

BPIC 2020: Analysis of reimbursement process

Georgy Buzikashvili¹

¹ PJSC Sberbank
buzikashvili.gn@gmail.com

Abstract. This paper is aimed on getting a better understanding of reimbursement process and analyzing its potential weak spot. As work travels sometimes can be expensive and time consuming, poor process management may lead to insufficient use of funds and time. On the other hand, both travels and reimbursement open up opportunities for fraud, which might be an interesting point to analyze. With use of Process Mining and other Data Science techniques we hope to answer these questions and therefore provide our insights on the reimbursement process.

Keywords: BPIC, Reimbursement process, Process discovery.

1 Introduction

BPIC is an annual process mining competition in which participants are invited to investigate some real process using all the data analysis methods available to them.

This time, the organizers proposed to analyze the process of reimbursement of funds spent on working trips. For this, the following data were provided:

- Requests for Payment: 6,886 cases, 36,796 events - requests for refunds;
- Domestic Declarations: 10,500 cases, 56,437 events - declaration of local trips that do not require prior permission from the employer;
- Prepaid Travel Cost: 2,099 cases, 18,246 events - prepaid travel;
- International Declarations: 6,449 cases, 72,151 events - declaration of international travel requiring the prior permission of the employer;
- Travel Permits: 7,065 cases, 86,581 events - travel permits.

Also, the organizers identified some issues that are of particular interest for analysis, namely:

- How long does it take to declare a trip from filing to payments;
- Is there a time difference between domestic travel and international travel;
- Is there a difference between the clusters of cases for declaration;
- How long each step of the process takes;
- Bottlenecks of the travel declaration process;
- The bottlenecks of the travel authorization process;
- How many declarations are rejected at various stages of the process and how many are never approved;
- And a number of other questions on more specialized topics.

In this paper, some of the questions posed by the authors will be explored. But to begin with, a more "raw" analysis of the data provided will be made.

2 Data

The data was presented in the form of 5 separate log files (mentioned above), each of which contains information on a specific sub-process. In addition to the data inherent in log files (case id, timestamp and the like), they also contain more specific information that may be of interest. Below we will consider such attributes for each of the files.

2.1 Requests for Payment

Among other attributes specified in this log file, the following may be of particular interest:

- Cost type;
- RequestedAmount.

The cost type turned out to be equal to zero in all cases, which leads to the conclusion that it is not of interest for use in further analysis.

RequestedAmount – the requested reimbursement amount. Using the pandas library, let's look at some statistics:

Table 1. Statistics of the RequestedAmount attribute.

Statistics	Amount
count	6 886
mean	406,73
std	11 858,85
min	-
25%	36,13

It turns out that most of the requested sums are no more than 253 currency units. However, there is a case when this amount turned out to be about a million, which begs the question - what is the reason for such a large number? We will explore this issue in the next block.

2.2 Domestic Declarations

As mentioned above, for travel within the country, a simplified version of the process is provided, which is why, among others, one can highlight, perhaps, which may be interesting for analysis - Amount (the amount spent on the trip).

2.3 International Declarations

For declaring trips abroad, a more complex procedure is provided, which requires prior approval from the employer. Moreover, in the process of declaring, the requested amount can be adjusted. Below are the most interesting attributes:

- Amount
- RequestedAmount
- AdjustedAmount
- Permit RequestedBudget

Here we immediately found out how often this amount is adjusted and by how much. As it turned out, this happened only in one case; in the future, we may consider this case in more detail.

2.4 Travel Permits

Here, from an interesting point, there is an OverspentAmount attribute, which contains information about how much the budget for the trip was exceeded (or vice versa, was not used).

Since this attribute has several explicit outliers, let's remove them first. After that, we display the distribution:

