**Prom Vs Disco:**

ProM is a very extensive tool which allows for dozens of different model types. It allows for many different input formats and it allows for many different types of process mining

Disco is simple, fast and easy to use. But in terms of functionality, coffers only smaller parts of this course. The advantages are easier and pleasant to use.

The data used in this assignment is available for download below. It contains an 'fbt' file to be opened in Disco, an 'xes.gz' file to be opened in ProM and a 'normativeModel.pnml' file to be used in a later question.

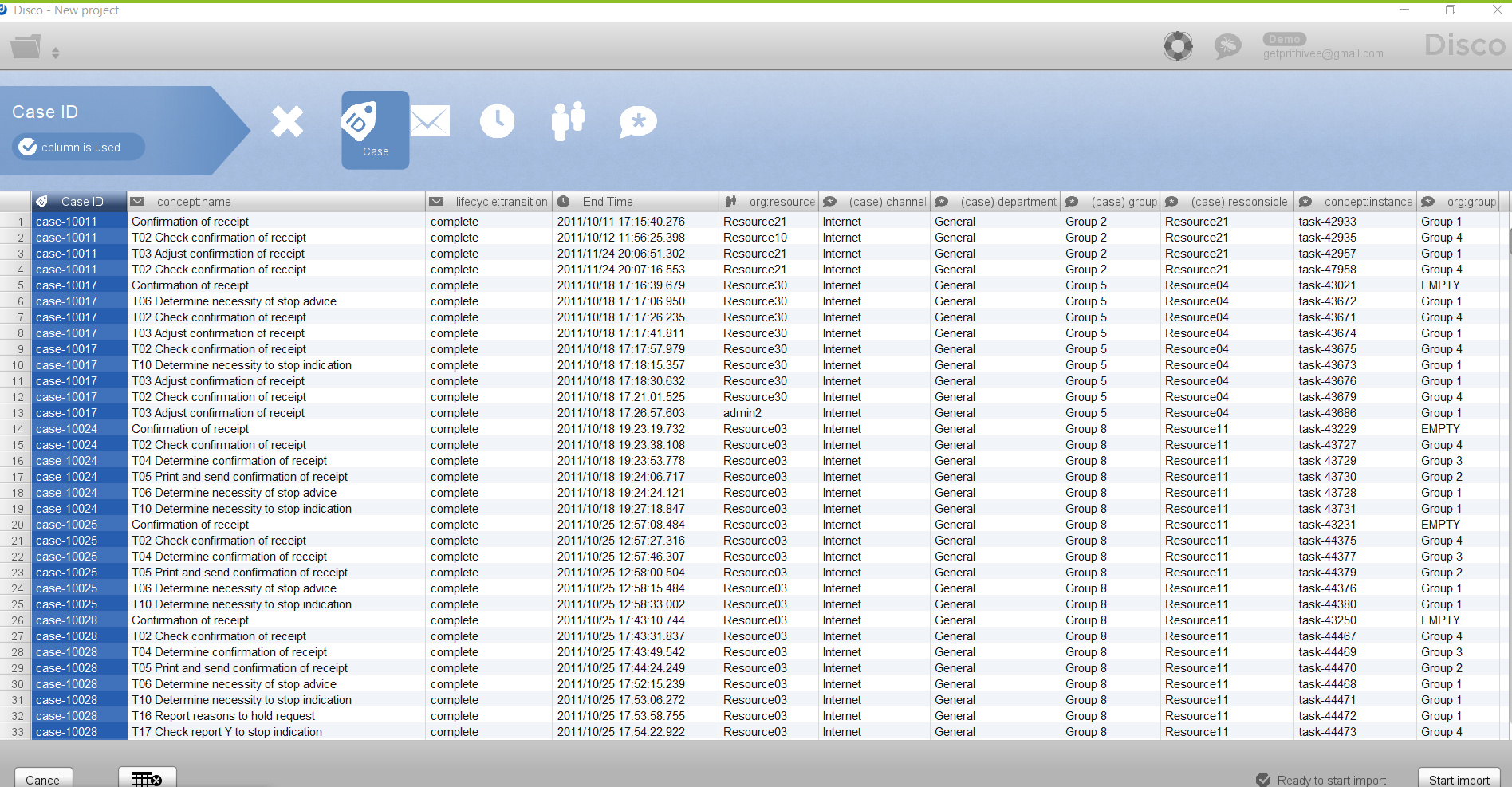
Open the event log ('Receipt phase of an environmental permit application process (\_WABO\_) CoSeLoG project.fbt') in Disco and switch to the 'Statistics' view.

Without switching to other views, use the statistics view to answer the following three subquestions:

How many events are there on average per case?

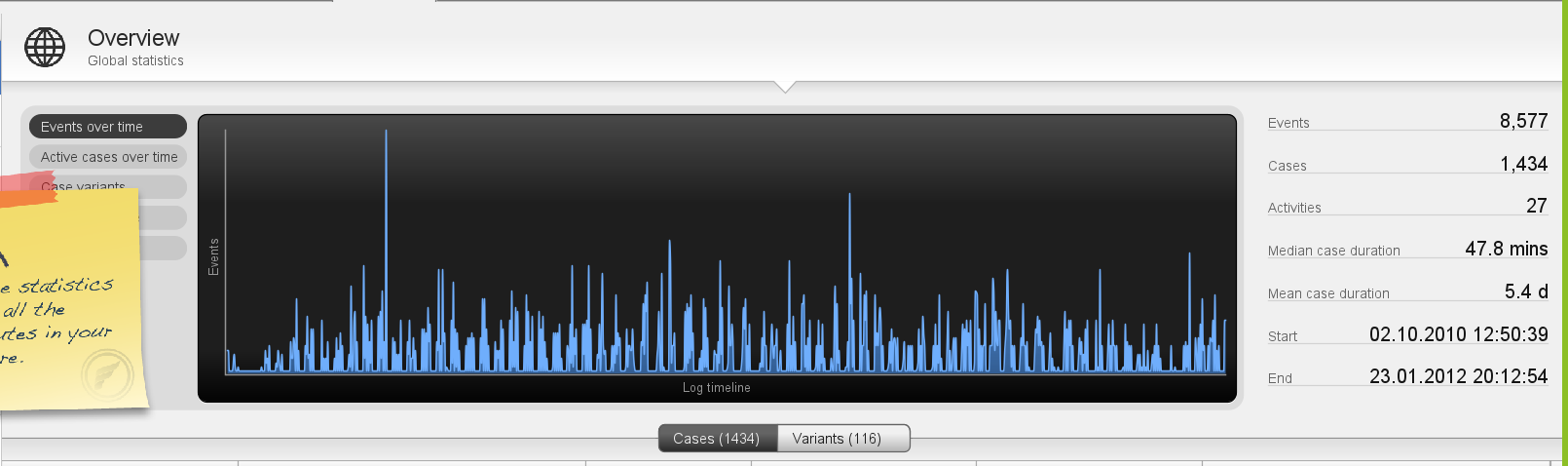
Can you indicate whether each case seems to be unique or whether many cases follow the same activity sequence?

What is the main observation that can be made from the 'Events over time' graph?



1.Open the fbt file using demo mode. ('Receipt phase of an environmental permit application process (\_WABO\_) CoSeLoG project.fbt'

2. Click import to open statistics view. Initially it was in tabular format. Opened statistics according to the question



3. The above image is from statistics page. From this we can answer the first question.

Question 1

How many events are there on average per case?

Total cases = 1434

Total Events = 8577

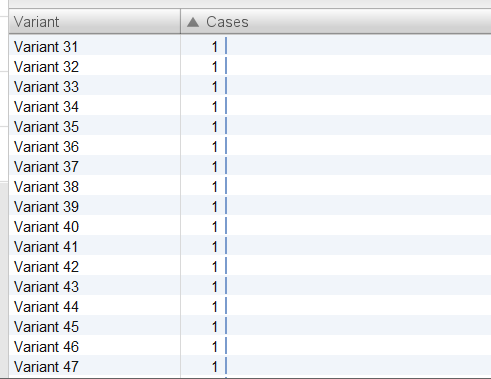
Number of events per case = 8577 / 1434 = 5.98

Question 2

Can you indicate whether each case seems to be unique or whether many cases follow the same activity sequence?

