**Term project report**

**16102284 Lee SungHo**

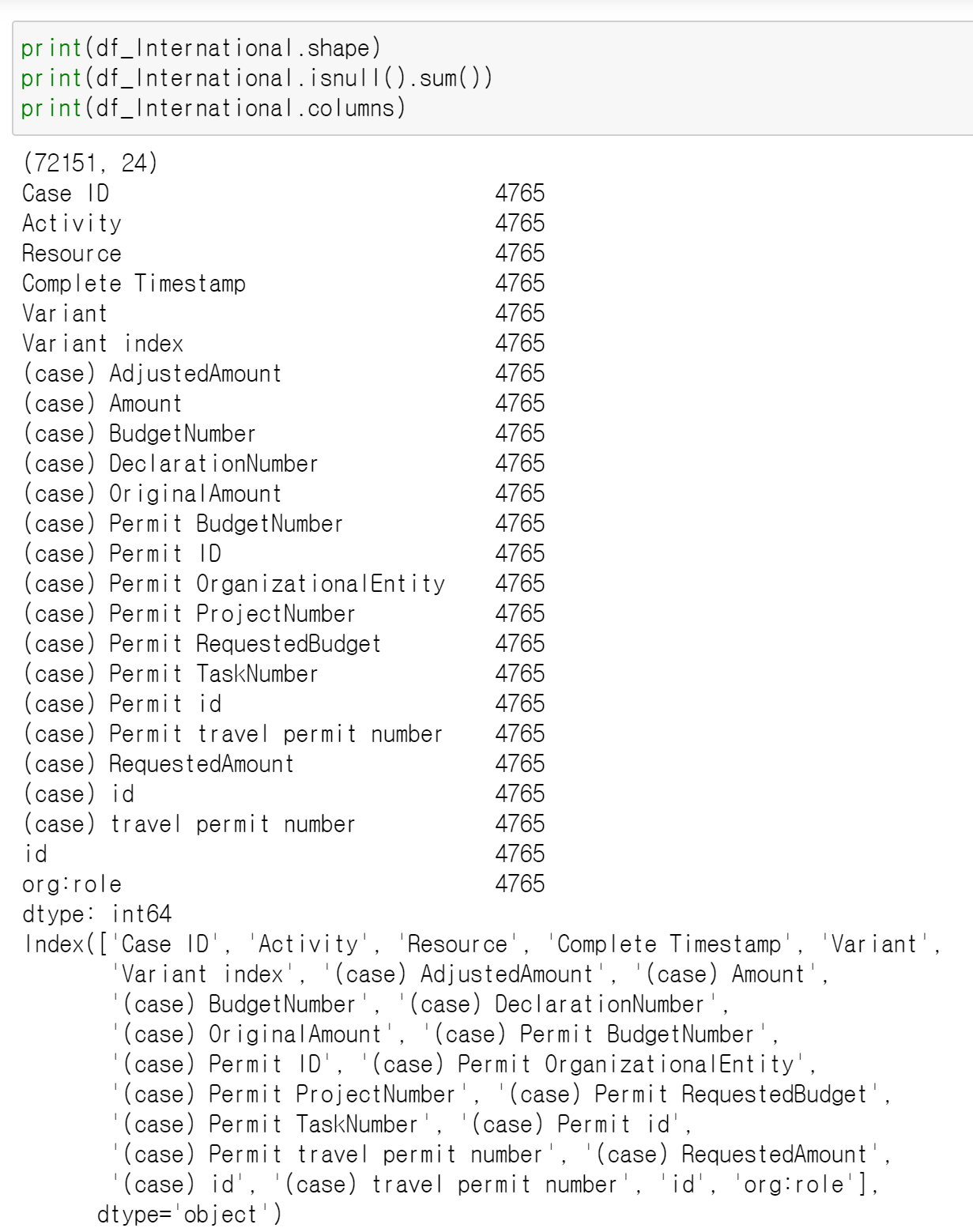
Data Preprocessing



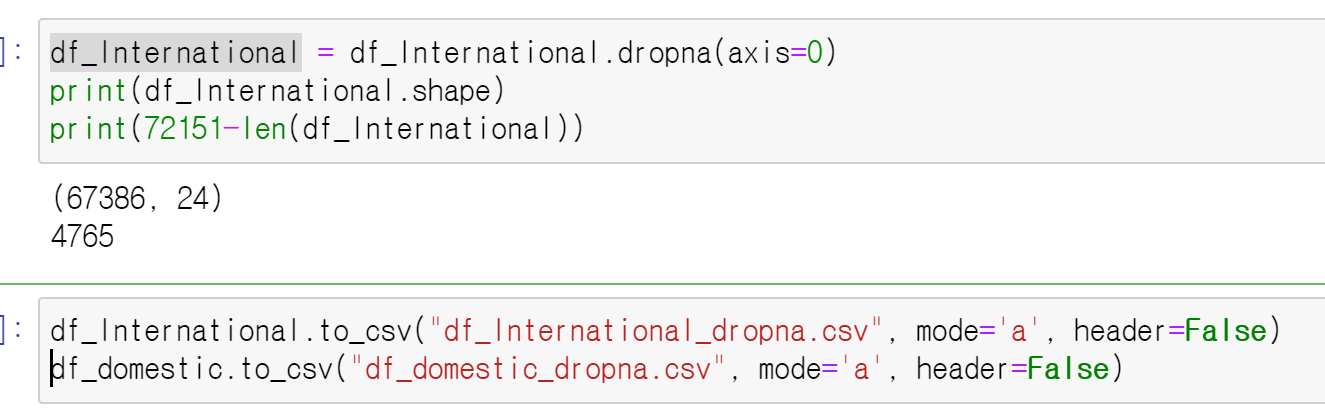
First, I analyzed the domain data. Once there are twelve columns, the row consists of 56162 events. It is important to note that there are a total of 105 missing values. So it was decided to eliminate missing values first.



As a result of removing the null value, we can see that a total of 105 data were erased and the missing occurred at once in one event.

