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

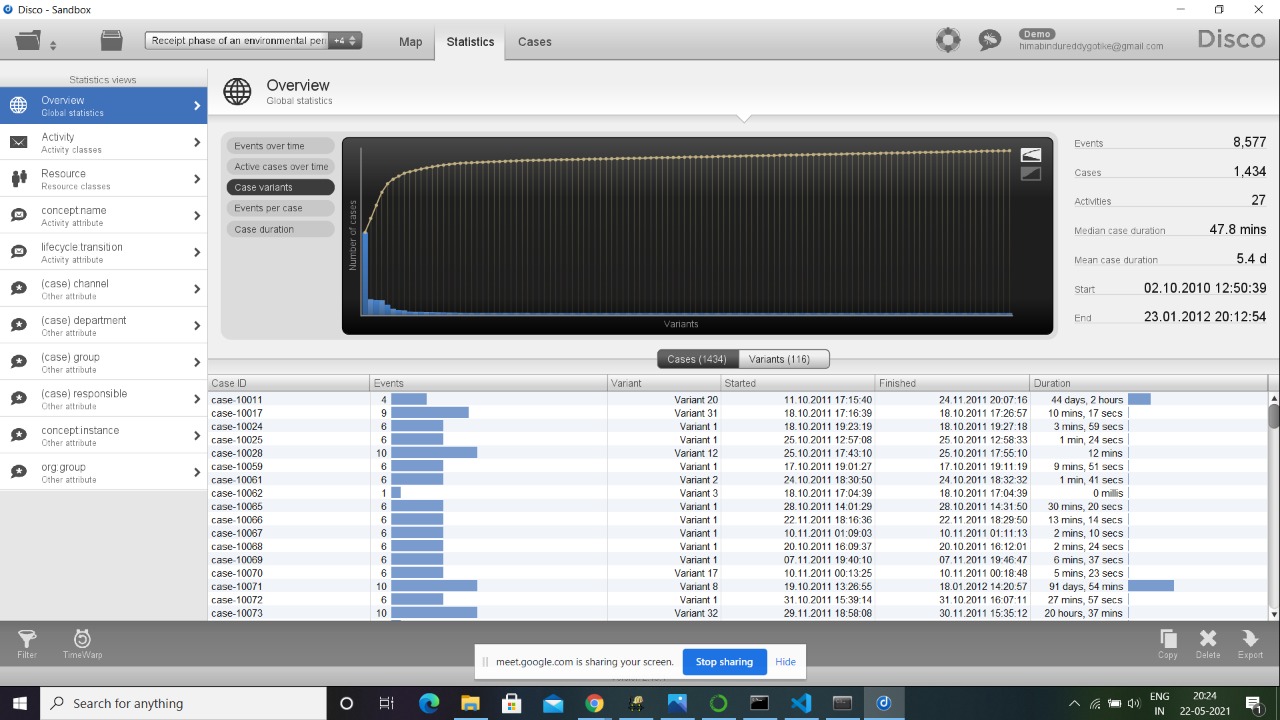
**Without switching to other views, use the statistics view to answer the following three subquestions:**

1. **How many events are there on average per case?**
2. **Can you indicate whether each case seems to be unique or whether many cases follow the same activity sequence?**
3. **What is the main observation that can be made from the 'Events over time' graph?**

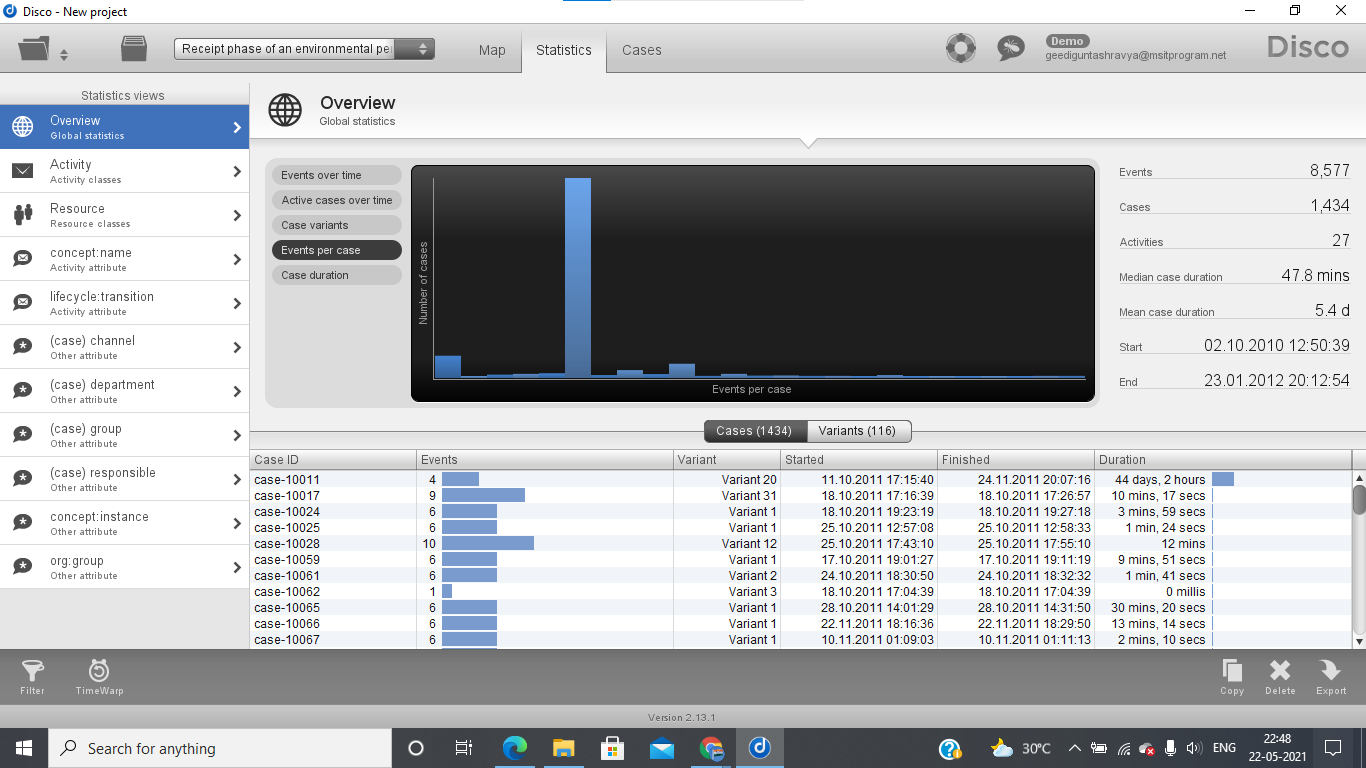
After loading the event log to Disk, I switched to the Statistics view and select Events over time and export the table to csv to perform statistical calculation of the average of events per case.