Total cases = 1434

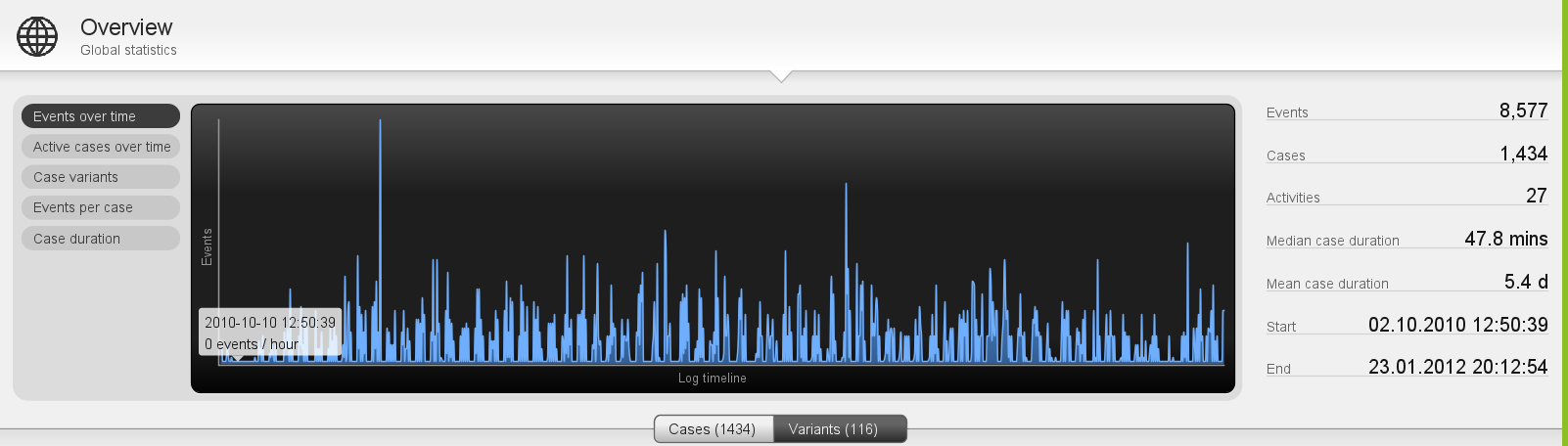
Unique Cases = 86



As we can see the total number of variants are 116. After the 31st variant everything is unique variant. So 116 - 30 = 86. So there are 86 unique cases. We can also verify this using the above graph. Where after 31 everything has a single variant

Question 3

What is the main observation that can be made from the 'Events over time' graph?



The above graph is events over time graph.

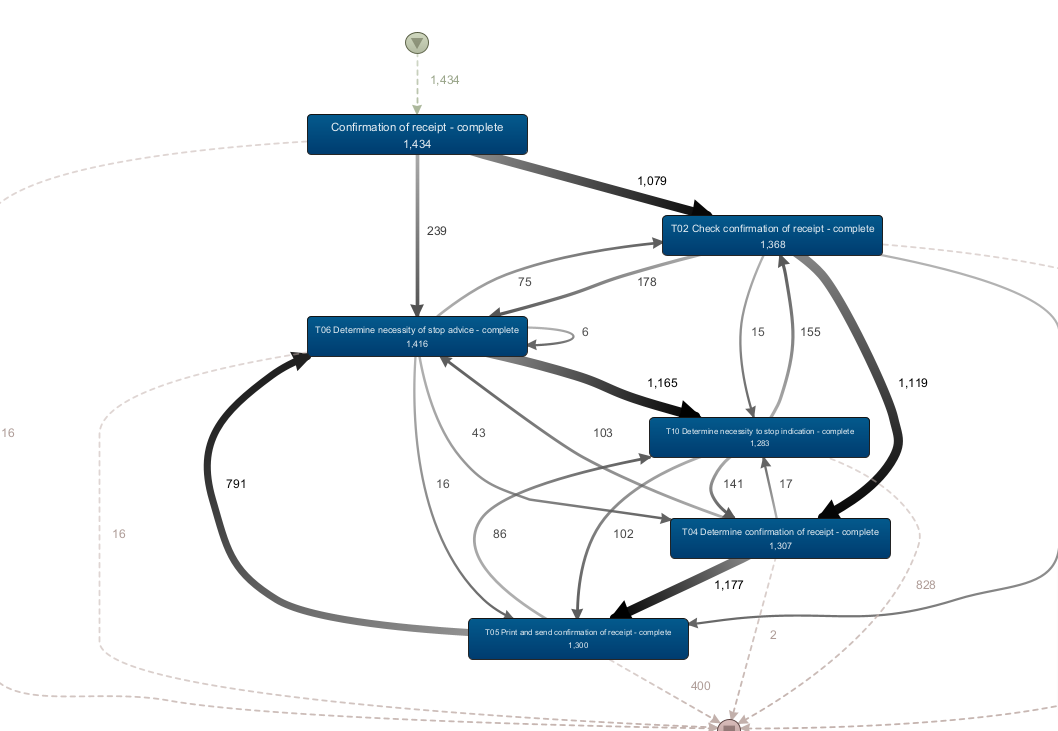
My observation from the graph is that it has lots of 0 event hours(Hours which have recorded zero events or Zero events for the given hour) . There are also a few outliers which appear randomly. The majority lies between 0 to 6 events per hour.

Using the map view, change the activity and path detail settings in order to create a comprehensible process map (e.g. a process map that could be printed on one A4 or letter paper or shown on a single computer screen while still being readable in full).

Discuss this process map, what is the main process?

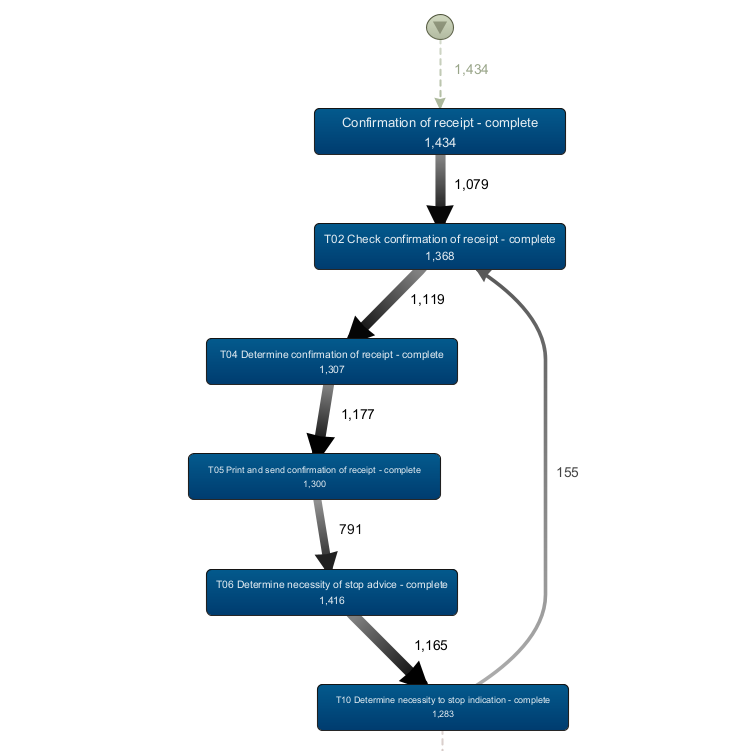
Adjust activities and paths sliders to get the below figure.

