPI's task is to analyze the data provided, answer questions of interest from the process owner, identify bottlenecks, establish the throughput capability of domestic and international declarations, and assess the effectiveness of all stages of the processes. Since the quality of the study of bottlenecks can be influenced by a set of data, they were first examined for informativeness.
The competition issues are expected to be resolved through Process Mining, which focuses on the relationship between business process models and event data. The main purpose of analysis is to find the correlated attributes of the event log data, which can provide insight into bottlenecks and find ways to resolve them. During the analysis of the data, correlations were not established, therefore, the main research tool was the expert analysis of logs for the registration of international and domestic business trips, as well as the graphs built on their basis.

2 Research and data preprocessing

In the data provided by TU / e, the process described refers to 2018, both for domestic and international travel.
Consequently, for the analysis of internal travel, not only the submitted and paid declarations in 2018 were selected by ID, but the declarations taking into account the transition of events on applications from 2017 to 2018, as well as taking into account the transition from 2018 to 2019. Thus, the time period for the analysis was: 02/15/2017 - 06/17/2019.

Table 1. Fields for Analysis by Domestic Travel

Field name	Field description	Quantity
Case ID	Employee's declaration number	8345
Activity	The stage of working with an application in the system of a specific user	46799
Resource	The identifier of the user who decides on the declaration	5
Complete Timestamp	Timestamp of the event	15.02.2017–17.06.2019

An analysis of the informativeness of the fields in the provided initial data showed that for further research, we need certain information from the fields of the unloading logs (Table 1).
For the analysis of international trips, not only submitted and paid declarations in 2018 were selected by ID, but also declarations taking into account the transition of events on applications from 2017 to 2018, from 2018 to 2019. Thus, the time period for the analysis was: 01/09/2017 - 08/19/2019.
An analysis of the information content of the fields in the provided data showed that for further research, we need certain information from the fields of log uploads (Table 2).

Table 2. Analysis fields for international travel

Field name	Field description	Quantity
case:id	Employee's declaration number	5500
concept:name	The stage of working with an application in the system of a specific user	62844
org:role	The identifier of the user who decides on the declaration	5
time:timestamp	Timestamp of the event	09.01.2017–19.08.2019

As part of data analysis, 3 main processes were identified:

1. Submission of an international declaration

   – filing an international declaration
   – obtaining a travel permission

2. Formation of an internal декларации
3. Payment for the trip

   – request and receiving prepayment
   – request for payment of completed trips

An important element of successful log analysis is the search for comparability in the unloading. Comparison of the original unloading of the TU / e company did not lead to the establishment of common columns by which it would be possible to combine files of different stages of the same process and dive into the more detailed analysis.
For example, the search for relationships in the process of issuing an international trip (files International Declarations, Permit Log) was carried out using the event id and the declaration number. As a result, no relationship analysis was identified. The budget item in the files International Declarations, Permit Log files also did not correlate (the budget number was different for all lines, the logic of assigning numbers was not established).
The search for relationships in the process of filling out the internal declaration was carried out by the event id and the declaration number. As a result, the analysis of the links was not revealed. The budget item was also not correlated (the budget number for all lines is identical and carries information for analysis).
Based on the above, the following unload files became the data source for the analysis:

– Domestic Declarations
– International Declarations

With the help of the Disco program, process graphs were built, visually demonstrating the throughput capability of each step, the number of cases returned to the previous step, and additional process steps deviating from the "ideal".

3 Analysis of process models

3.1 Domestic travel

Throughput capability is the maximum data transfer rate along a given path or a sequence of related actions aimed at achieving the set business objectives. This is an important indicator of the process functioning, which makes it possible to assess its efficiency and development zones.
The business process of formalizing an internal trip does not imply preliminary approval of the trip, but immediately begins with the registration of a declaration and subsequent receipt of reimbursement. At TU / e, like any other, the process is technologically structured, but during its operation so-called deviations from the “ideal model” occur.
It is logical to calculate the throughput according to the "ideal model", which will allow to quickly and accurately determine the "falling" steps. The throughput capability of the inner declaration is determined by the column Activity:

- step "Declaration submitted by the employee" (9241 cases)
- step "Payment is being processed" (7982 cases)

When building a graph, the main part of the declarations goes from the beginning of the process to its end in 5 steps. This is the “ideal process” of movement of the internal declaration from creation to payment (Fig. 1). Based on the "ideal process", it is possible to calculate the total throughput capability of the system from the creation of an internal declaration to its payment (looking at the "ideal model", it is equal to 86.4%).