set the activity details slider to 100% (the highest) and the routes slider to 0% (the lowest). Shows all activities but only the most frequent routes between them. Based on the data table the following results were obtained

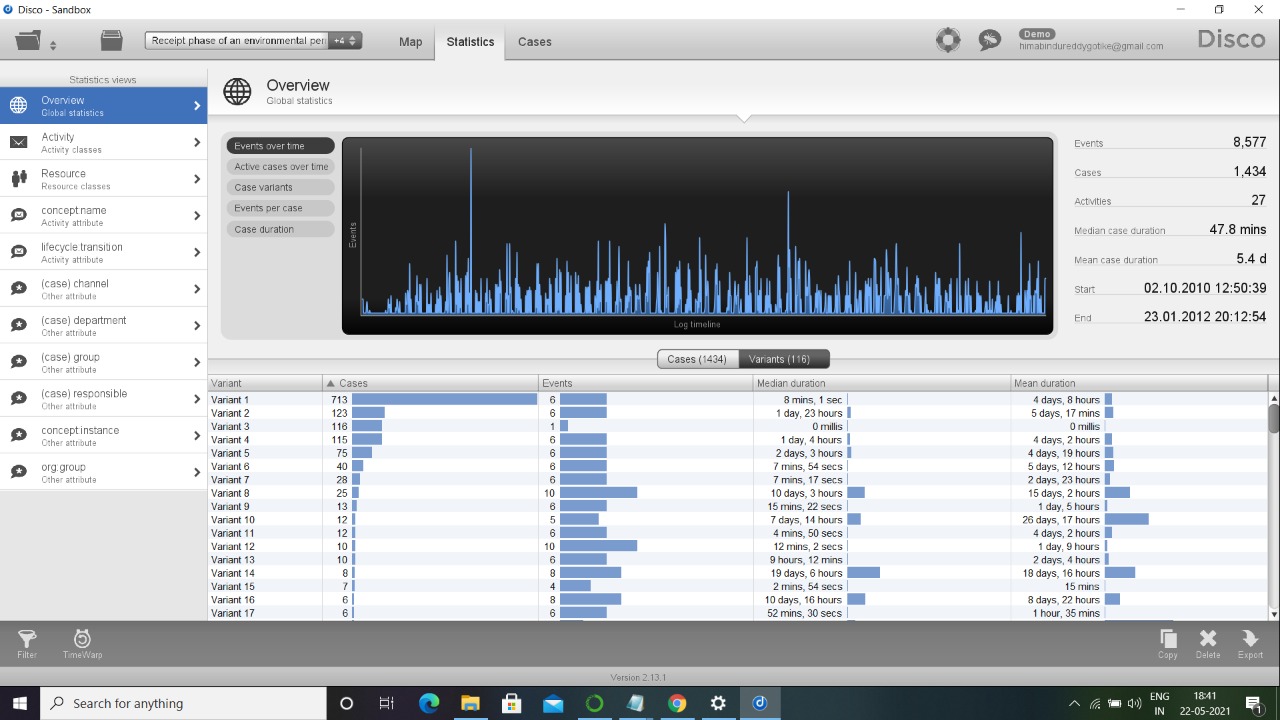
1. There are 5.98 events on average per case.



2. There are 116 variants where the cases follow the same sequence of activities.



3. At least 0 events are held per day and a maximum of 147.39 events per day



There are days in an interspersed way where some days have zero events and other days it has more than one event