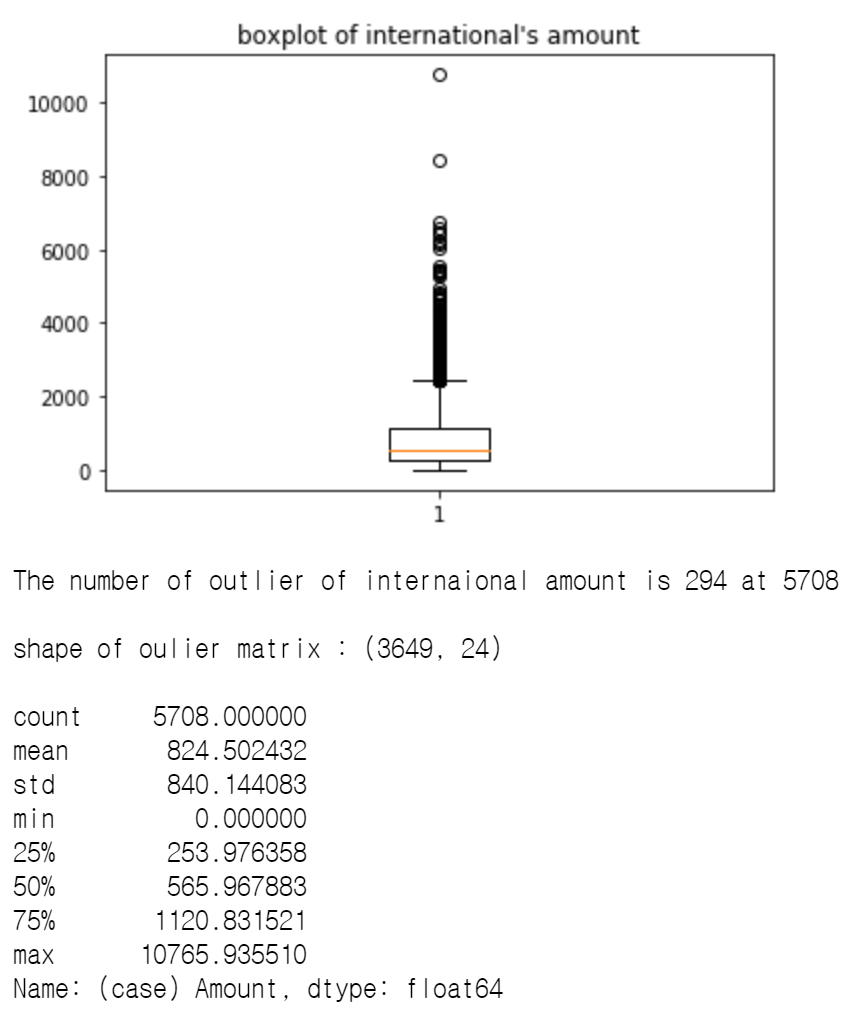
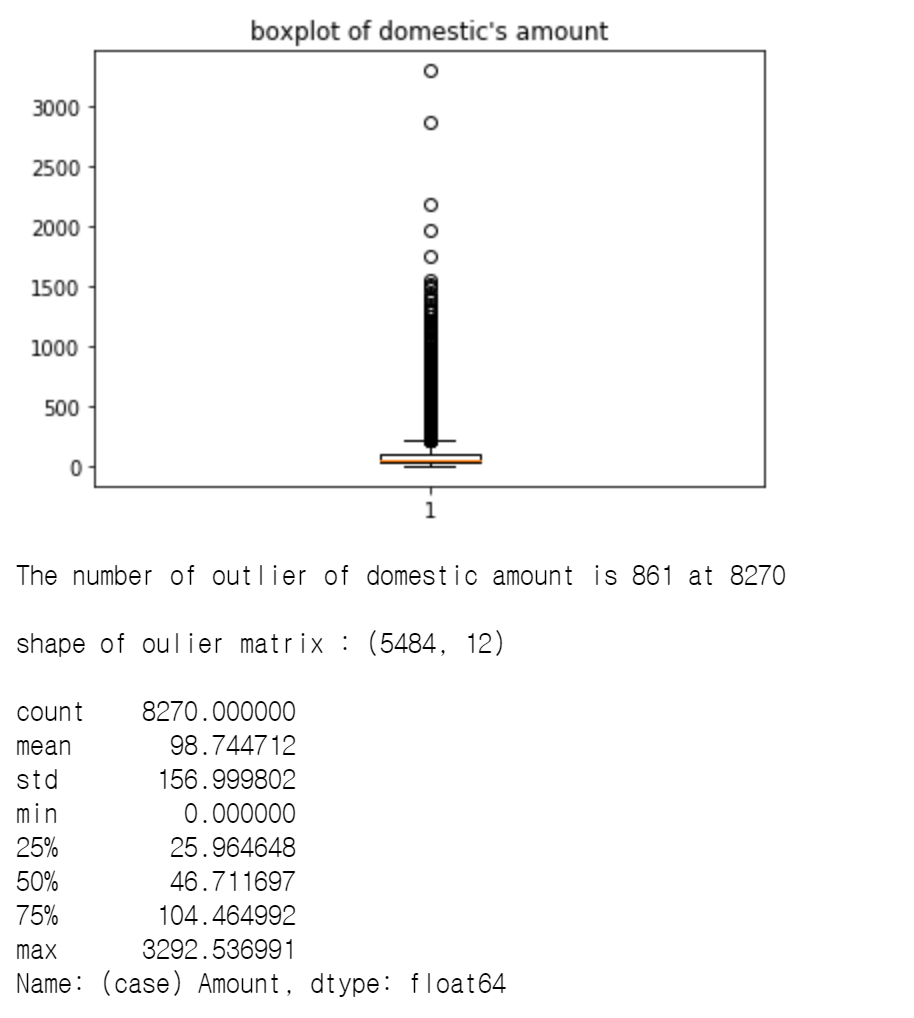


Similarly, the international data was investigated in the same way, and a total of 4,765 missing values were present, and the same was decided to be eliminated.



As a result, the international data also shows that all the defects occurred at one event. In the future, I will proceed with process mining using data with missing values removed.

% further analysis %



Additionally, I pre-processed the ‘amout’ column that exists in each data frame. As a result, the values of the outlier occurred in each data. These data values may find new insights on social issues such as luxury business trips(<https://www.hankyung.com/society/article/201602027605H>), so I will also proceed with process mining.

Model Fitting

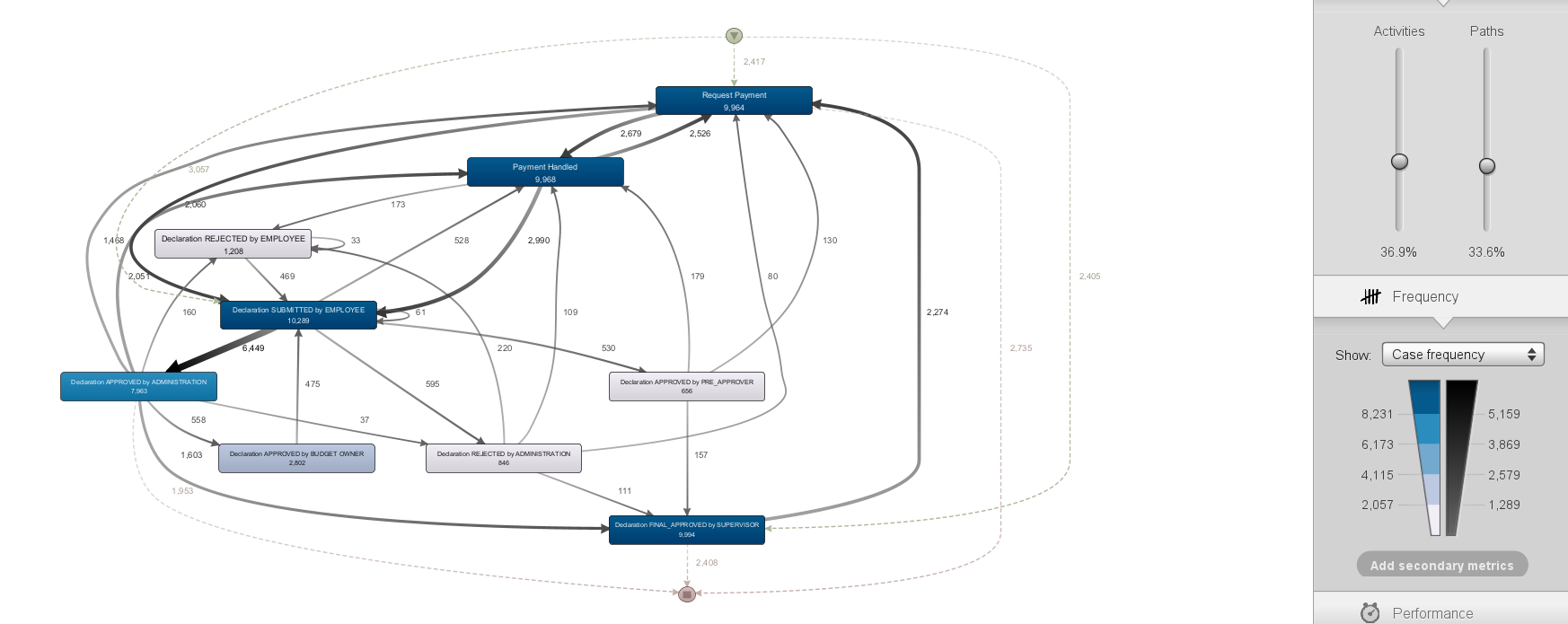


Fig 1. Domestic flow chart

First, I used all events except null to draw a flow chart like Figure 1. The reason for drawing these graphs was that first, I wanted to know the difference between the Outlier events about the cost of use, so I used all the data without filtering to compare fairly. Then, set activity and path to 0% to make them underfit, and then, by raising the two figures at the same time, using the central limit theorem, the figures were adjusted so that all pathways with significant statistical significance have more than 30 samples.