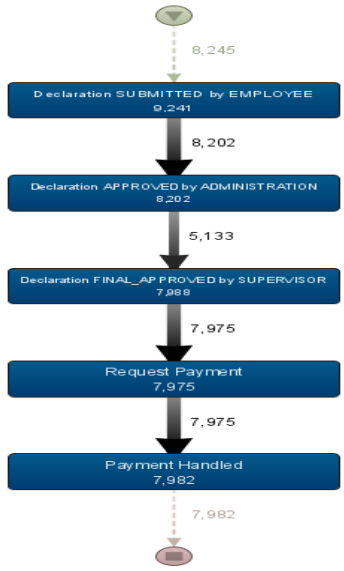

Fig. 1. Graph of "ideal process" movement of the internal declaration

Detailed understanding of the "falling steps" in the process system should not be limited to calculating only the total throughput value. The calculation of the capacity is required at each stage of the process of moving the internal declaration in the "ideal model". Using the graph (Fig. 1), the following values are obtained.

1. Declaration Approved by Administration (8202 events) / Declaration Approved by Employee (9241 events) = 88,77%
2. Declaration Final Approved by Supervisor (5133) / Declaration Approved by Administration (8202 events) = **62,58%** (at this step, there is a significant decrease in throughput)
3. Request Payment (7975) / Declaration Final Approved by Supervisor (7988) = 99,83%
4. Payment Handled (7975) / Request Payment (7975) = 100%
5. Paid (7975) / Payment Handled (7975) = 100%

The calculation showed that the most significant decrease in throughput is observed at the stage of transition of cases from the stage "Declaration Approved by Admin-istration" to the stage "Declaration Approved by Supervisor" and is 62.58%. In a detailed analysis of the reasons for the decrease in the throughput of the above step via the full process graph in Disco (Fig. 10), an intermediate step “Declaration Approved by Budget Owner” (Fig. 2) was identified, to which 35% of cases of the total the number of cases that left from the "Declara-tion Approved by Administration" step. To compare the processing speed of cases at the main stage in the "ideal model" and the intermediate stage in the "non-ideal model", it is necessary to determine the average value of the time spent on one case at each of the stages. Disco software calculates the time automatically. Time costs are shown in the list below.

- 65% cases went straight to the "Declaration Approved by Supervisor" step, where an average of 20.9 hours was spent on processing one case
- 35% cases we switched to the above step through additional approval from the Budget Owner, spending on average 66.6 hours per case (the sum of the time of entry and exit of cases at the stage is taken into account)

Thus, if we exclude the step of additional agreement with the Budget Owner (Fig. 2) from the process of registration of the internal declaration after agreement with the Administration, then it will be possible to reduce the time it takes for each case in the system to pass on average by 3.2 times.

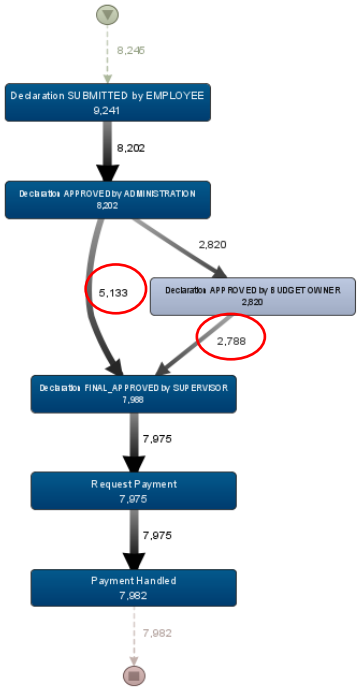

Fig. 2. Graph of the main deviation in the process of movement of the internal declaration

The movement of an internal declaration, like any document of strict reporting, involves stages of approval. Coordination is carried out by different resources - these are the roles of specific employees involved in the process. If there are errors in the documents, the cases cannot be reconciled and must be returned to the previous step, therefore the number of events is not equal to the number of cases. In addition, there are so-called "missing" employees who cannot complete the approval process and bring the case to a payment order and, as a result, receive this payment.