**While still in Disco, switch to the 'map' view to display a process map.**

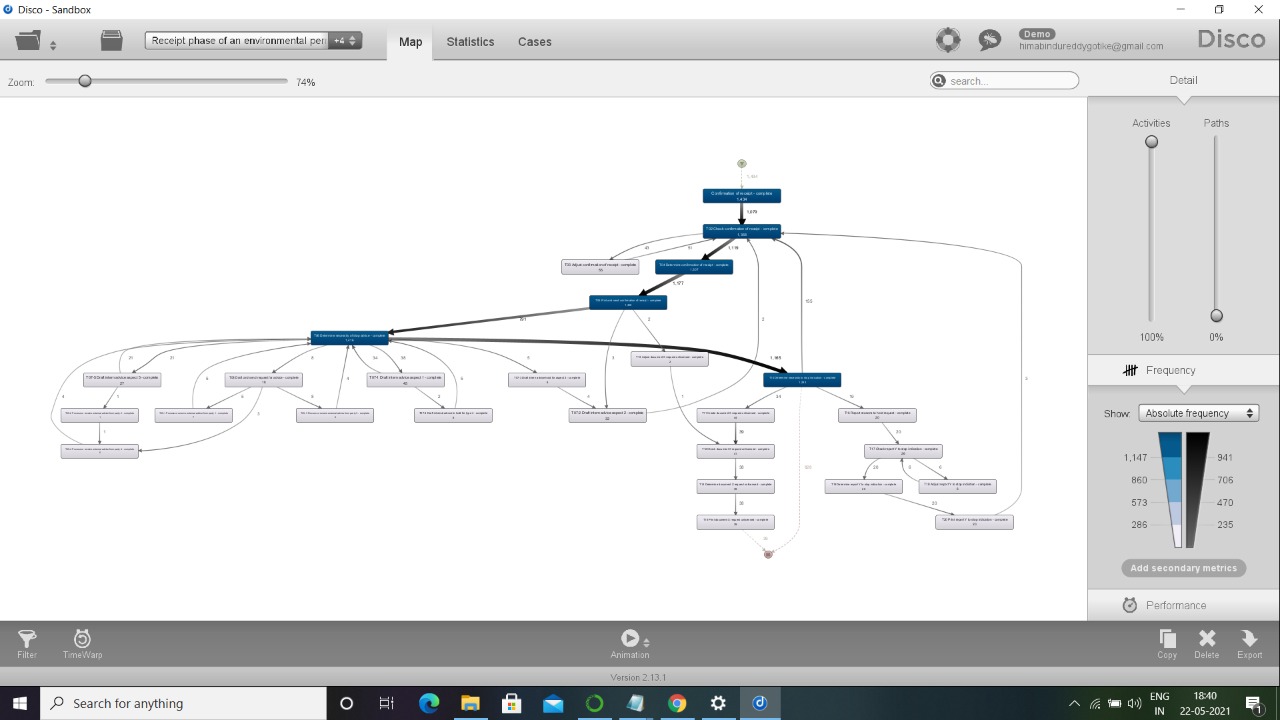
**Using the map view, change the activity and path detail settings in order to create a comprehensible process map (e.g. a process map that could be printed on one A4 or letter paper or shown on a single computer screen while still being readable in full).**

1. **Discuss this process map, what is the main process?**
2. **Which activities and paths between activities are frequent?**

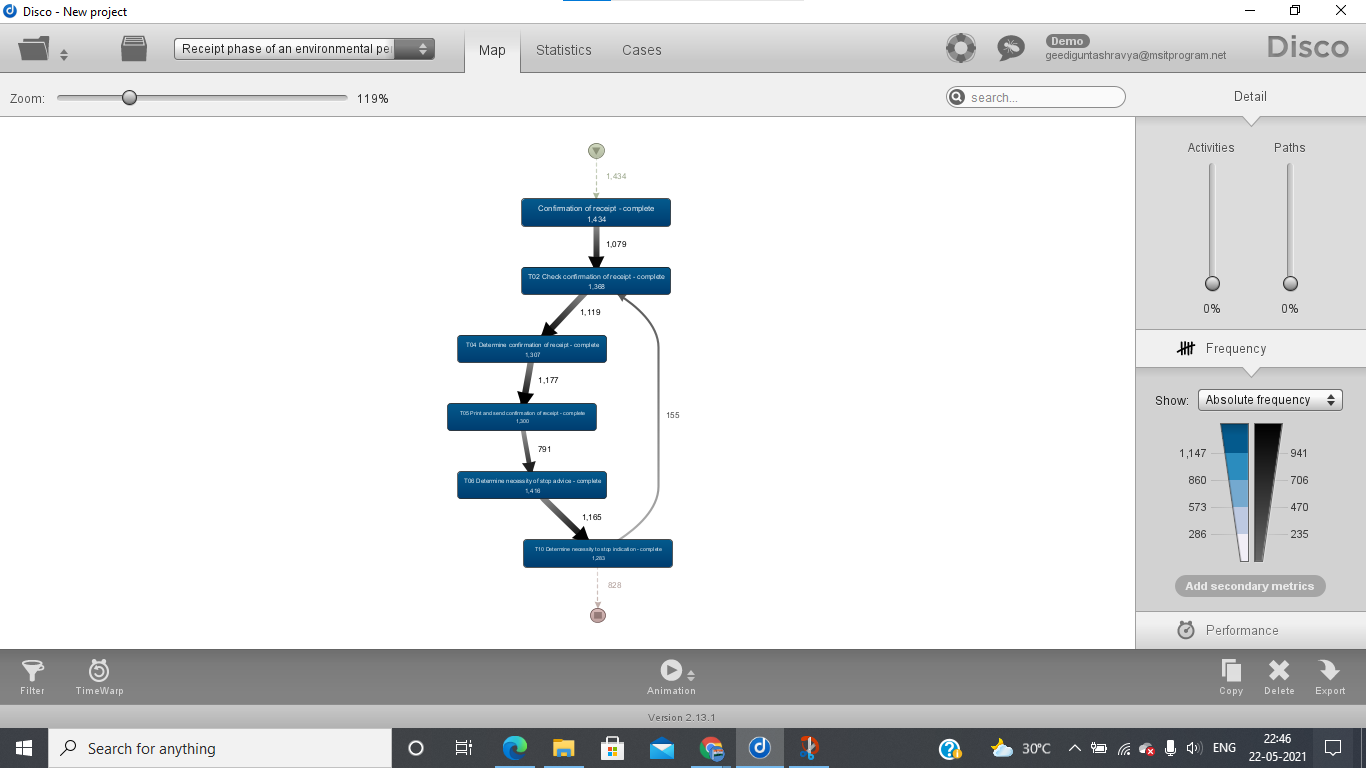
**In your answer, include the settings you used for both the activity and path sliders.**

Then I switched to the ‘Map’ view in ‘Frequency’ detail and set the activity details slider to 20% (highest) and the routes slider to 0% (lowest). Shows all activities, but only the most frequent routes between them.