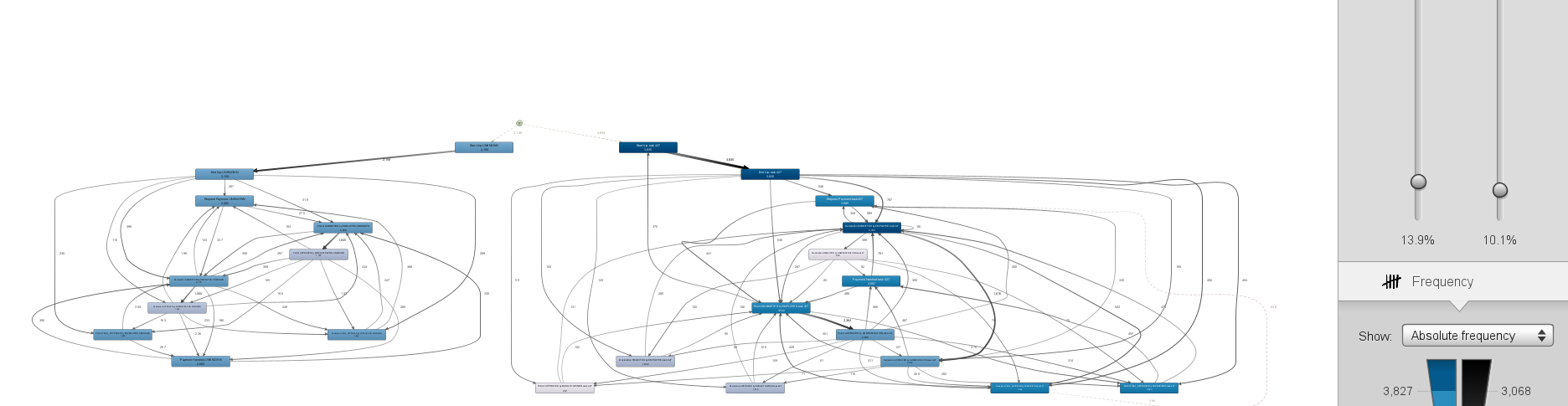


Fig 2. International flow chart

I drew International in the same way as above. The result shows that there is a fairly complex process compared to the domain stick.

* What is the throughput(or flow time) of a travel declaration from submission (or closing) to paying?

:

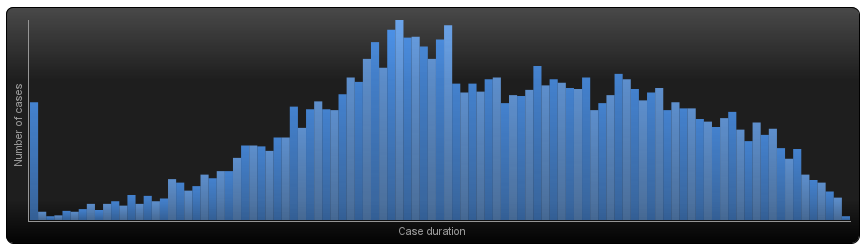


Fig 3. Domestic case duration’s describe & histogram

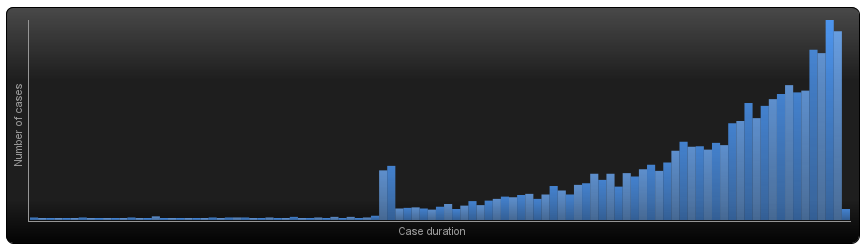
 

Fig 4. International case duration’s describe & histogram

Fig3 and Fig4 are the median, average, and histogram of the duration of the previously produced flow chart. First of all, as you can see by looking at Fig 3, the graph is slightly positive skewed in a form similar to the normal distribution. Thus, statistically speaking, the median is the value representing the distribution rather than the mean. Thus, the representative value of the domestic case duration is 33.5 hrs. In contrast, Pig 4 is clearly negative skewed. Thus, the representative value of the International case duration is 51.5 hrs, the average.

* Is there are difference in throughput(or flow time) between national and international trips?

As previously explained, domestics required 33.5 hrs and international 51.5 hrs, which required an average half-day approval period. This can be inferred from the much more complex processes, as shown by the previous work on Fig1 and 2.

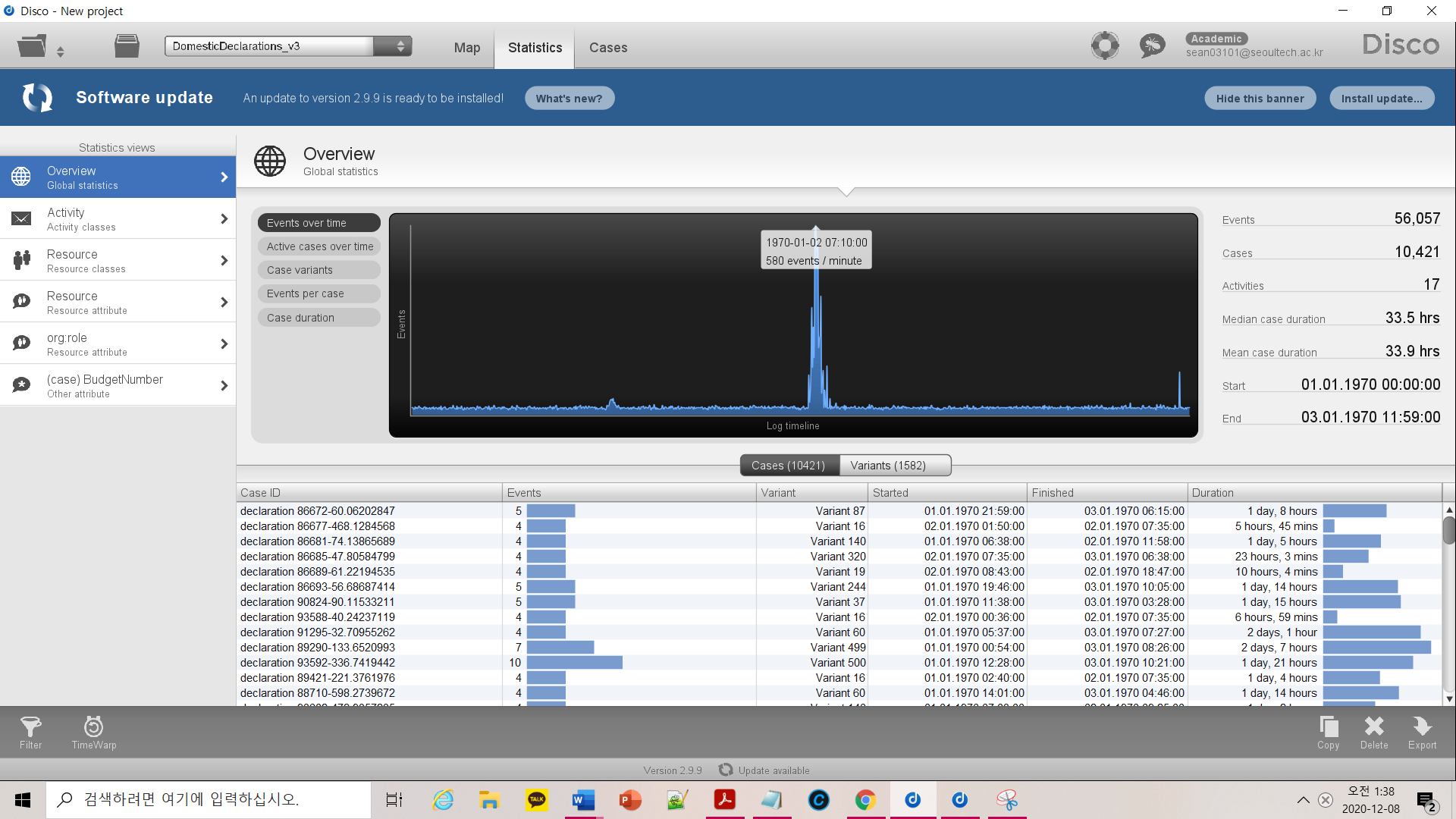


Fig 5. Domestic event case overtime distribution

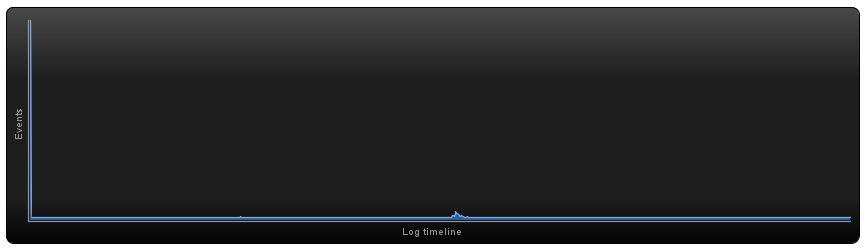


Fig 6. International event case overtime distribution

And Fig 5 and 6 are the overtime distribution graphs of each event case. At Fig 5, overtime events are largely grouped and distributed in three places, especially near 31hr. And Fig 6 is the shape in which the data are mainly distributed to the left. In other words, the international case's activities were well executed on time, but the domestic case shows that many other activities were not executed on time.