A detailed analysis of the number of rejected cases for internal declarations is presented in Table 3. At the same time, a total of 1,092 unique values were rejected for internal declarations.

Table 3. Statistics of deviations (returns) for internal declarations for the previous step of the process

Step	Number events	Number of cases
Application rejected by missing persons	6	6
Application rejected by pre-approved	4	4
The ad was rejected by the administration	952	846

Application rejected by an employee	1221	1078
Application rejected by supervisor	231	224
Declaration rejected by Budget owner	59	58

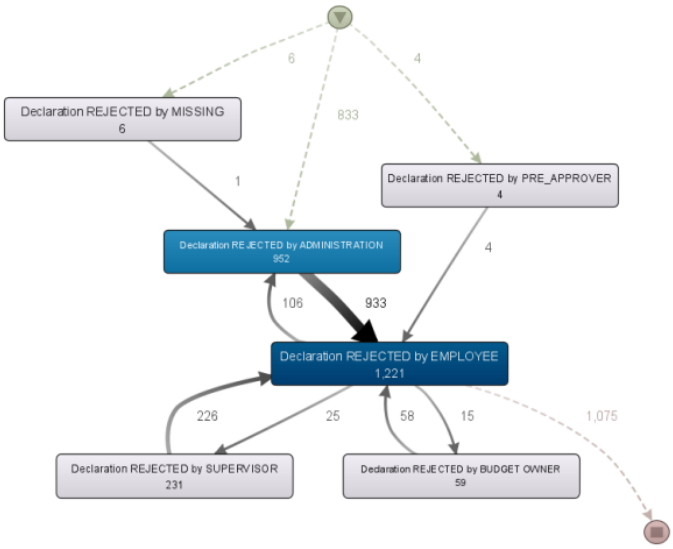

Fig. 3. Count of rejected cases for internal declarations

It is important to note that in the process there are declarations that will never be approved - these are the declarations of the so-called “missing” employees (Fig. 3).
At the stage of processing payments for internal declarations, facts of double payment were also established. In total, 7975 unique values (declarations) were received for payment, while 7982 unique values (declarations) were paid. This fact is confirmed by Fig. 1.
The well-built system of the process for internal declarations does not allow the authorized person to account for the traveler. For this reason, 100% of the declarations were submitted by the traveler (i.e. Employee), which confirms Fig. 4. Authorized persons are involved only in the approval / approval / rejection process. Thus, the responsibility for the formation of the declaration and receipt of the payment rests with the traveler.

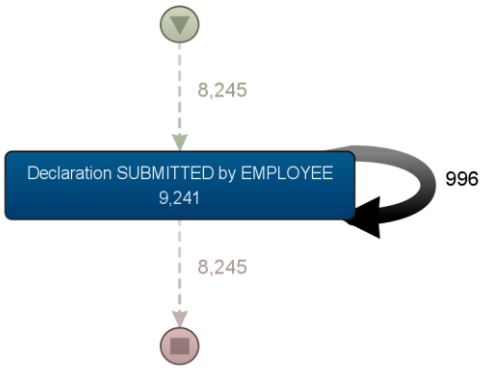

Fig. 4. Internal declaration graph

3.2 International travel

The business process for international travel formalization is different from domestic travel. A distinctive feature is that international business trips require preliminary approval of the departure, and only after receiving it, the declaration begins to be drawn up to receive prepayment before departure or compensation for money for an already completed departure. The calculation of the system capacity of the international travel process, by analogy with domestic travel, was carried out on the basis of the "ideal model" (Fig. 5), which made it possible to quickly and accurately determine the "lagging" steps, as well as compare the processes of work with internal ones. and international declarations. The throughput for international declarations is determined by the column Activity:

- step «Permit Submitted by Employee» (5250 cases)
- step «Payment Handled» (5239 cases)

When building a graph, the main part of the declarations goes from the beginning of the process to its end in 9 steps. This is the "ideal process" of the international declaration movement (Fig. 5). Based on the "ideal process", you can calculate the total throughput from the creation of an international declaration to its payment (if you look at the "ideal model", then it is equal to 99,79%).