1. The a process map with 11 activities and 13 arcs that connected them. The process map was clearly legible. common activities are a darker blue (6 activities) and the most frequent paths had thicker arrows (5 arcs).



2. Based on the process map, can conclude that the main path consists of the sequence ‘Acknowledgment of receipt’-> ‘T02 Verify receipt receipt’-> ‘T04 Determine receipt receipt’-> ‘T05 Print and send receipt confirmation’ ,this is presented with three different paths infrequent in the activity ->’T10 Determine the need to stop the indication’ , one returns to ->’T02 Verify receipt confirmation’; two are followed by less frequent activities in the following sequence->’T11 Create document X request without license’->’T12 Verify document X request without license’ ->’T14 Determine request for document X without license’->’T15 Print document X request without license’-> ‘Terminate’ and three directly->’Finish’



**While still in Disco, and while using the same process map (e.g. do not change the activity and path settings), switch to the performance projection.**

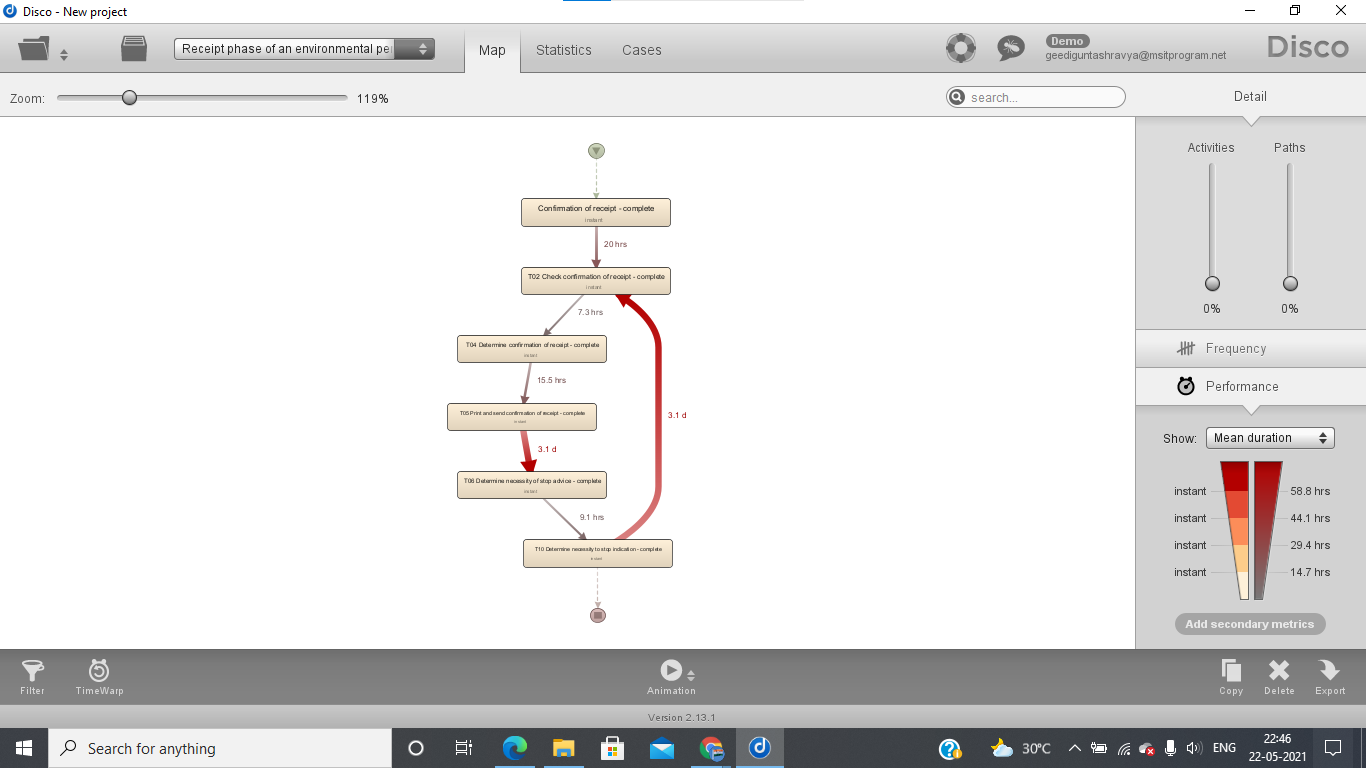
**Discuss where the process takes most time, e.g. where there are possibilities for improvement. Relate these times (of the bottlenecks) to the time spent in other parts of the process. In other words, discuss how severe the bottleneck is with respect to the time spent on other activities.**

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

Then I switched to the Map view in Performance detail and set the activity details slider to 20% (highest) and the routes slider to 0% (lowest). Shows all activities, but only the most frequent routes between them.