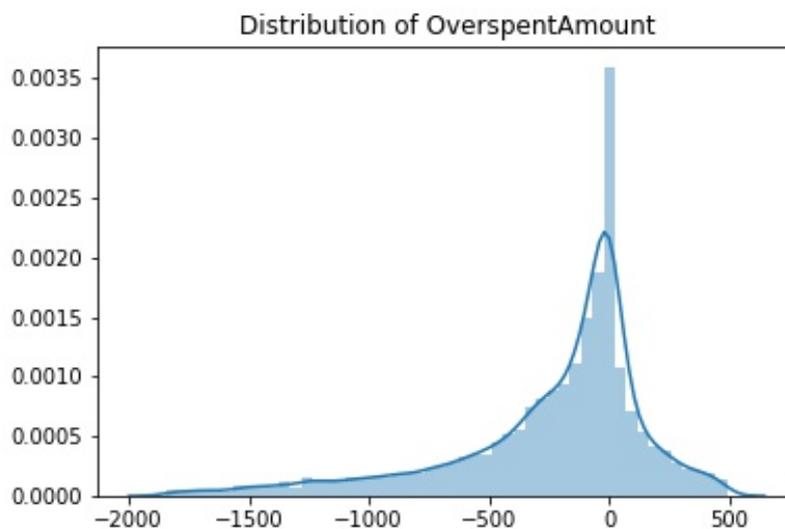


Fig. 1. Distribution of OverspentAmount.

As you can see (except for cases when the value is equal to zero), basically the budget is not spent to the end. This may indicate that there is an opportunity to optimize the process in terms of estimating the required budget for the trip.

3 Competition questions

Next, we will consider some of the issues that are of greatest interest.

3.1 Q1: How long does it take to declare a trip from submission to payment

Consider the declaration process for international travel:

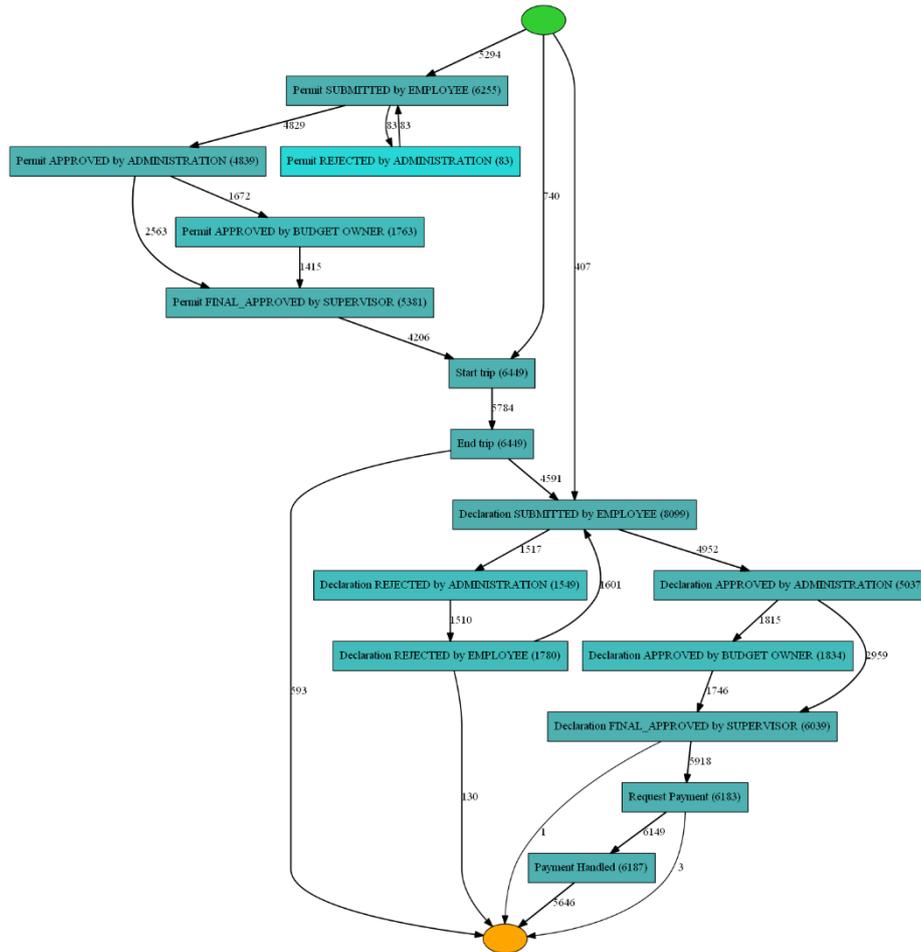


Fig. 2. Graph of international declaration process.

In this case, the `heuristics_miner` of the `pm4py` library was used. The `dependency_thresh` parameter was set to 0.999, so that the graph was not unnecessarily cluttered with process route paths, which are relatively rare for requests.

When declaring international travel, prior travel authorization is required.

References

1. PM4PY Python library, <http://pm4py.pads.rwth-aachen.de/>, last accessed 2020/08/30
2. BPIC 2020, <https://icpmconference.org/2020/bpi-challenge/>, last accessed 2020/08/30.