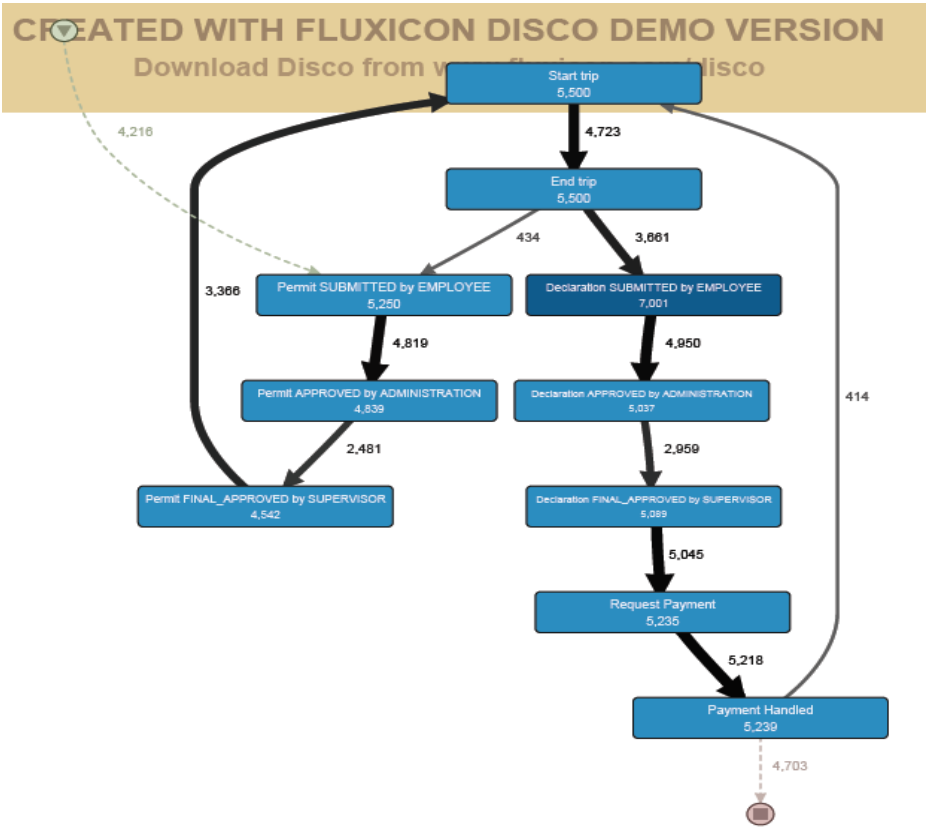

Fig. 5. Ideal process graph for international travel

Analysis of the throughput of the systems for the movement of the internal declaration (subsection 3.1) and external declarations showed a significant deviation in relation to the number of payments on them. The percent of unpaid internal declarations is 13.6%, while for international declarations the value was only 0.21%. It should be noted that the presence of an additional stage for international travel (we are talking about obtaining approval for the exit) did not negatively affect the throughput of the system. In addition, during the analysis of international trips, the presence of a significant number of so-called “missing persons” was established - 27 cases were rejected for this reason. As a reminder, for domestic travel, the number of “missing persons” was 6 unique cases.

Thus, the capacity losses for domestic declarations are significantly higher than the capacity losses for international declarations. This suggests that TU / e management should revise the domestic travel payments processing.

To understand the "lagging steps" in the system of the international travel process, calculating only the value of the total capacity should not be limited. It is required to calculate the throughput at each stage of the process in the "ideal model" (Fig. 5)

1. Permit Approved by Administration (4819)/Permit Submitted by Employee (5250) = 91,79%
2. Permit Final Approved by Supervisor (2481) / Permit Approved by Administration (4839) = **51,27%** At this step, there is a significant reduction in throughput
3. Start Trip (3366) / Permit Final Approved by Supervisor (4542) = 74,1%
4. End Trip (4723) / Start Trip (5500) = 85,87%
5. (Permit Submitted by Employee (434) + Declaration Submitted by Employee (3661))/ End Trip (5500) = 74,45%
6. Declaration Approved by Administration (4950) / Declaration Submitted by Employee (7001) = 70,7
7. Declaration Final Approved by Supervisor (2959) / Declaration Approved by Administration (5037) = **58,74%**
   At this step, there is a significant reduction in throughput
8. Request Payment (5045) / Declaration Final Approved by Supervisor (5089) = 99,13%
9. Payment Handled (5218) / Request Payment (5235) = 99,67%

The calculation showed that in the process of movement of international declarations there are two "bottlenecks", with a capacity of less than 60%.
When analyzing the reasons for the decrease in throughput from the processing stage "Permit Approved by Administration" to the stage "Permit Final Approved by Supervisor", an intermediate step "Declaration Approved by Budget Owner" (Fig. 6) was identified, to which 36% of cases of the total number were transferred. cases exited from the "Declaration Approved by Administration" step. To compare the processing speed of cases at the main and intermediate stages, the average value of the time spent on one case should be determined. Disco software calculates the time automatically. The data is listed below.