The time that takes the longest on the path that takes the longest with respect to performance is between the activity 'TO5 Print and send acknowledgment of receipt - >'T06 Determine the need for a stop warning' that has an average duration of 3.1 days. It also has the same duration in case of returning from the activity T10 Determine the need to stop the indication TO2 Verify 14124 receipt confirmation tod due to this situation the duration of the process is lengthened more for each time the loop it is tour. Therefore, it creates a large bottleneck.

The performance metric chosen is the average duration, in order to average the times between activities. In order to have an average of the average duration of the environmental permit application process.

****

**Now load the original event log in ProM. Visualize the event log using the Dotted Chart or XDottedChart visualizer (by pressing the 'eye'-icon with the event log selected and switching to the Dotted Chart or XDottedChart visualizer).**

**Using the Dotted Chart, answer the following questions:**

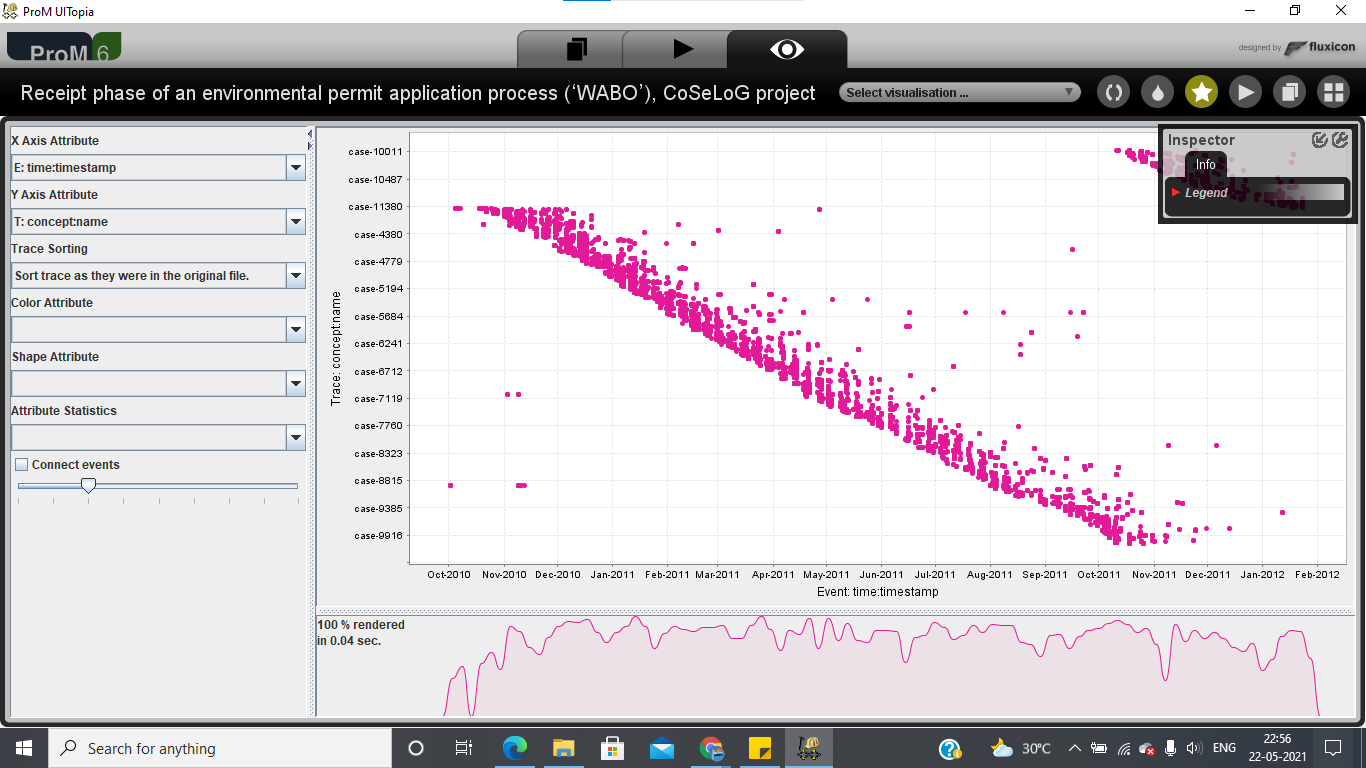
1. **Is the arrival rate of new cases constant? If not, when are there fluctuations? If yes, how can we see this from the Dotted Chart?**
2. **Can you observe a change in the global process?**