T02 - Check confirmation of Receipt(based on number of repetitions and connections) - Complete is the main process



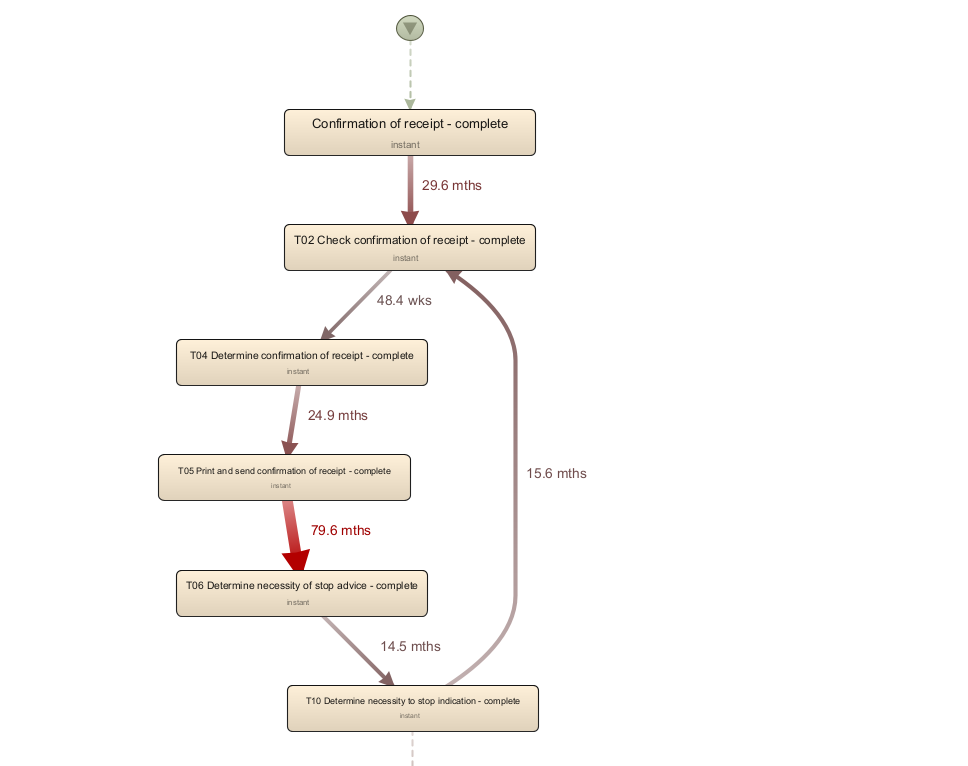
Which activities and paths between activities are frequent?

T02, T04, T05, T06 and T10 are frequent. Evident from this picture



After this performance projection.

Discuss where the process takes most time, e.g. where there are possibilities for improvement. Relate these times (of the bottlenecks) to the time spent in other parts of the process. In other words, discuss how severe the bottleneck is with respect to the time spent on other activities.

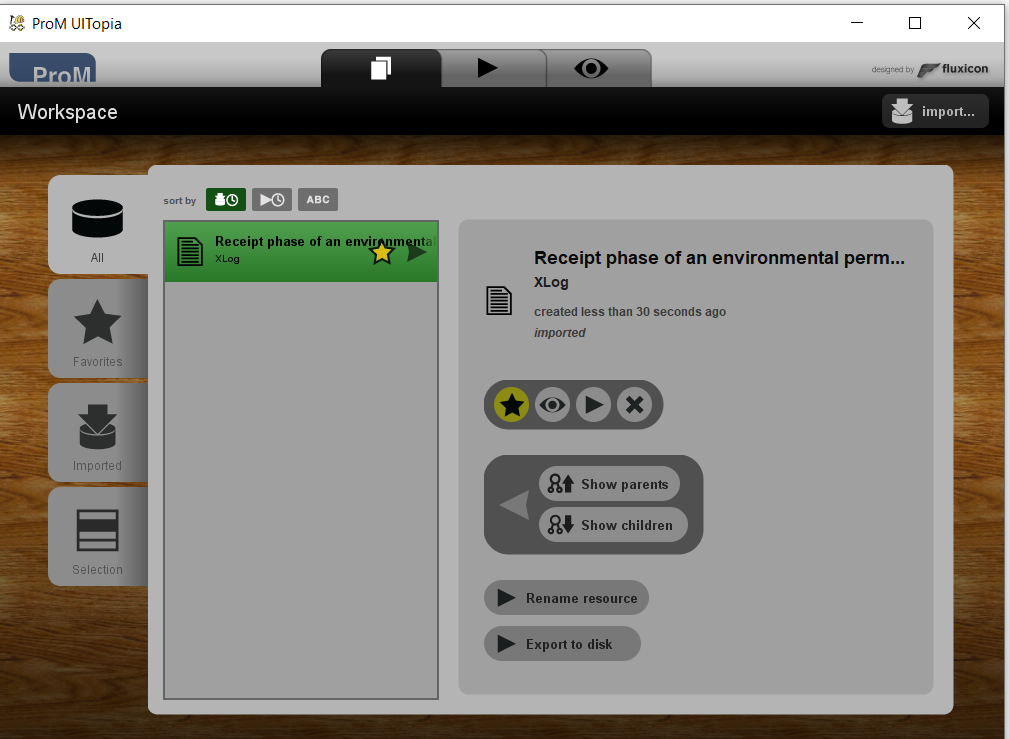


T05 - To6 is the issue which causes a delay of 79.6 months. This is a major bottleneck which avoided could increase the productivity by many folds. T06 - T10, T10 - T02 have comparatively little time because of the issue with T05-T06. This also causes staling of other process indirectly.

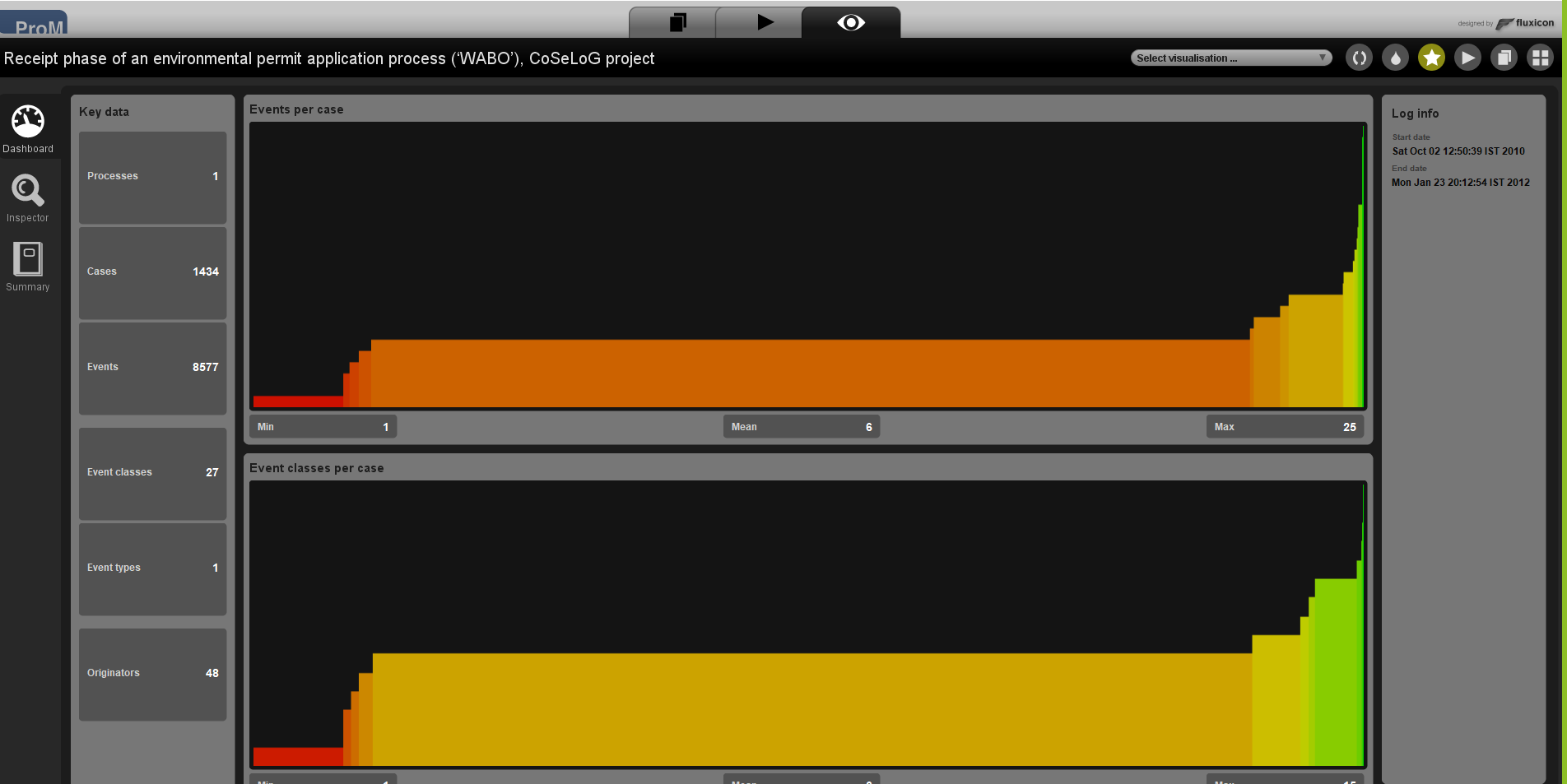
Also explicitly mention the performance metric chosen (e.g. total, mean, median, or max) and why you have chosen this setting.