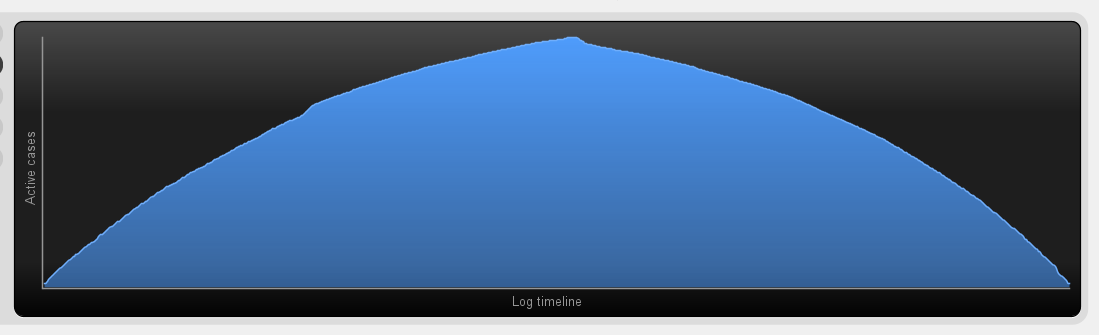


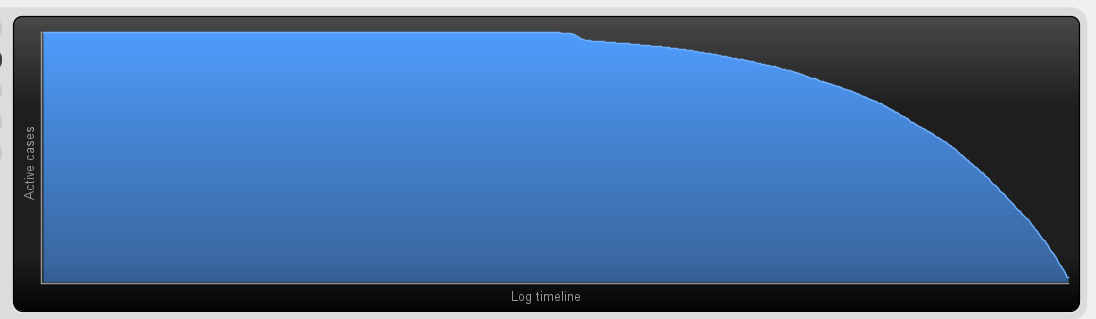
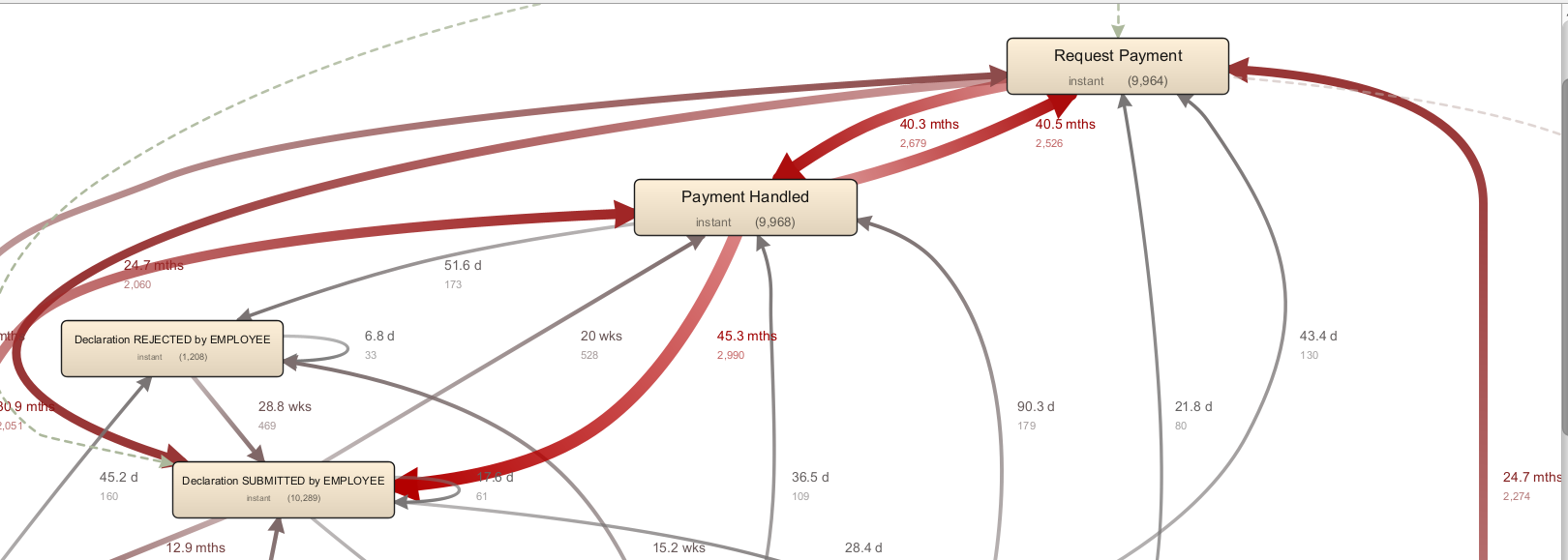
Fig 7. Domestic active case overtime distribution 

Fig 8. International active case overtime distribution

This trend is similar in the case-overtime graph. Figure 7 is a distribution of domestic active cases, which is shown to be evenly distributed according to the log time of the active case, with a distribution shape similar to the normal distribution. In contrast, the International active case shows that the data are skewed to the left at figure 8. It means, in the case of international declarations, 95.94% of the 6449 reports were paid, while the rate of 1,500 domestic reports was slightly slower at 95.62%. The remaining cases are saved as drafts that have not been approved or submitted.

* Where are the bottlenecks in the process of a travel declaration?



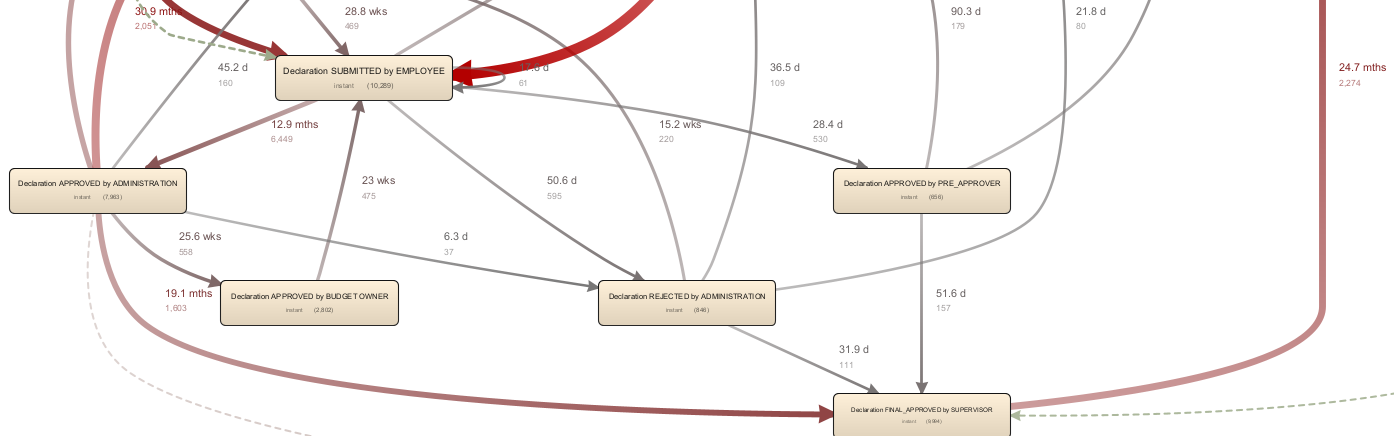


Fig 9. Domestic performance flow chart

Domestic's botlenecks in the process of a travel detail is quite simple. Fig 9 shows that bottlenecks are occurring in many events, but the bottleneck is caused in the ‘declared sub-by-employee’ related to declaration. In addition, the final process, the ‘declaration final approved by supervisor’, is occurring significantly, if not as much as the bottleneck that occurred earlier.