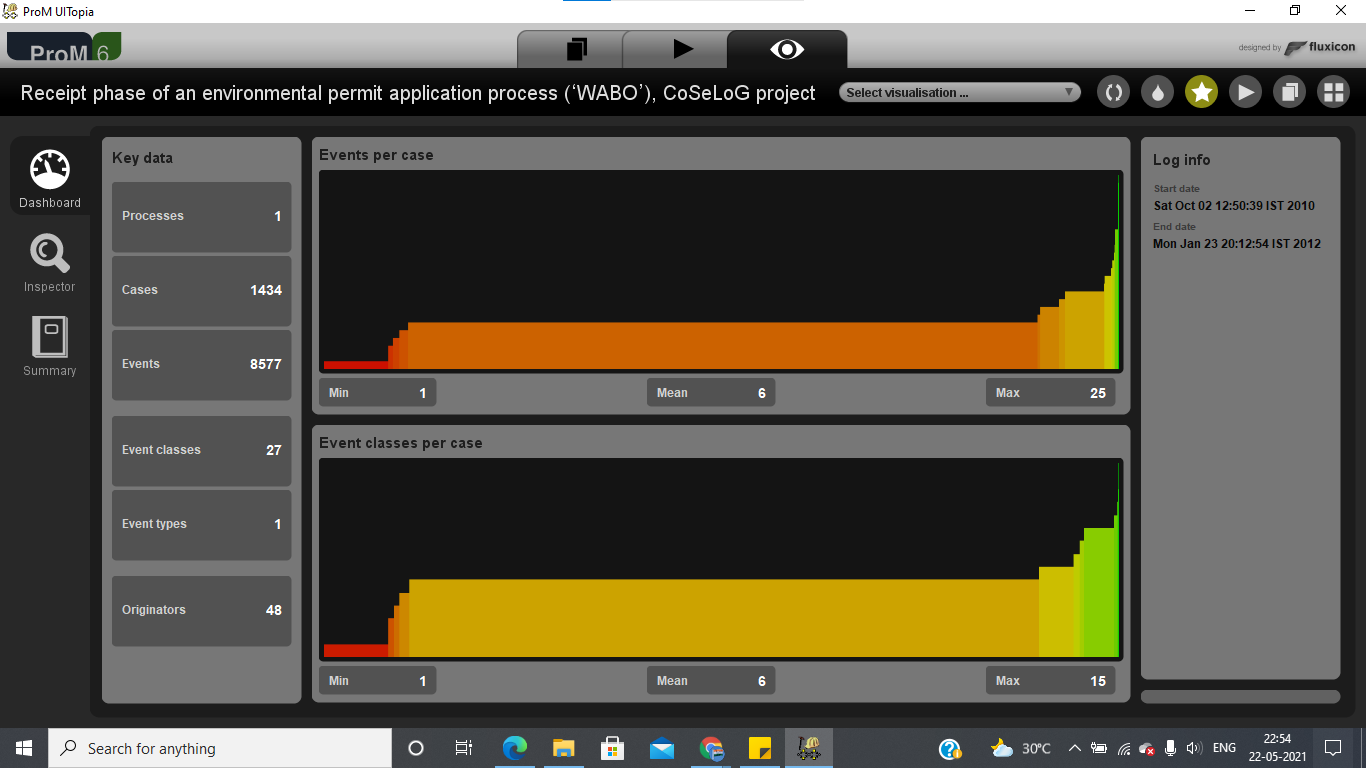
**Note that you don't need to change the component, time or coloring settings. You can however re-sort the traces on the time of the first event, and zoom in or out if you want.**

**The Dotted Chart is explained in lecture 4.8: 'Exploring Event Data'.**

The event log was imported and the event log was viewed using 'Dotted Chart (LogProjection); where (x) axis is time. (y) axis is cases, and the colored dots are the sequence of events that have happened.

1. It tends to be a straight diagonal line with which the arrival of cases can be considered constant. However, there are about three fluctuations in June, September and December. The way to see this in the graph is to draw a diagonal line in the time of arrival of the cases which varies.





2. A global change is observed due to a gap where the points are grouped into two groups, where the cases are happening at different times, which is due to the order of the traces and therefore of the cases.

**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.**

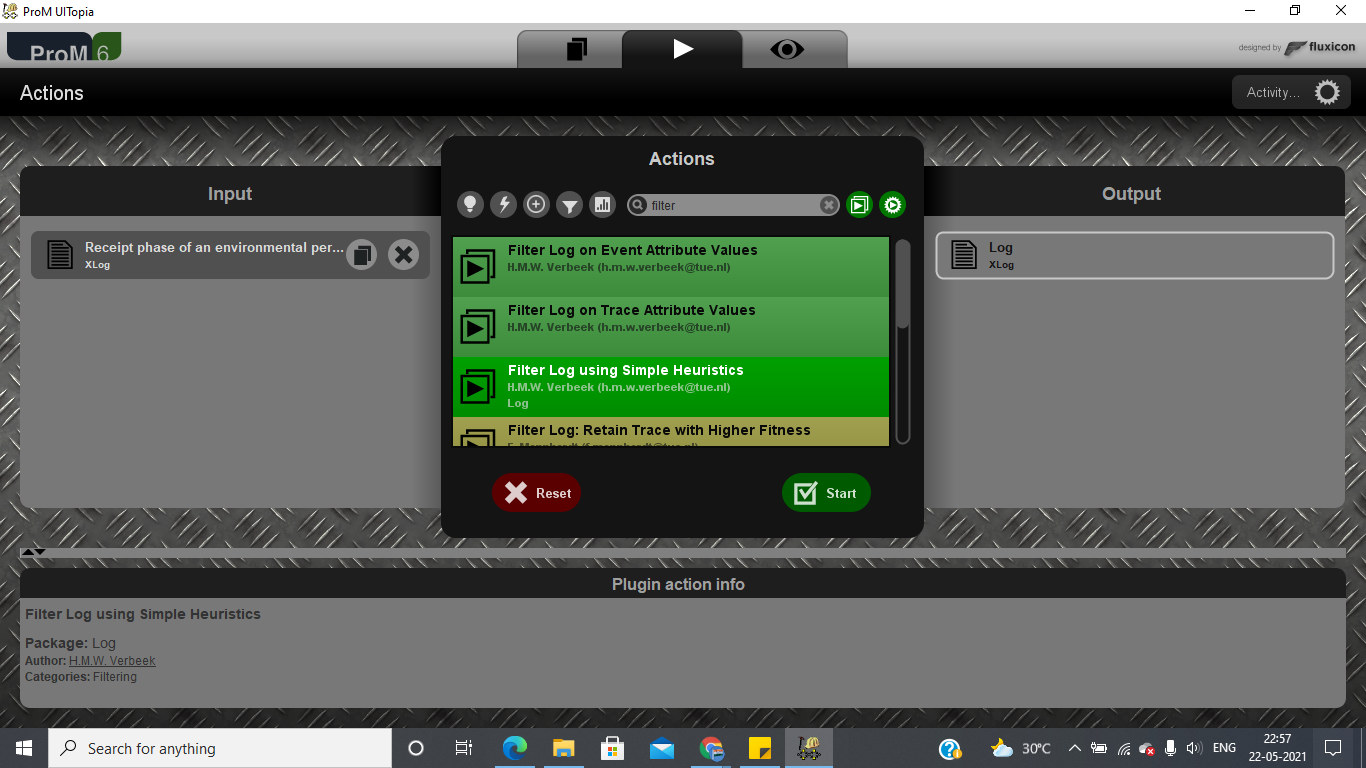
1. **Clearly indicate which settings you have used for the 'Filter log using simple heuristics' plug-in.**
2. **Explicitly motivate the filtering settings chosen, why did you pick this percentage or selection of activities?**
3. **Discuss and argue which plug-in (or chain of plug-ins) you have used to discover a Petri net, for instance by comparing two or more plug-in results and arguing why one of the Petri nets is better.**
4. **Explain the (best) Petri net: what is the main process and what are notable parts of the Petri net?**