- 64,4% cases went straight to the “Declaration Approved by Supervisor” step, where an average of 21.2 hours was spent on processing 1 case
- 35,6% cases we switched to the above step through additional Budget Owner approval, spending an average of 69.5 hours per case

Thus, if you exclude from the process the step of additional coordination with the Budget Owner after coordination with the Administration, you can reduce the time for passing each case by an average of 3 times.

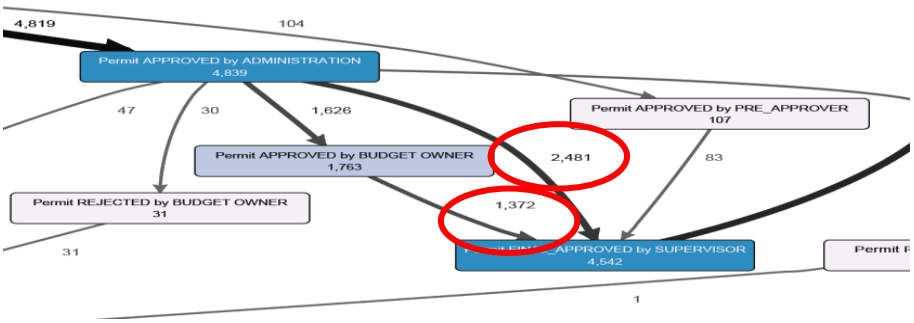

Fig. 6. Bottleneck at Step 2 in the International Declaration Movement Process

Another bottleneck is the transition of cases from the "Declaration Approved by Administration" processing stage to the "Declaration Final Approved by Supervisor" stage (Fig. 7), where the throughput is 58.74%. The average time spent on final Supervisor approval for a single case is 26.6 hours.

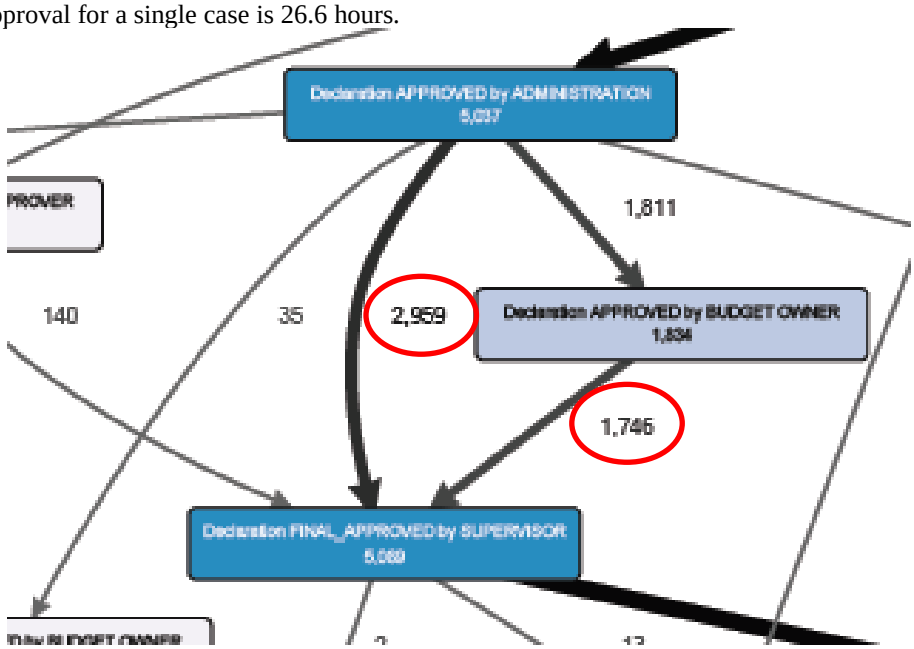

Fig. 7. Bottleneck at step 7 in the process of moving the international declaration

When analyzing the reasons for the decrease in throughput from the "Declaration Approved by Administration" processing stage to the "Declaration Approved by Su-pervisor" stage, an intermediate step "Declaration Approved by Budget Owner" was identified (Fig. 7), to which 37.1% of cases passed from the total number of cases that left from the "Declaration Approved by Administration" step. To compare the speed of processing cases at the main and intermediate stages, the average value of the time spent on 1 case should be determined. The Disco software calculates the time automatically. The data is listed below.