The performance metric I chose is Total Duration to know the actual time taken for each and every process. THis would be an ideal way to calculate the discrepancies.

ProM

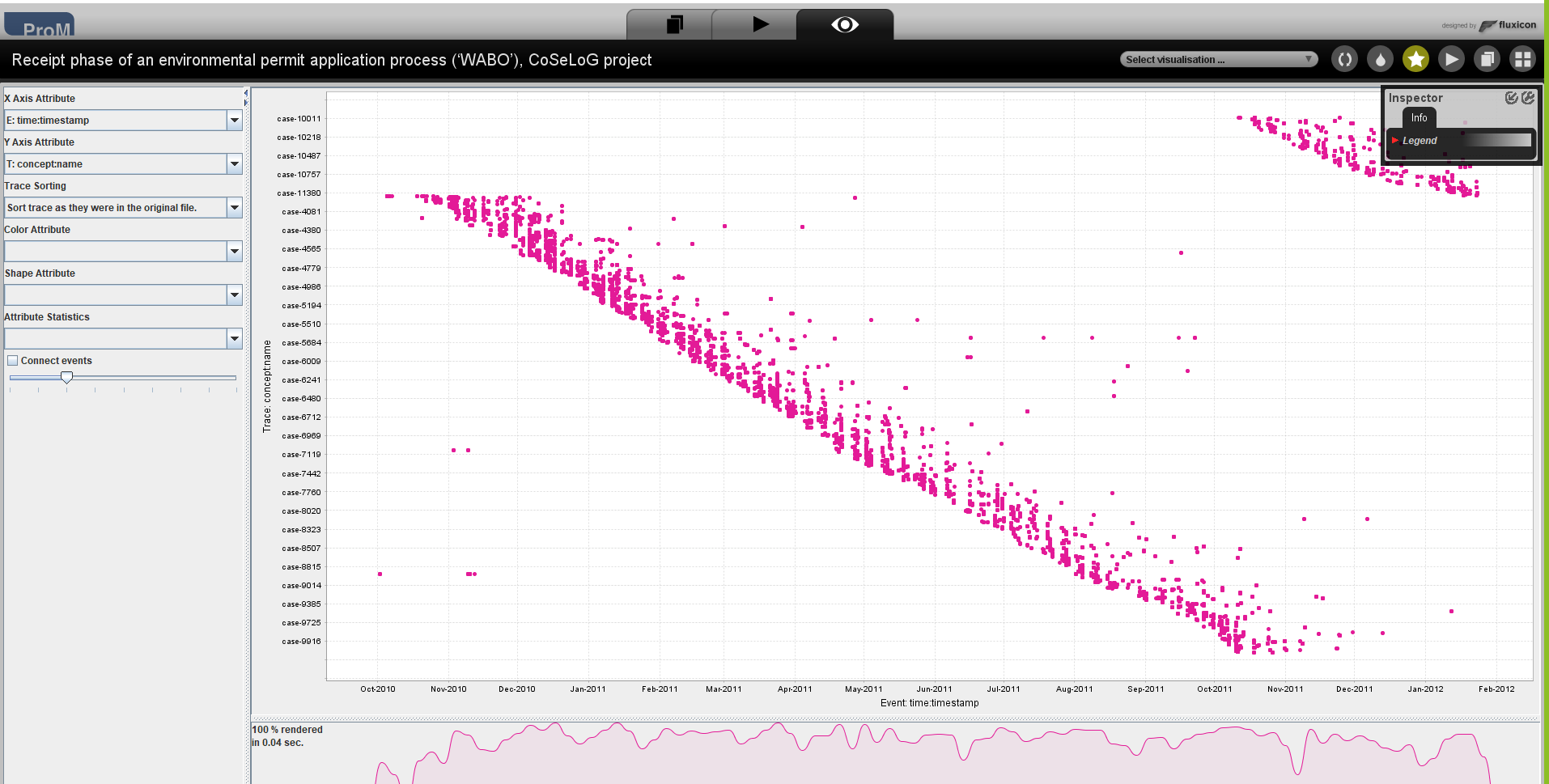


Click on eye icon which is near the star



Select the select visualisation option on top right

Click on dotted projection



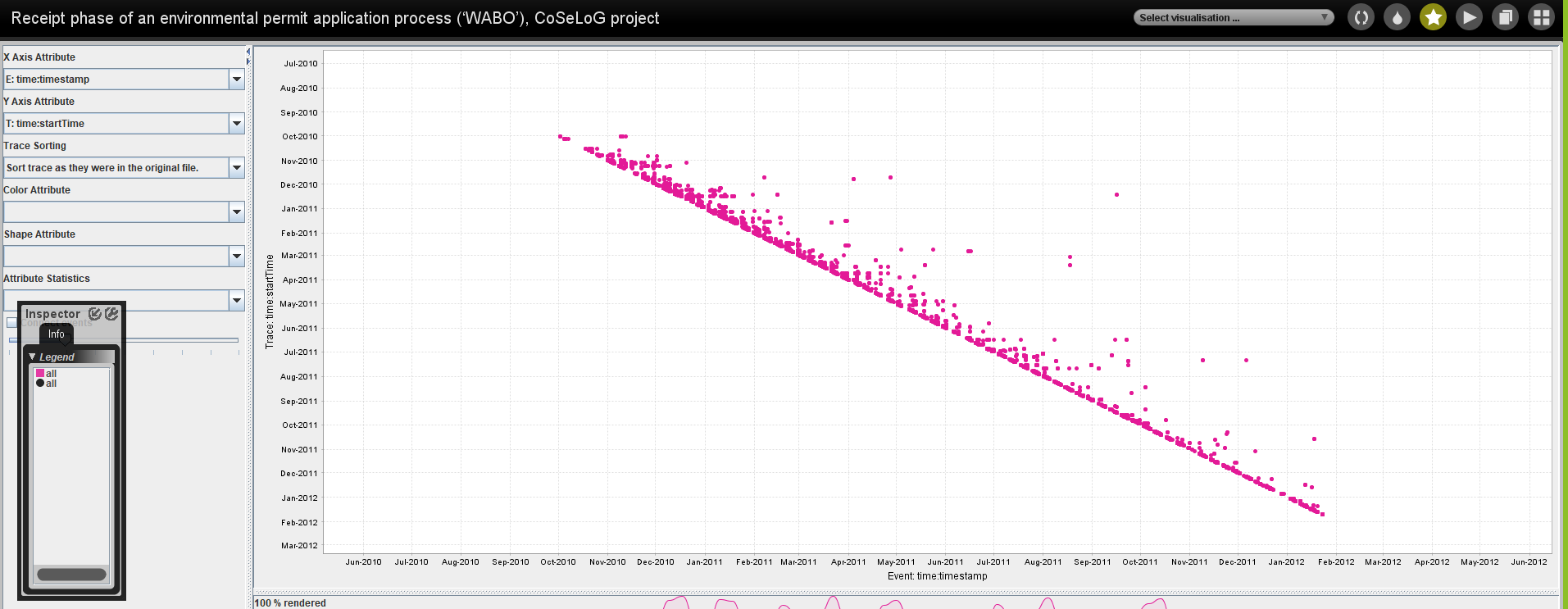
Using the Dotted Chart, answer the following questions:

1. Is the arrival rate of new cases constant? If not, when are there fluctuations? If yes, how can we see this from the Dotted Chart?