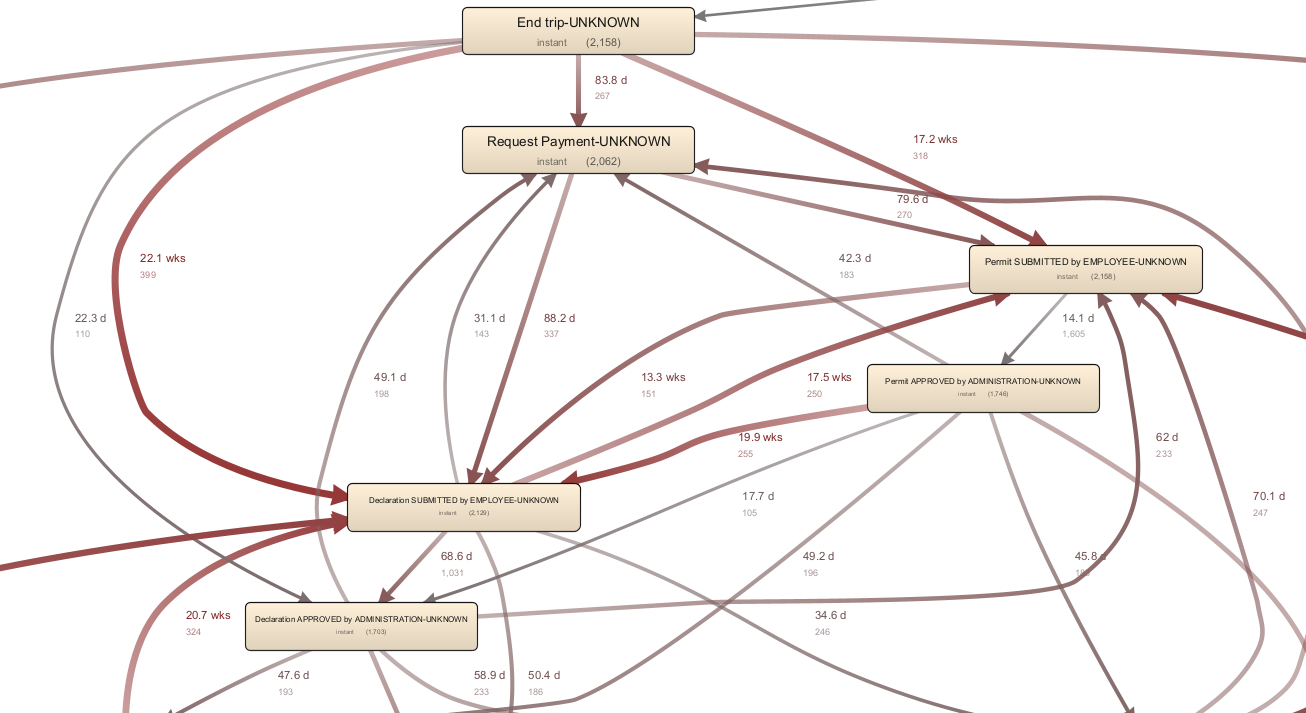


Fig 10. International performance flow chart not relate task 427

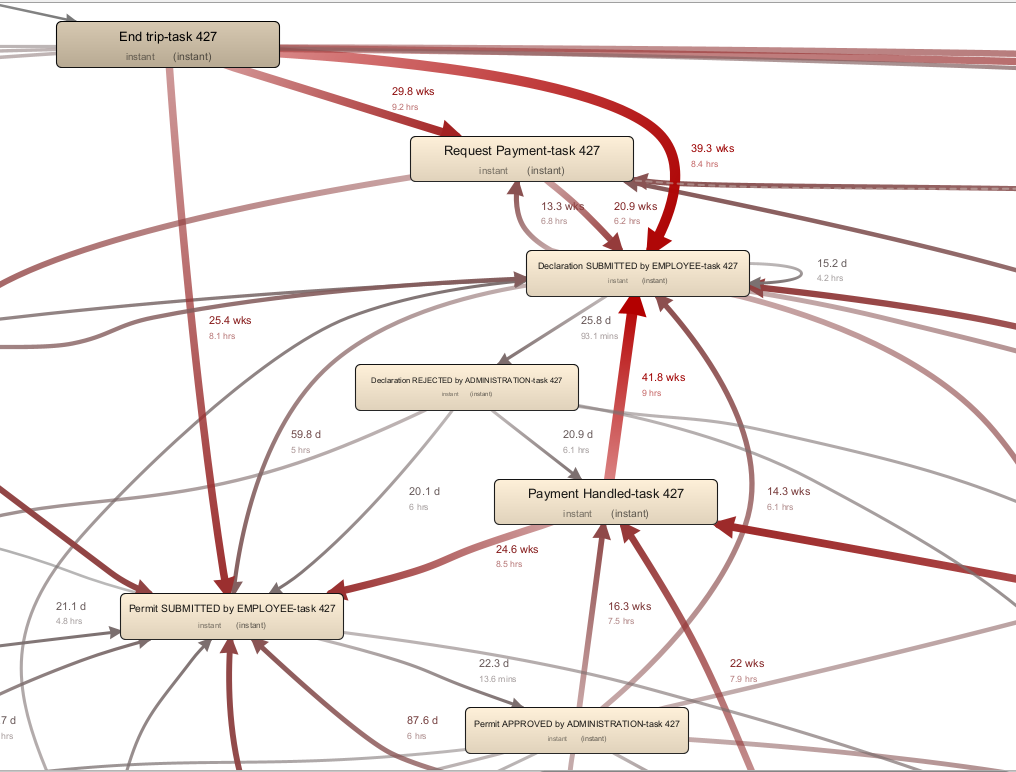


Fig 11. International performance flow chart relate task 427

As a result of the process fitting, International was taught as a very complex model and found something unusual. International has an unusual structure in which flow charts related to task 427 and non-related flow charts are split into two sides and approved for payment.  
First of all, if you look at fig 10, the 'declared sub-by-employee' part comes out as a bottle snack. This is similar to the results obtained earlier through fig 9. Also, the events created are filled with events similar to the previous domestic flow chart.  
According to Figure 11, the Bottlenack phenomenon is worse than the previous analysis of Figure 10. In particular, "declared by deployee task 427," which appeared for the first time, is the most severe. In addition, the previous analysis of 'declaration final applied by supervisor' shows a similar tendency to Bottlenack. In other words, this shows that international has various events in delclaration, so there seems to be a big difference, but in conclusion, the parts where Bottlenack occurs are both similar.

* Where are the bottlenecks in the process of a travel permit (note that there can be mulitple requests for payment and declarations per permit)?

As you can see from the previous analysis of Figure 11, ‘payment handled‘ and ‘request payment’ cause serious bottlenecks in the domain case in the permit section. And all of the analyses related to declaration that we analyzed earlier are connected. In other words, it can be seen that the two activites, the permit part, become a major role in bottlenecks.