- 63% cases went straight to the step "Declaration Final Approved by Supervisor", where an average of 26.6 hours were spent on processing one case
- 37% cases we switched to the above step through additional Budget Owner approval, spending an average of 74 hours per case

Thus, if we exclude from the process the step of additional coordination with the Budget Owner after coordination with the Administration, it is possible to reduce the time for passing each case by an average of 2.8 times.

The movement of the international declaration supposes the coordination of cases, which are spent by different resources (these are the roles of specific employees involved in the process). If there are errors, the cases cannot be reconciled and must be returned to the previous step, therefore the number of events is not equal to the number of cases. In addition, there are so-called “missing” employees who cannot complete the approval process and bring the case to payment. A detailed analysis of the number of rejected cases is presented in Table 4. At the same time, a total of 1520 unique values were rejected according to international declarations.

Table 4. Statistics of deviations (returns) for international declarations for the previous step of the process

Step	Number events	Number of cases
Declaration Rejected by Pre_Approved	28	28
Declaration Rejected by Employee	1707	1414
Declaration Rejected by Missing	28	27
Declaration Rejected by Supervisor	109	105
Declaration Rejected by Administration	1549	1287
Declaration Rejected by Budget Owner	40	40
Declaration Rejected by Director	3	3

It is important to note that in the process there are declarations that will never be approved - these are declarations of the so-called "missing" employees (Fig. 8).

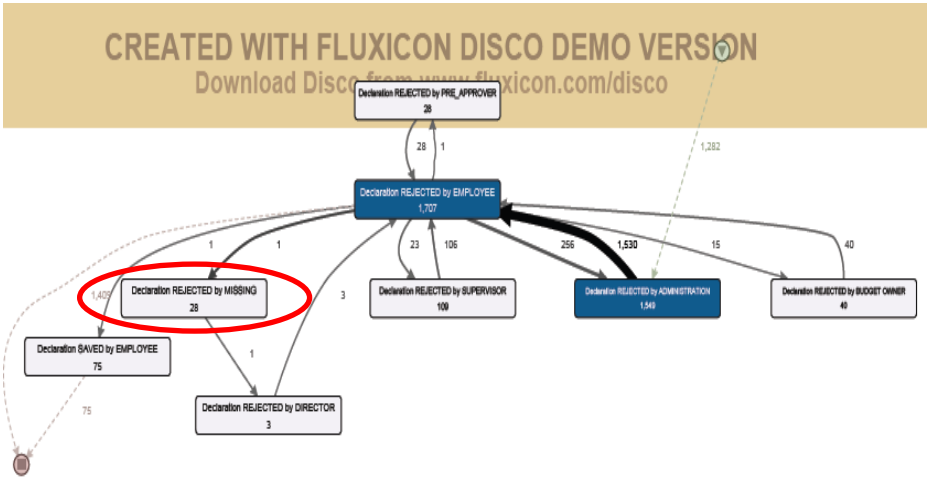

Fig. 8. Count of rejected cases for international declarations

Coordination of international business trips, in contrast to domestic ones, necessarily requires obtaining permission to leave. Analysis of the graph (Fig. 5) showed that all declarations passed the stages of approval for travel authorization. It turns out that receiving an advance payment or compensation for money spent is carried out on a legal basis and the process in this part does not require adjustment.

At the stage of processing payments for international declarations, no facts of double payment were established. In total, 5218 non-unique values (cases) were received for payment, while 4703 unique values (declarations) were paid. This fact confirms Fig. 5.