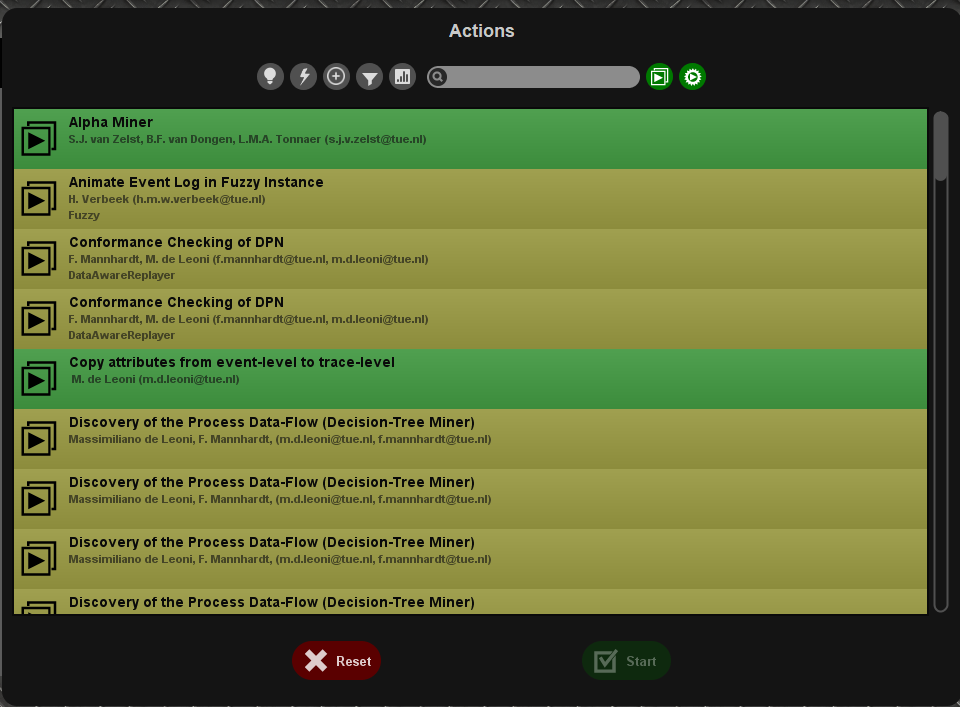
As we can see a pattern of a crossed line which is gradual the cases are constant without any fluctuation.

1. Can you observe a change in the global process?

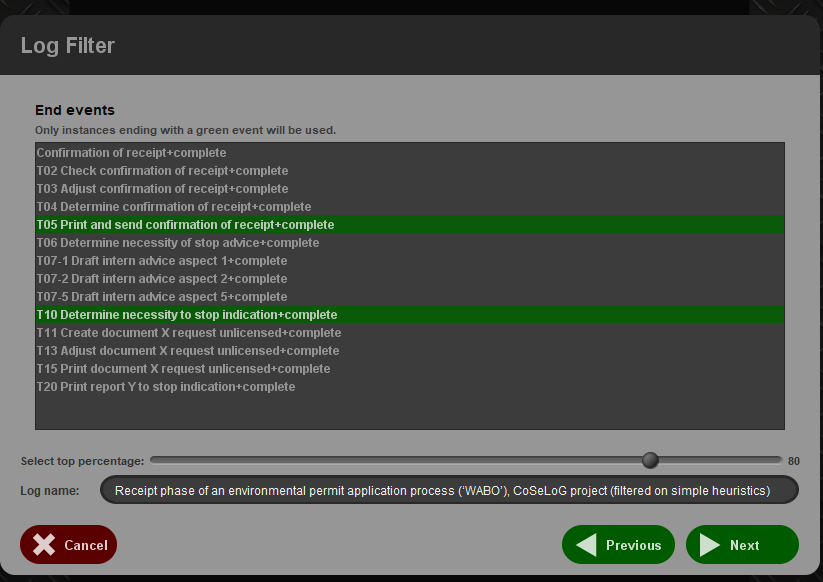
Since there is only one colour we can declare that there is no change in global process.



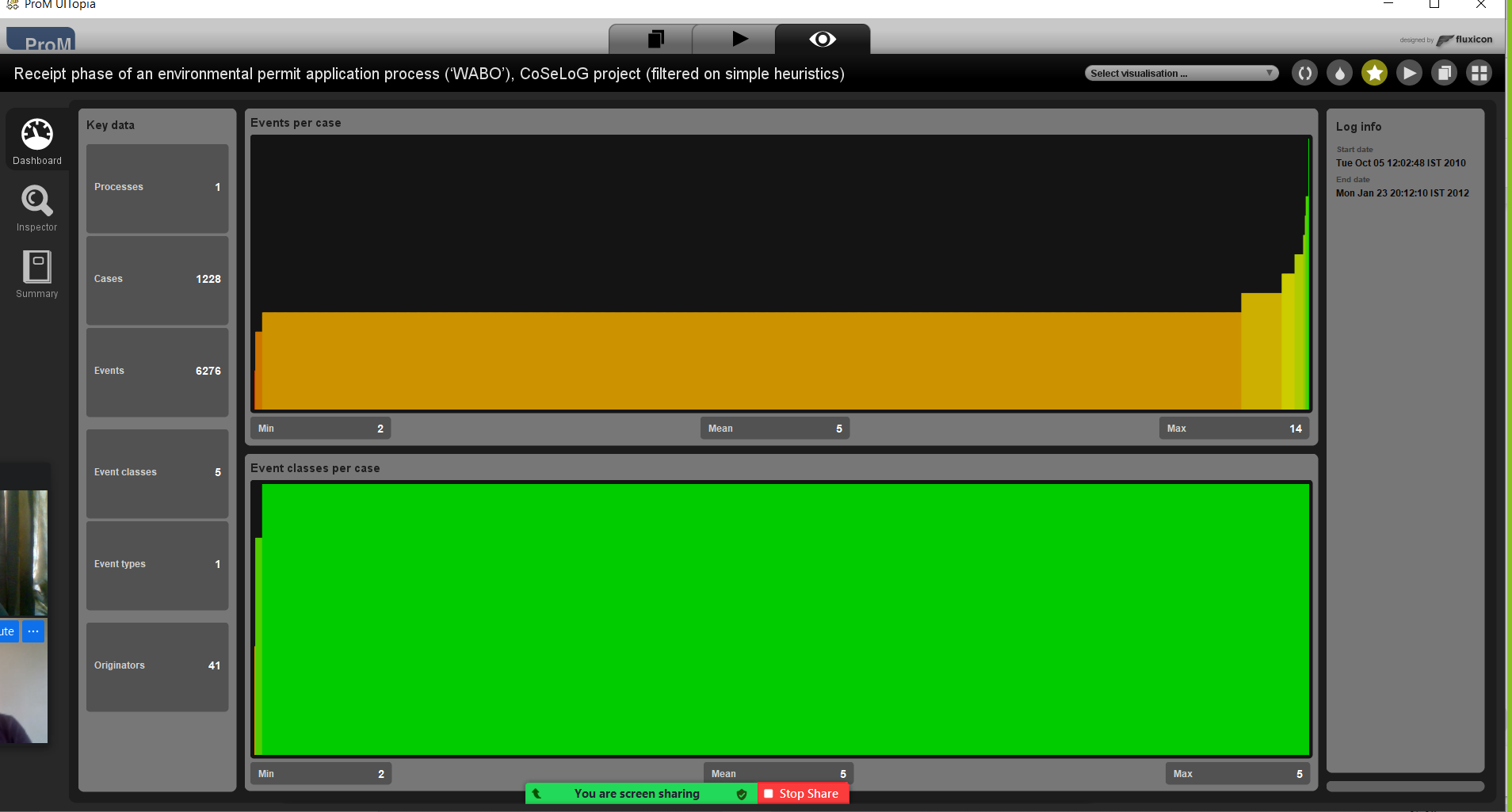
To get this line change Y axis attribute to t:time:starttime. Since we need to check whether the new cases are constant we have to plot against start time