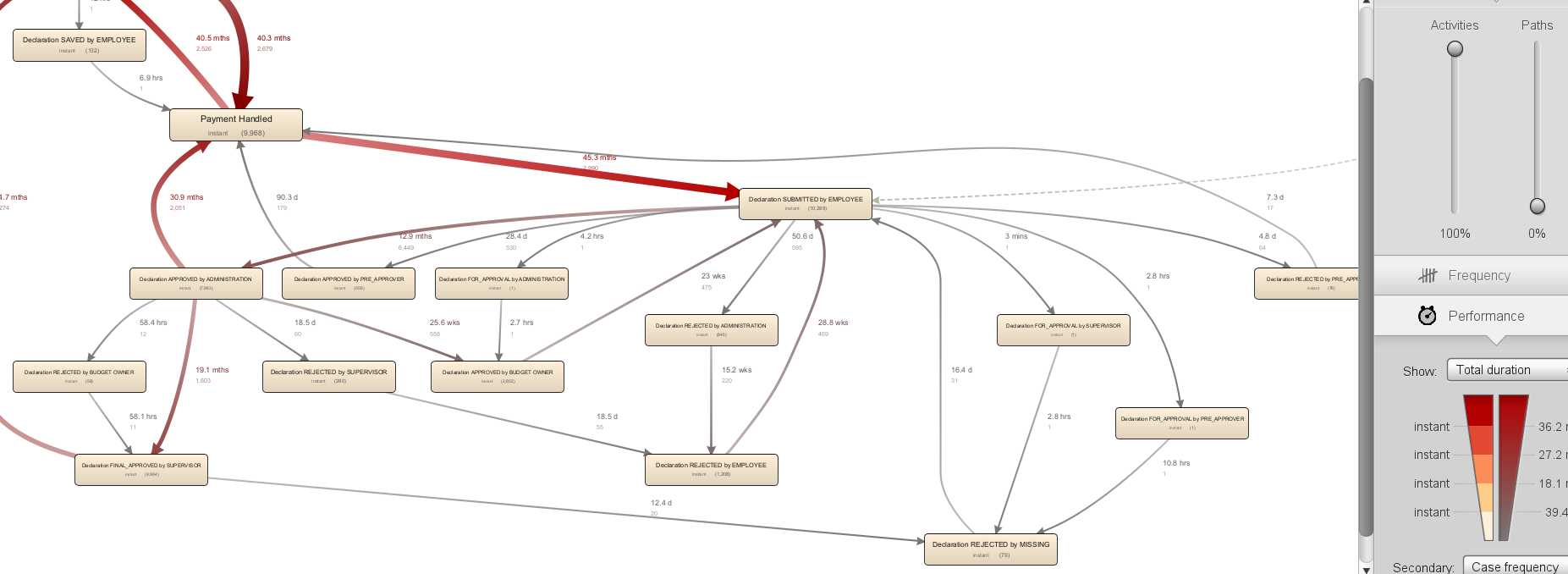


Fig 12. 100% activities domestic flowchart fitting

Also, there is no activity with permit written on it, except for two activites, even though fitting at 100% like Fig 12.

In contrast, international tends to be different. As you can see from Figure 10, there are many activities related to permits such as 'permit SUBMITTED EMPLOYEE'. In particular, as you can see from Fig 12, there are activities such as 'permit Submitted by Employee', 'payments Handle Task' and 'permit final applied by supervisor' and these contribute greatly to Bottlenack. In particular, 'payments handled' was applied as the main cause of Bottlenack, as previously analyzed domestics.

Outlier model Fitting

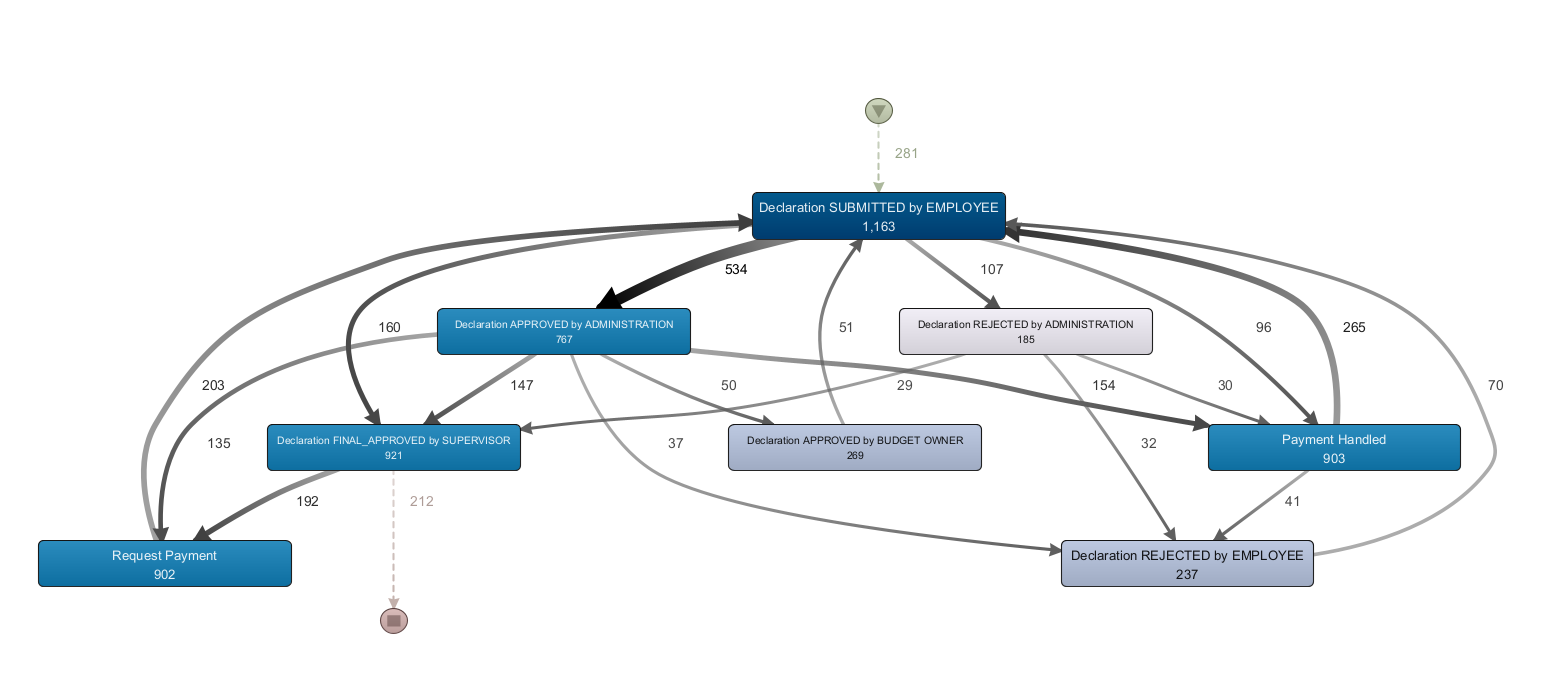


Fig 13. Domestic amount outlier model

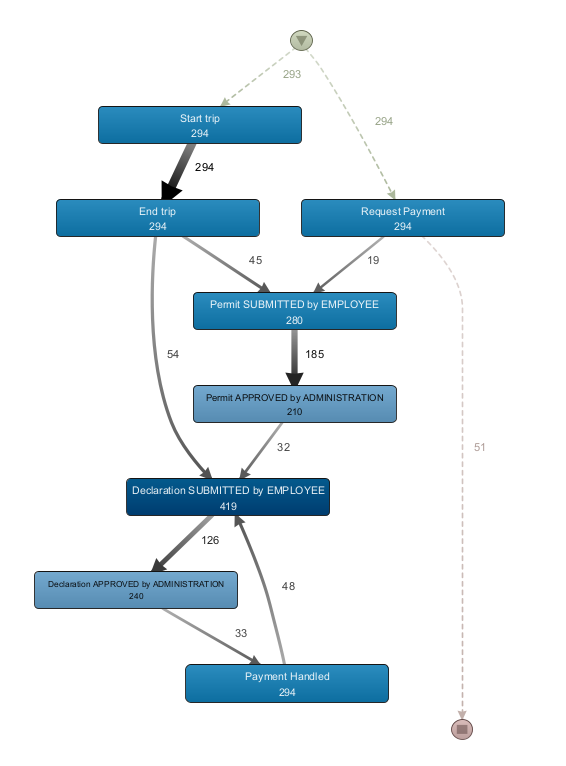


Fig 14. Domestic amount outlier model

It was carried out in the same way as the model fitting that was carried out earlier. Here, however, one of the samples of the critical path is 19, which violates the central extreme theorem, but it proceeded with model fitting.