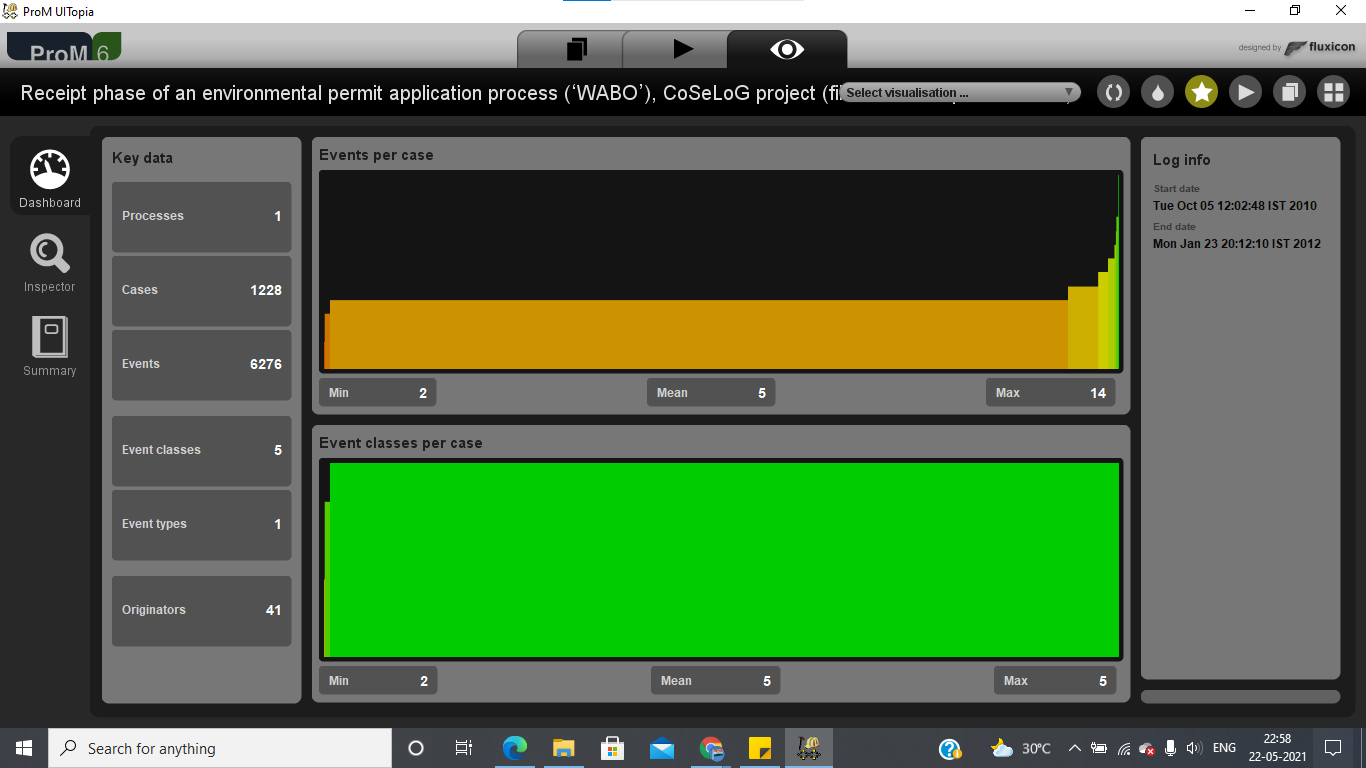
**Note that this question requires you to experiment with different filtering settings and discovery plug-ins. You are not required to describe *everything* you have tried but found unsuccessful. Only describe the successful combination of plug-ins and its result(s) and argue why your final result is 'good'.**

**Suggested list of plug-ins or plug-in chains to produce a Petri net:**

* **Mine for a Petri Net using Alpha-algorithm**
* **Mine for a Petri Net using ILP**
* **Mine for a Heuristics Net using Heuristics Miner *followed by* Convert Heuristics net into Petri net**
* **Mine for a Petri net with Inductive Miner**

1.When executing the action ‘Registry filter using simple heuristics', the following configuration has been used:1.1) Record filter: complete. 1.2) start events: 90%. 1.3) final events: 95%. 1.4) event filter: 95%.