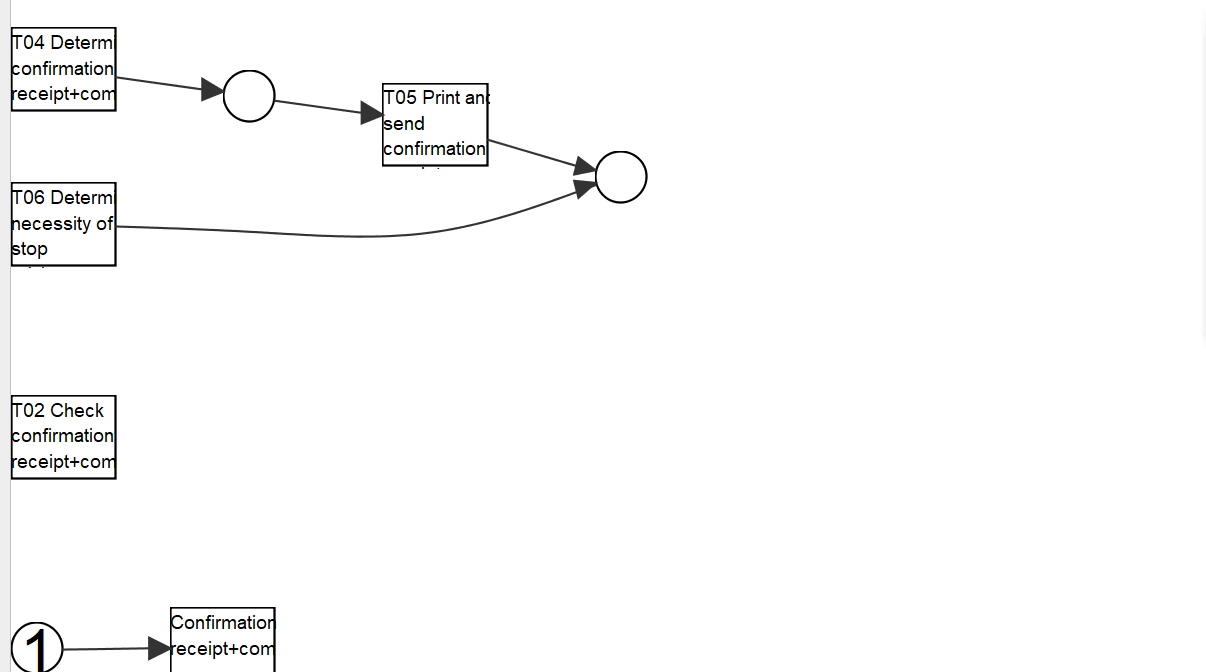
After importing click on the play and the above dialogue block pops out. In this select “Filter log using simple heuristics”.



We will get this dialogue box.

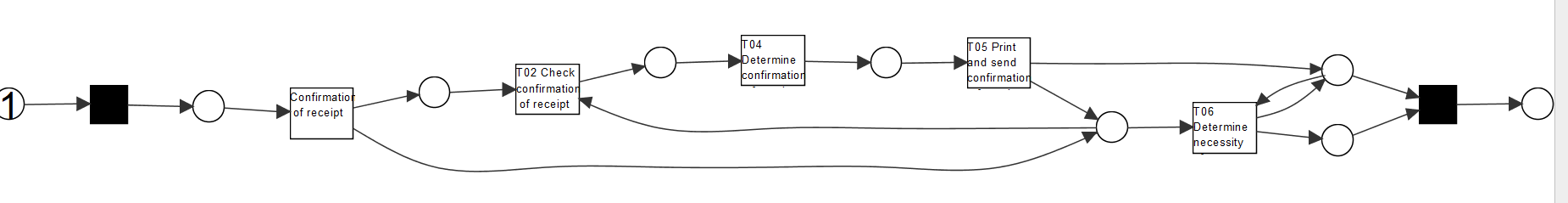


Then select alpha miner in the files option.

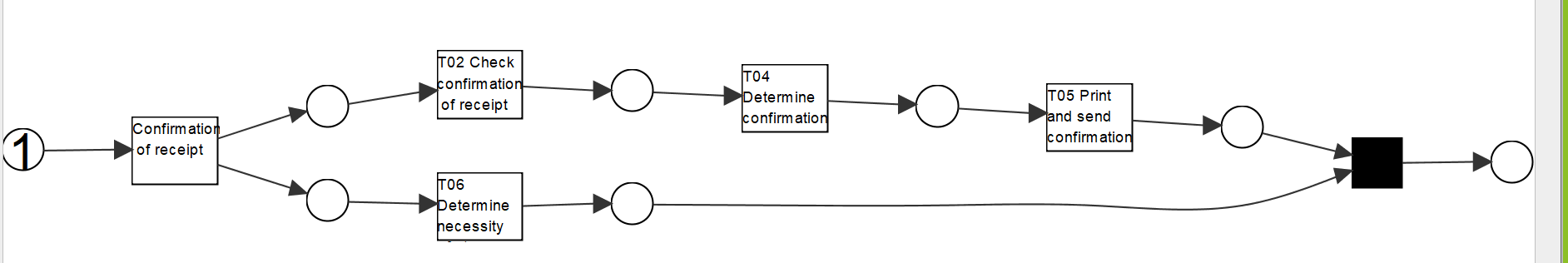


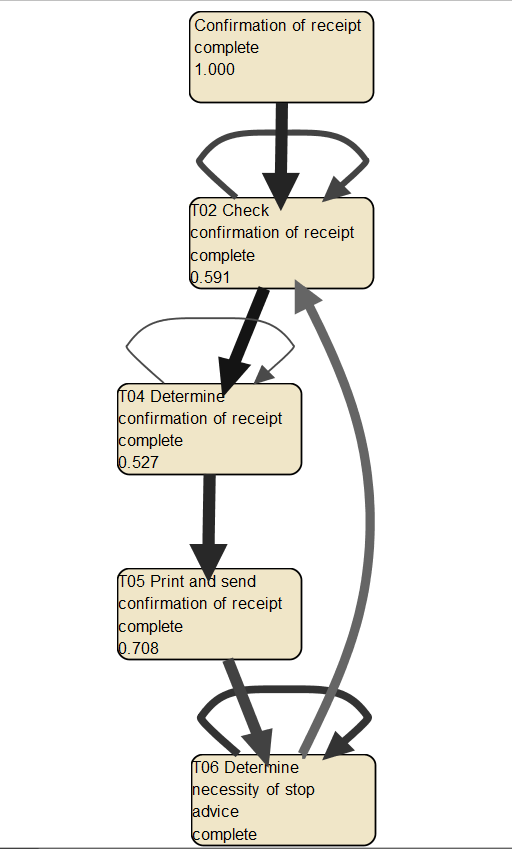
ILP based process discovery(Express)

ILP stands for integer linear program.



Mine petri net with inductive miner





Mine for a fuzzy model

You are now asked to discover a Petri net on the event log. However, the unfiltered event log results in an incomprehensible Petri net. Therefore, you are allowed to run the 'Filter log using simple heuristics' plug-in once on the original event log to discover a Petri net on the filtered event log.

There are different algorithms which take the csv file and output the model. Like alpha algorithm, Inductive Miner, ILP based process discovery and Fuzzy model which have been discussed above.

Clearly indicate which settings you have used for the 'Filter log using simple heuristics' plug-in.

Explicitly motivate the filtering settings chosen, why did you pick this percentage or selection of activities?

‘

The settings which I have used for are the default. I changed the data size to entire data

Discuss and argue which plug-in (or chain of plug-ins) you have used to discover a Petri net, for instance by comparing two or more plug-in results and arguing why one of the Petri nets is better.

The screenshots are pasted above.

Alpha Miner is not well connected.

ILP and Inductive Miner both are well connected. But ILP is elaborate

Fuzzy model talks about the main parts.

Explain the (best) Petri net: what is the main process and what are notable parts of the Petri net?