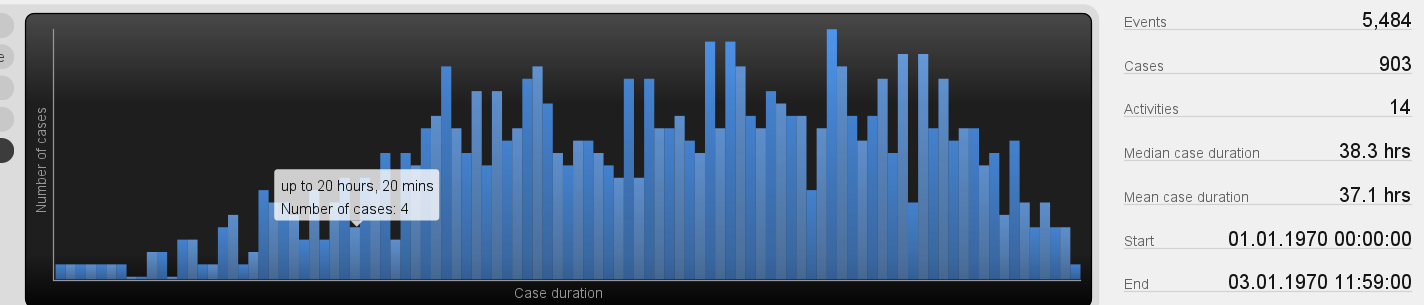


Fig 15. Domestic amount outlier duration describe & histogram

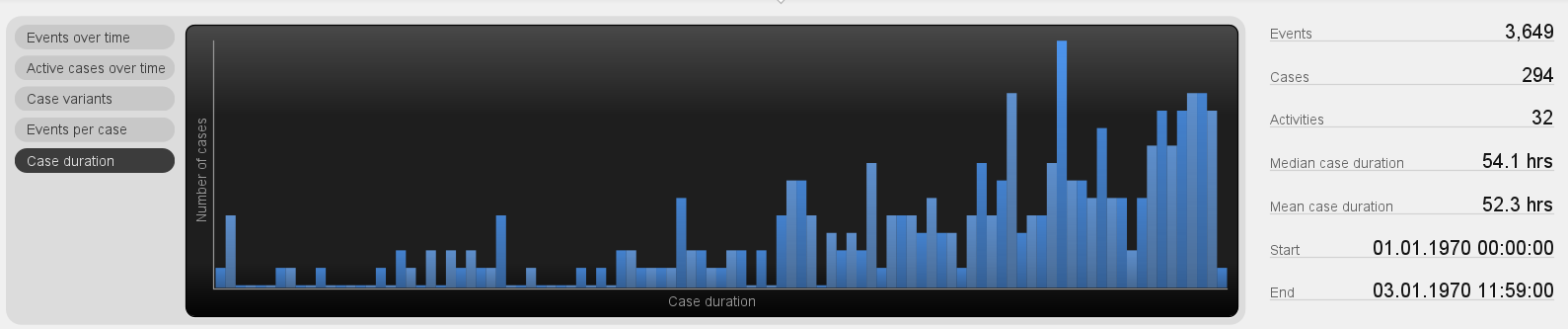
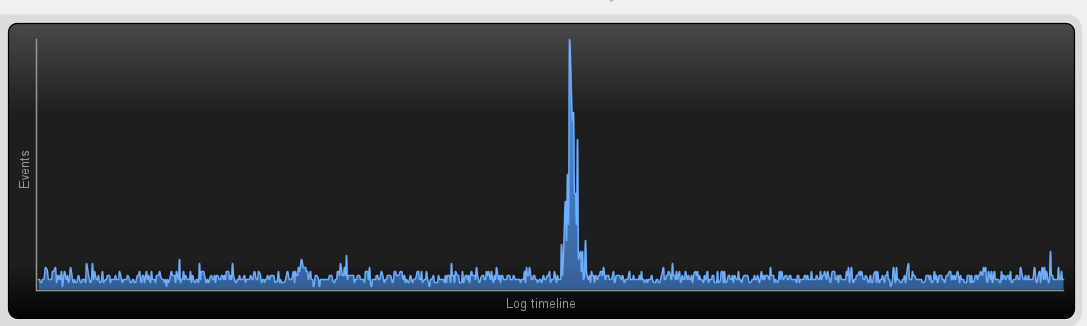


Fig 15. International amount outlier duration describe & histogram

The representative value of the Dominic account outlier distribution is 37.1 hrs and the International account outlier distribution is 52.3 hrs. This is all higher than the values previously obtained. This shows that the average duration of the outlier values takes longer.



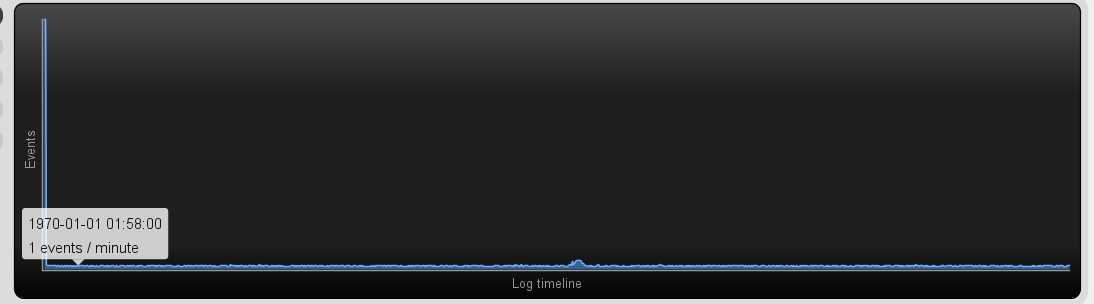
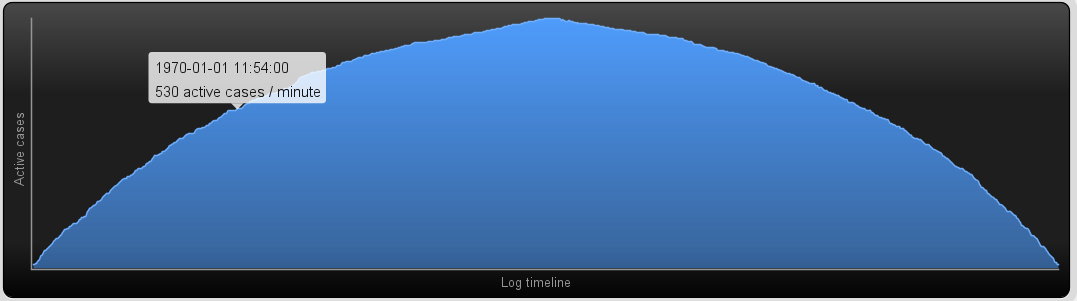


Fig 16. Domestic & International event case overtime distribution



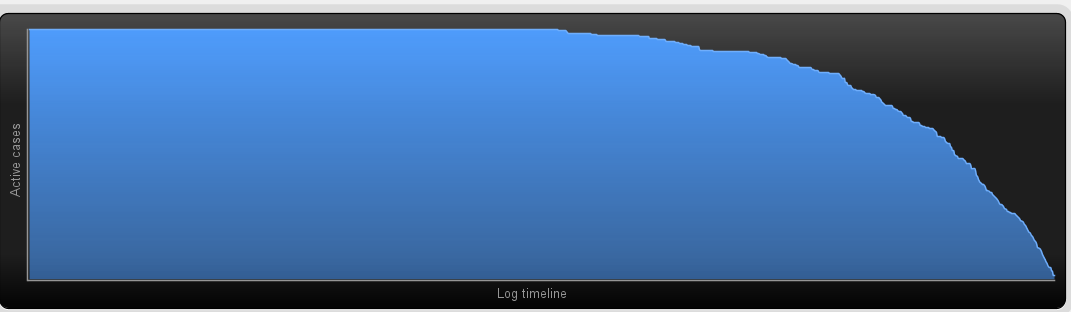


Fig 17. Domestic & International active case overtime distribution

As previously conducted, each activity and event was investigated for an additional time like fig 16 and 17. This is quite similar to the results of the previous survey. In other words, this means outliers are statistically similar except for an increase in average duration.