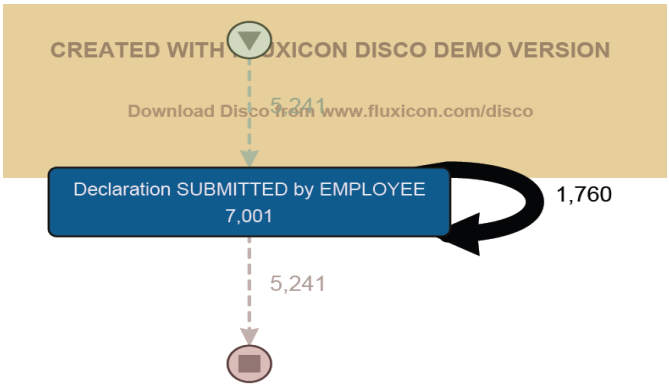

Fig. 9. Schedule for filing international declarations

It is important to note that, as in the case of internal declarations, the built-up system of the process for international declarations does not allow the authorized person to account for the traveler. For this reason, 100% of the declarations were submitted by the

traveler (i.e. Employee), Fig. 9. Authorized persons are involved only in the process of approval / approval / rejection. Thus, the responsibility for receiving the payment rests with the traveler..

4 Conclusion

The analysis compiled within the BPI competition can be grouped into three parts.
In the first stage of the analysis, three main processes were identified in the initial offloads of TU / e:

1. submitting an internal declaration
2. submission of an international declaration
   a. obtaining a travel permission
   b. formation of an international declaration
3. payment for the trip
   a. request and receive prepayment
   b. request for payment of completed trips

Data analysis was carried out in the Disco software. For the analysis, general graphs of the processes of movement of cases for international and domestic business trips, as well as graphs of the "ideal" process and graphs of "bottlenecks" were built.
At the second stage of the analysis, a sample of data was made, which became the basis for the analysis by time stamps.
At the third stage, an analysis of the graphs of the processes of registration and payment of domestic and international trips in general was carried out, as well as an analysis of bottlenecks. The bottlenecks were found to be identical in both domestic and international travel: the Budget Owner additional approval stage. This not only lowers the throughput of the system, but also increases the processing time of cases.
Thus, the results of the analysis can help the owner of the process in its understanding and scope of changes for improvement.