The main process is T02 check confirmation of receipt

The notable parts of the petri net are

T02 check confirmation of receipt

T04 Determine confirmation

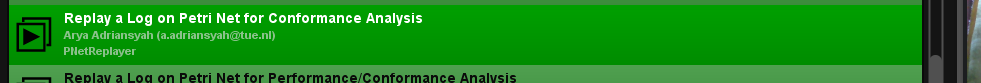
T05 Print and send confirmation

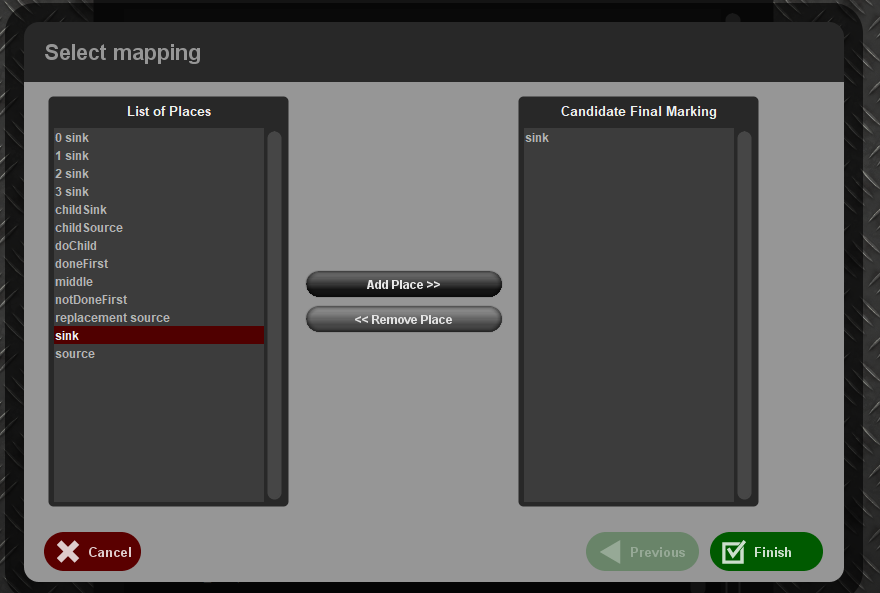
T06 Determine necessity

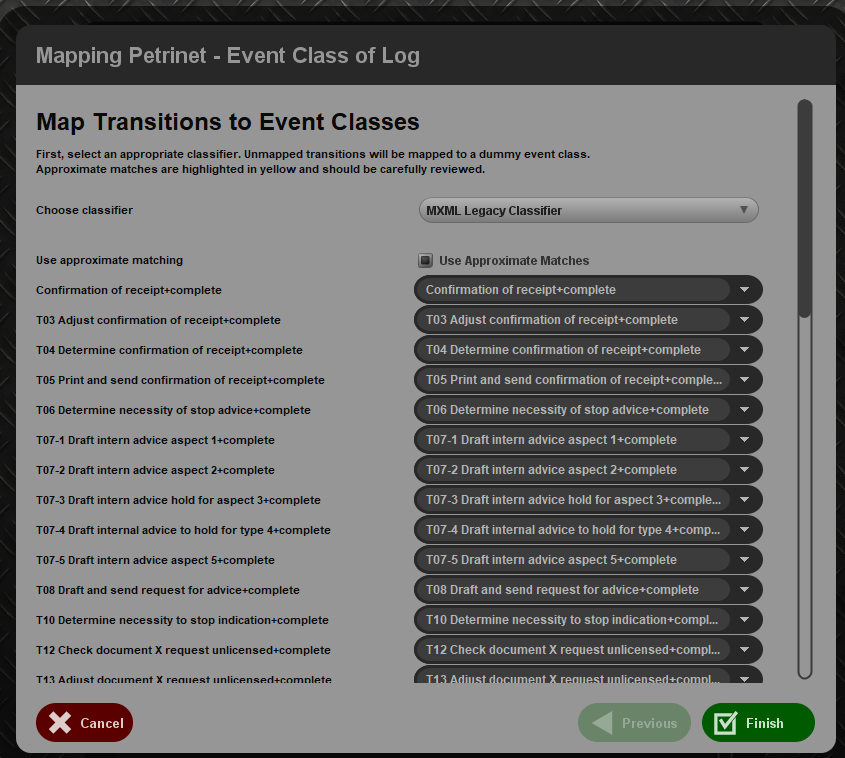
I think the best petri net is ILP because it well connected, precise and self explanatory

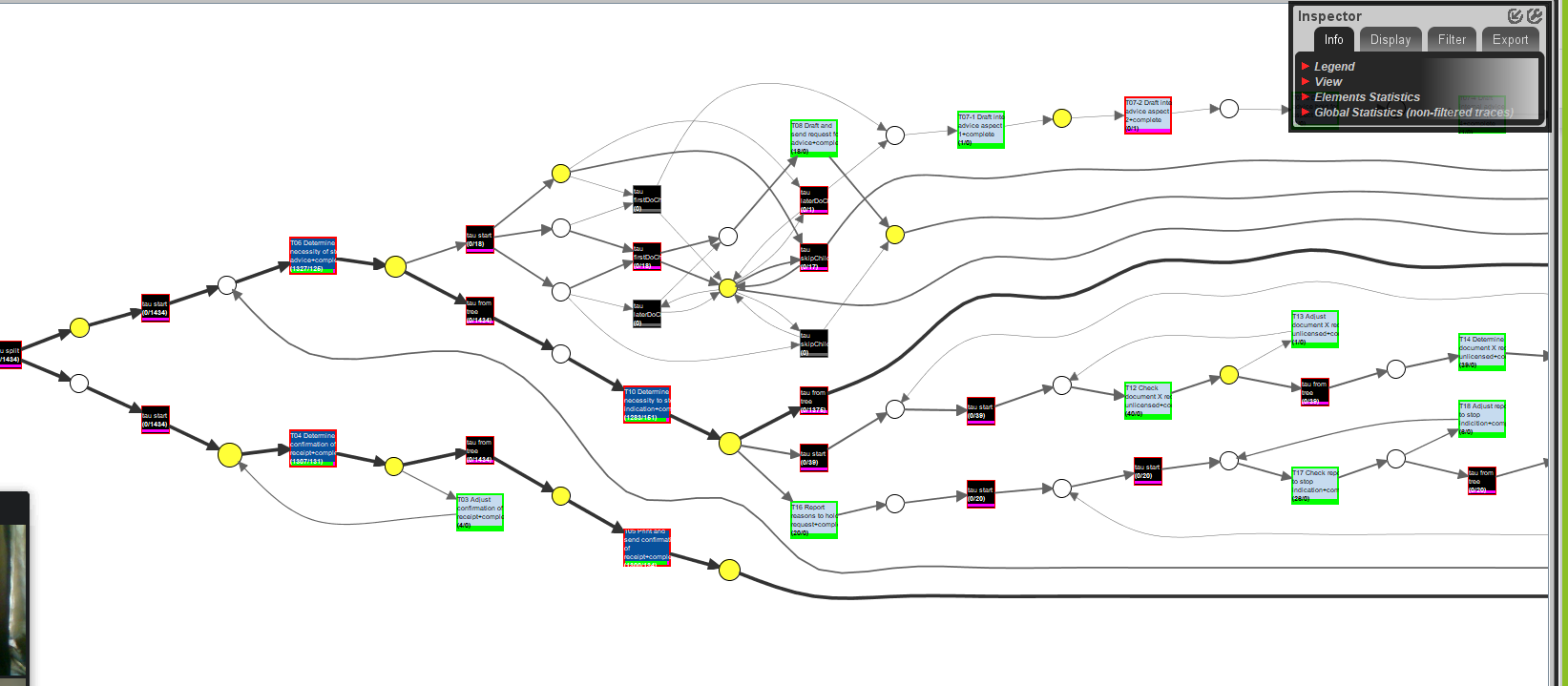
The organization has a process model that describes the 'should be' process (i.e. a normative process model). Load the file 'normativeModel.pnml' into ProM and apply conformance checking on this process model, and on the full unfiltered original event log.

Include a screenshot of the part of the normative process model, with the conformance information projected onto it, that shows where most of the deviations occur.





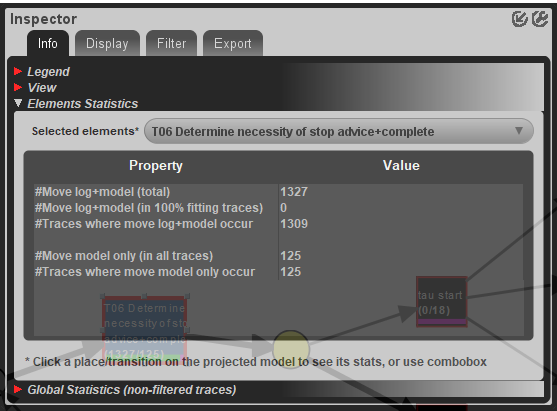
Final answer is below



What is the replay fitness (the 'trace fitness' statistic) of the event log on the normative process model?

Whether the resulting model can reproduce the same events in the log is called replay fitness

Select the transition 'T06 Determine necessity of stop advice+complete' (on the top left of the model) and discuss its element statistics: how many times is the transition executed correctly and how many times incorrectly?