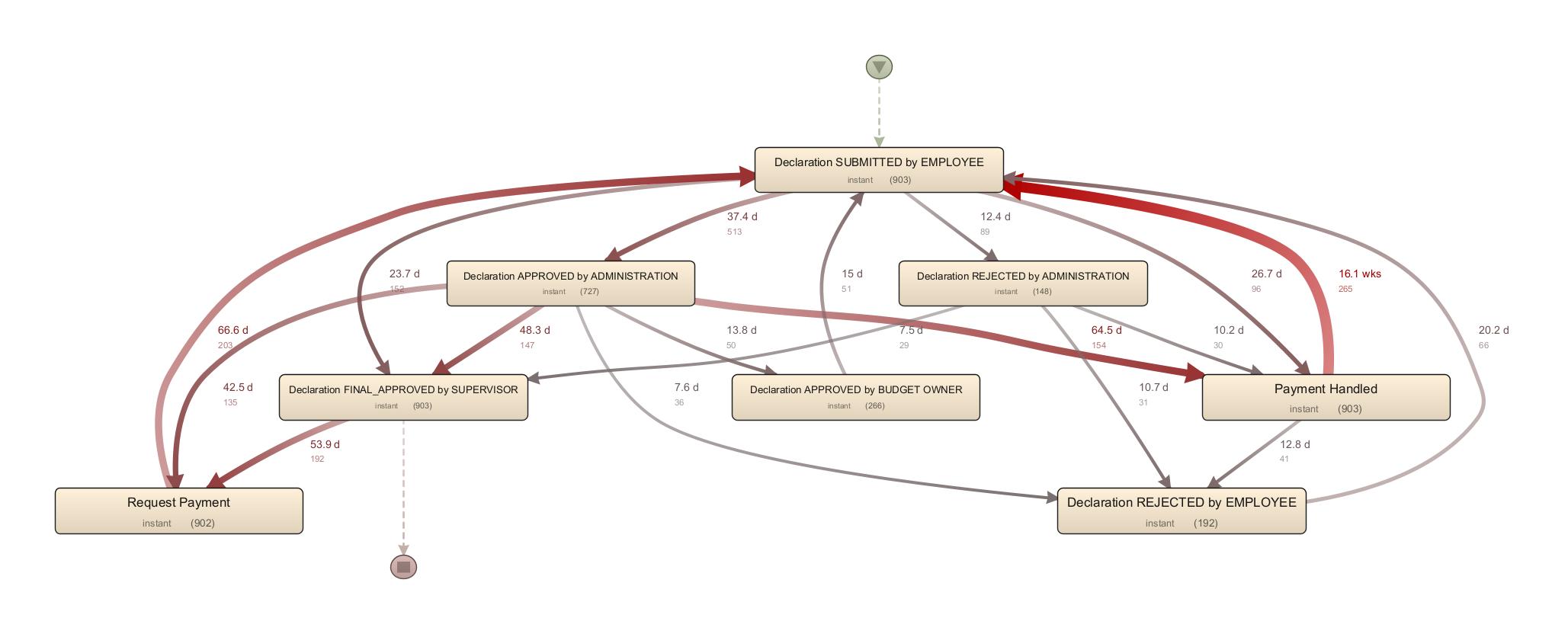
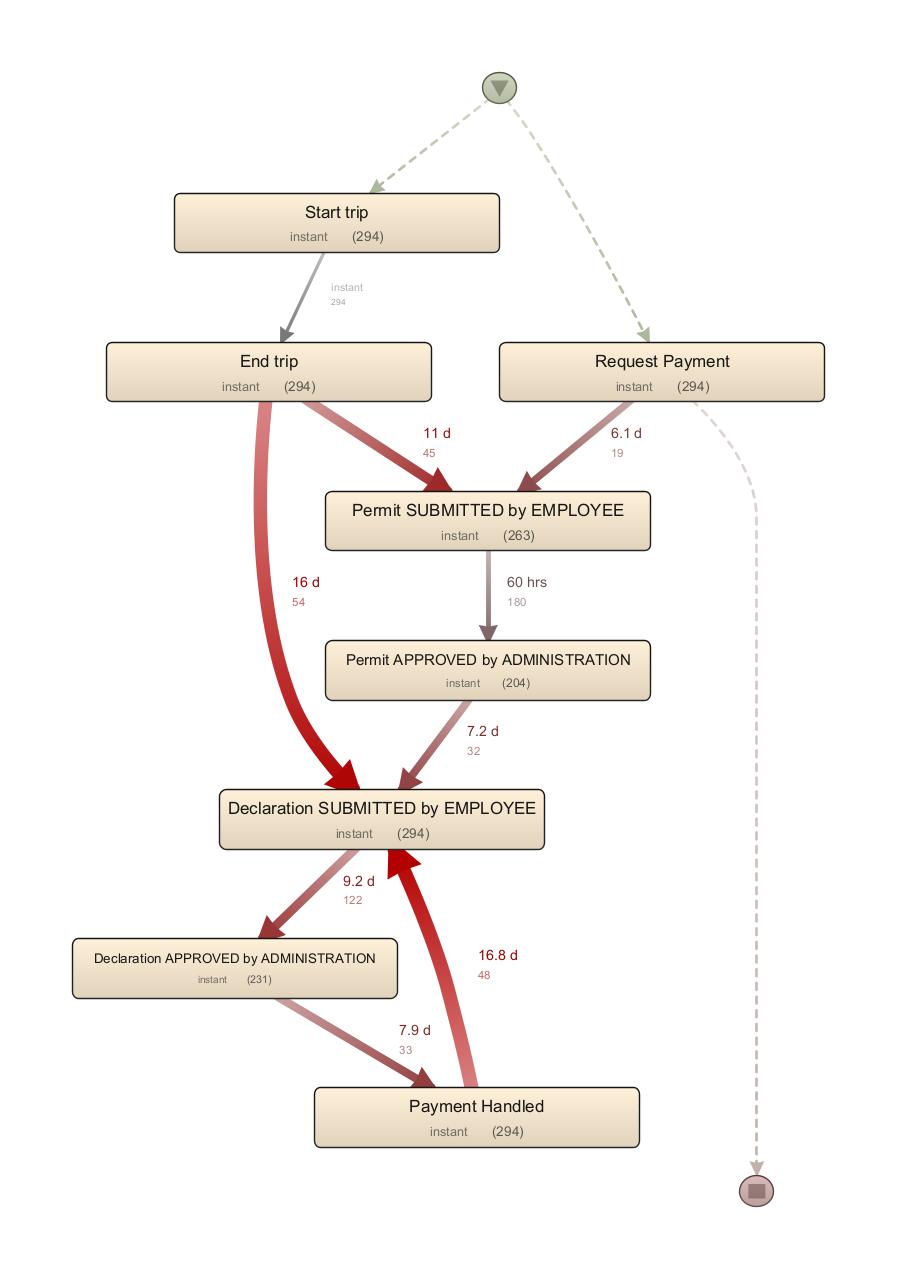


Fig 18. performance flow chart



Domestic & International performance flow chart

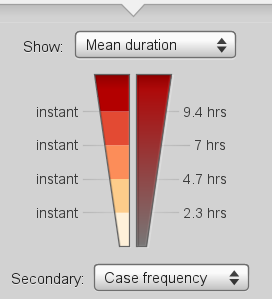
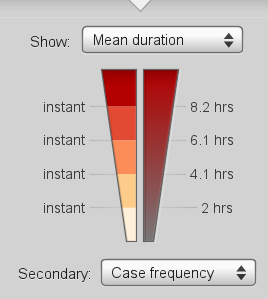
 

Fig 19. Compare to range of waiting time domestics (right:outlier)

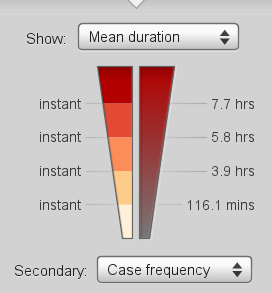
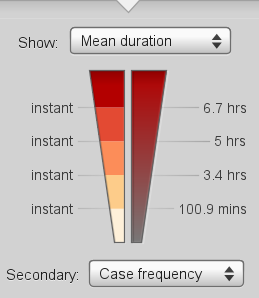
 

Fig 20. Compare to range of waiting time domestics (right:outlier)

As you can see in Figure 18, a fairly simple model has emerged, especially in international, unlike the model previously created using overall data. Because the data has shrunk, it seems that it couldn't learn the model elaborately. Except for that, the Bottle Neck phenomenon occurred equally in similar parts. However, as you can see from the Figs 19 and 20, the average waiting time is longer for the entire data.

Conclusion

Thus, in conclusion, through Fig 15,19,20, it can be seen that the outlier model has a longer average duration period and a shorter bottle-neck wait time. That is to say, high-cost business trips are a major concern and are intended to be handled as quickly as possible, and there are many details that are not expressed in the process model flow, so it can be assumed that duration time is long. This meticulous administrative process shows that a luxury business trip, which had previously been questioned, is virtually impossible, and it was just a costly business trip.

In order to improve the process efficiently, it is also necessary to resolve the common activity of domestic and foreign groups, called 'declaration final applied by supervisors'.