5 Appendix

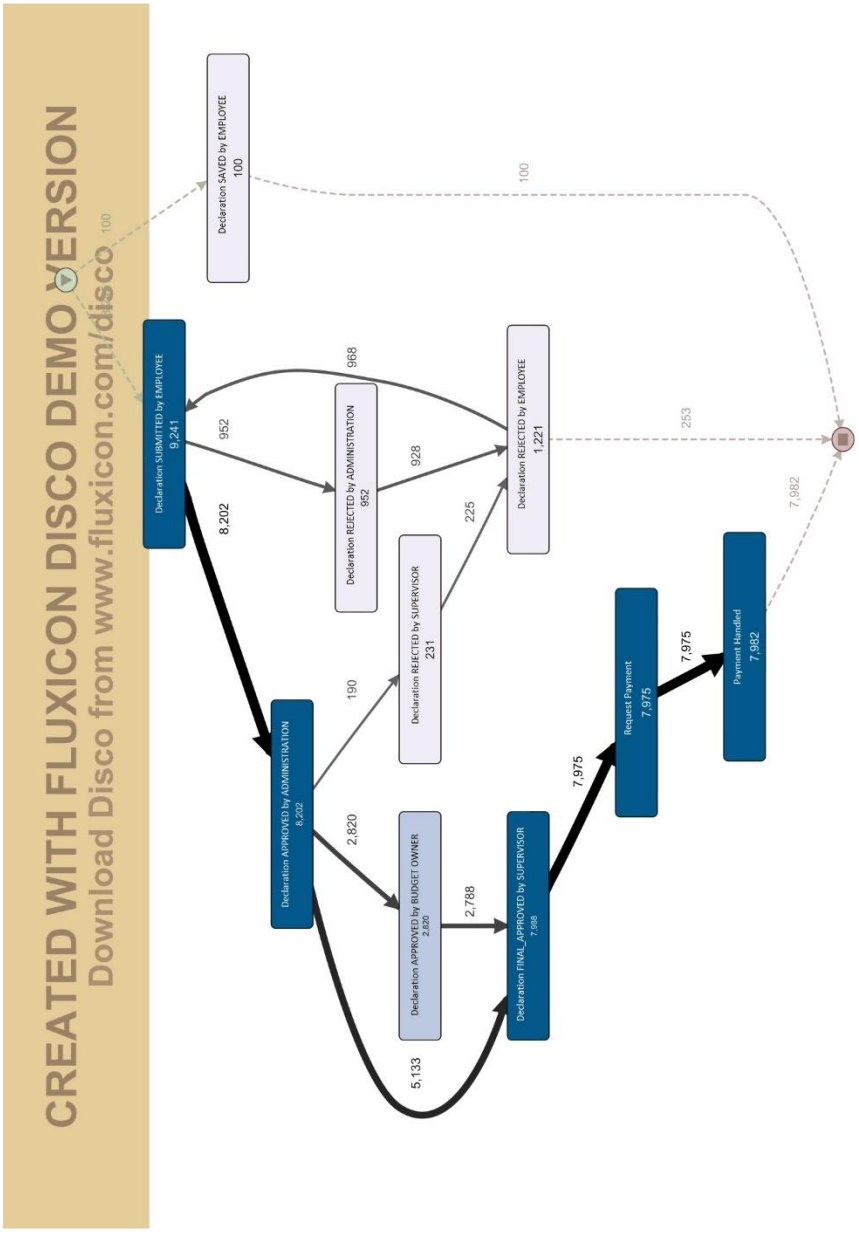

Fig. 10. Complete graph of the process of working with internal declarations

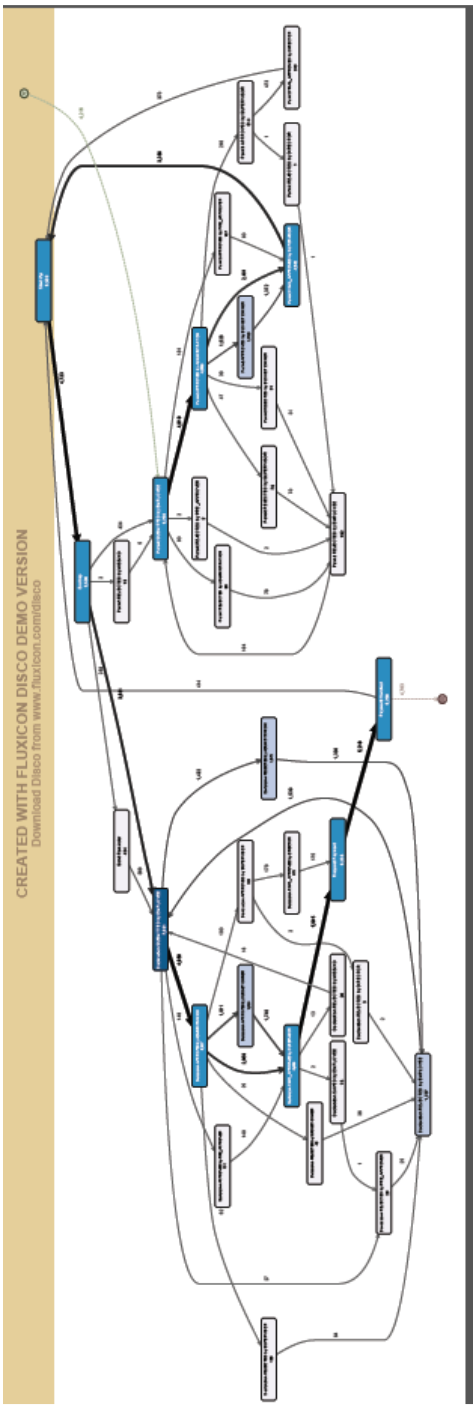

Fig. 11. Complete graph of the process of working with international declarations

References

1. https://icpmconference.org/2020/bpi-challenge/
2. My first acquaintance with Process Mining. https://vc.ru/ml/116313-moe-pervoe-znakomstvo-s-process-mining
3. How does it work Process Mining. https://kachestvo.pro/kachestvo-upravleniya/protsessnoe-upravlenie/process-mining/
4. N. Sitnikov: Process Mining as a method of increasing the efficiency of processes in a company.https://zen.yandex.ru/media/id/5bc8a6b4ad99e300ab487906/process-mining-kak-metod-rosta-effektivnosti-processov-v-kompanii-5be1958900bbe300a935b5b4