Element statistics of T06. Click on T06 and look for element statistics in the inspector

Correct execution - 1309

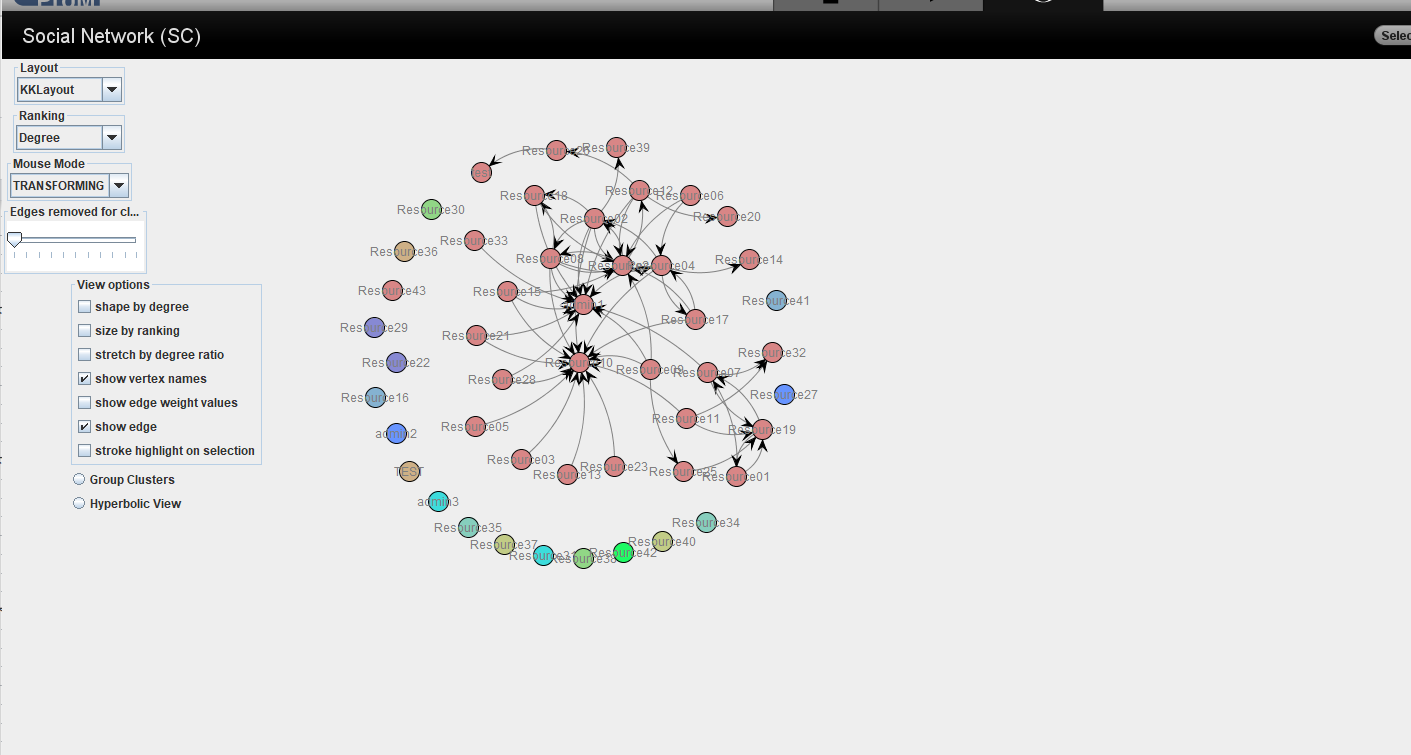
Incorrect execution - 125

Using the element statistics of transition 'T06 Determine necessity of stop advice+complete', what can you say about the (in)correct execution of this activity?

Out of 1327 only 1309 the log+model occurred. That’s why it is in red color, All the green ones like T03, T08, T16 have full completion.

The final analysis you have to perform on the original event log is a resource analysis, e.g. looking at the user behavior in the event log.

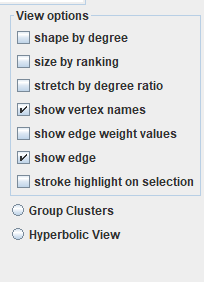
1. Use the plug-in 'Mine for a Subcontracting Social Network'. Note that subcontracting means that if individual *j* frequently executed an activity in-between two activities executed by individual *i*, then individual *i* subcontracted work to individual *j*. Answer the following question using this view: Can two or more groups of users be distinguished?



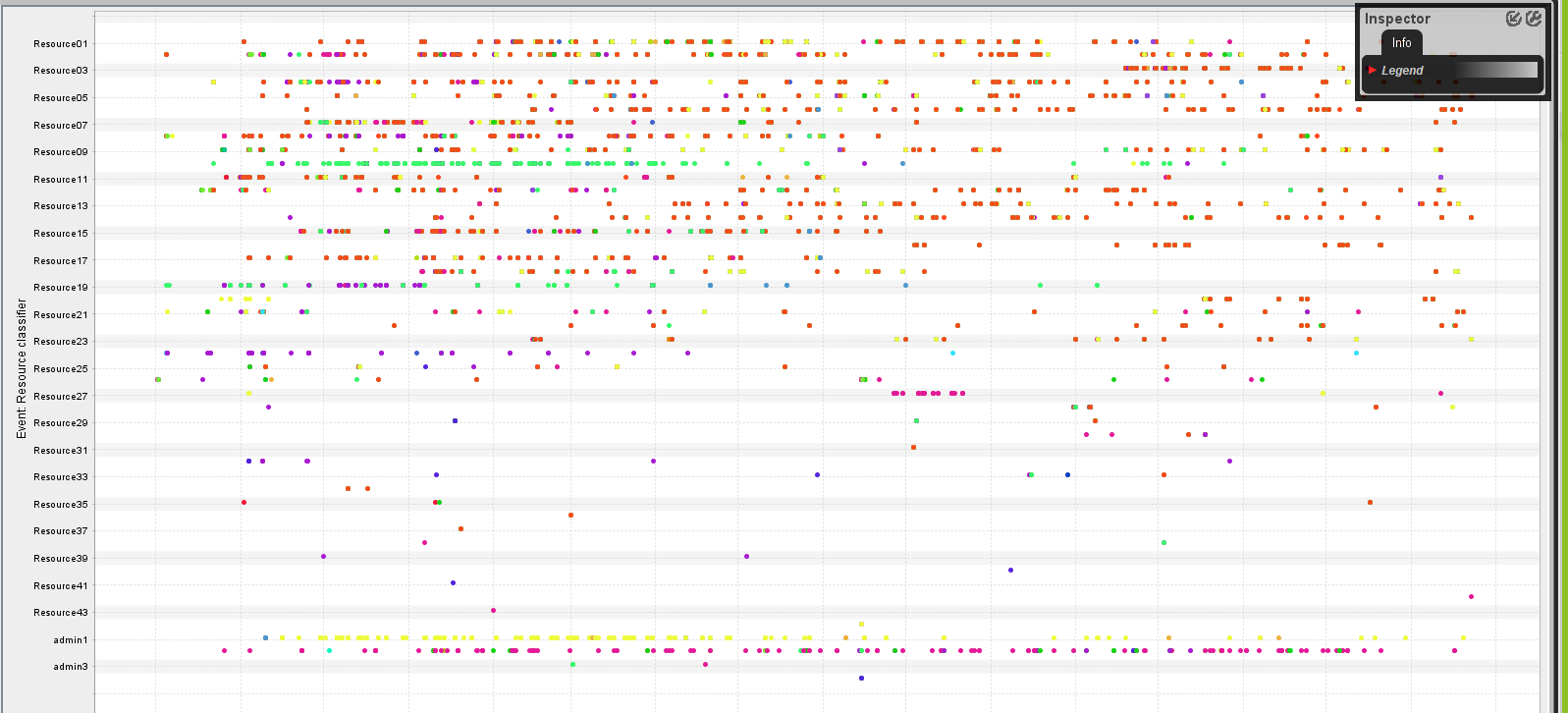
As seen from the above image everything can be distinguished clearly

1. Explicitly discuss the settings you have used in the resulting visualization.

Default settings was used. We can alter the below table to make changes in the settings



1. Again use one of the two Dotted Chart plug-ins. For the XDottedChart change the component type to 'org:resource'. If you use the Dotted Chart visualizer change the 'Y Axis Attribute' to 'C: Resource classifier' and the color attribute to 'C: Activity Classifier'. Answer the following two questions using this view:



1. Are all users executing activities from the start of the event log, or are some users joining later?

As we can see from the screenshot above they are joining in different points. We can see that with respect to x axis

1. Are users mainly executing particular activities or are most users executing most of the activities?

Most of the users are executing most of the activities. For a given resource people are in the same line