2. The percentage of start of events was chosen as it is only a start event, which another value more or less will not have major changes. Regarding final events, the percentage was chosen because four possible events have been identified to finish. And the filter of events was chosen the percentage when identifying six activities that are the most frequent 'Confirmation of receipt. T02 Check confirmation of receipt. T04 Determine confirmation of receipt. TOS Print and send confirmation of receipt T06 Determine necessity of stop advice and T10 Determine necessity to stop indication’.

3. The plug-ins have been used: 3.1) 'Mine for a Heuristics Net using Heuristics Miner. 3.2) 'Convert Heuristics net into Petri net. And I compare them with the plugins: 'Alpha Miner'. "Mine for a Petri Net using ILP Mine Petri net with Inductive Miner. Therefore and ' 'Convert Heuristics net into Petri net has better described the model, unlike for example Mine Petri net with Inductive Miner' which provides an overfitted model. And in another case, for example, it allows too many behaviors.

4. The main process is Confirmation of receipt. T02 Check confirmation of receipt T04 Determine confirmation of receipt TOS Print and send confirmation of receipt T06 Determine necessity of stop advice and T10 Determine necessity to stop indication and the most notable parts is the existence of unlabeled transitions between the following activities:

4.1) Confirmation of receipt' and 'T02 Check confirmation of receipt.

4.2) T02 Check confirmation of receipt and T04 Determine confirmation of receipt'.

4.3) TO2 Check confirmation of receipt and T02 Check confirmation of receipt.

4.4) 'T04 Determine confirmation of receipt' and 'T05 Print and send confirmation of receipt.

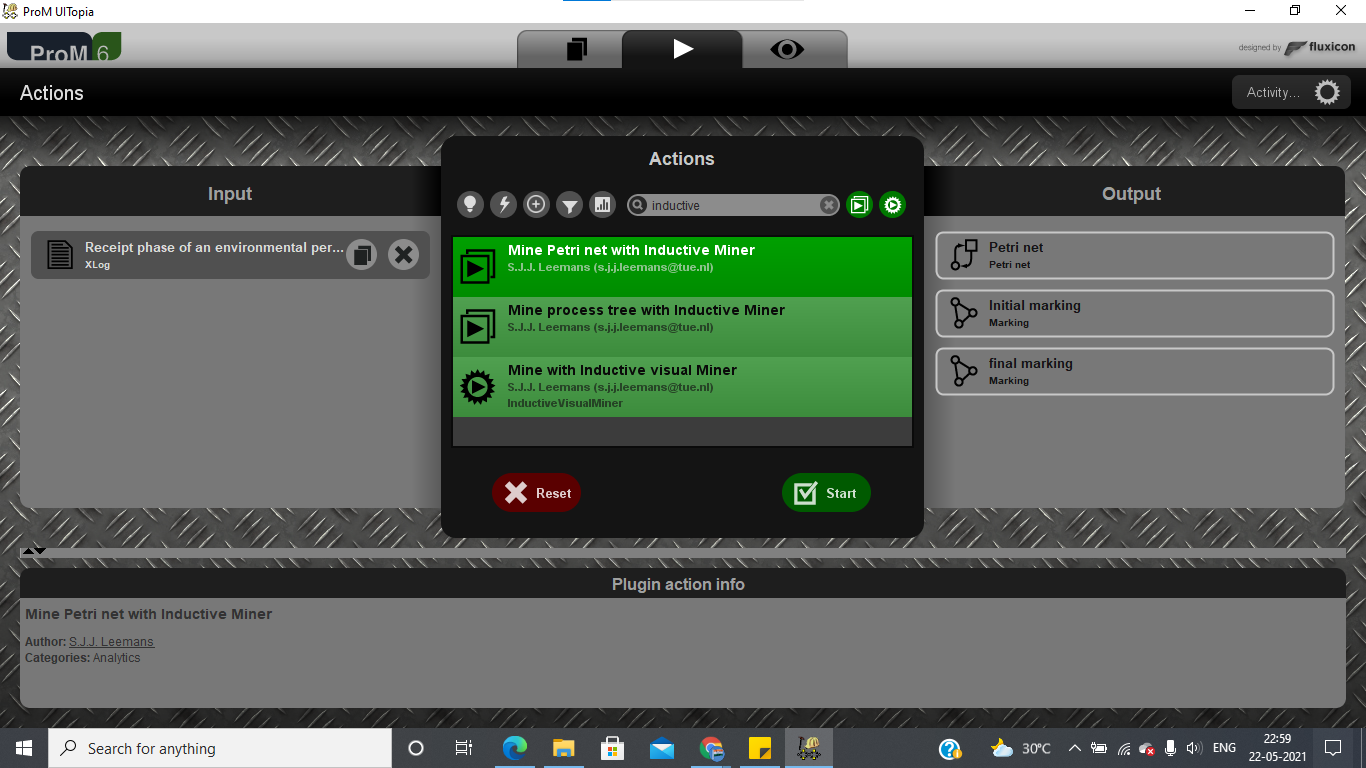
4.5) TO5 Print and send confirmation of receipt and T06 Determine necessity of stop advice.

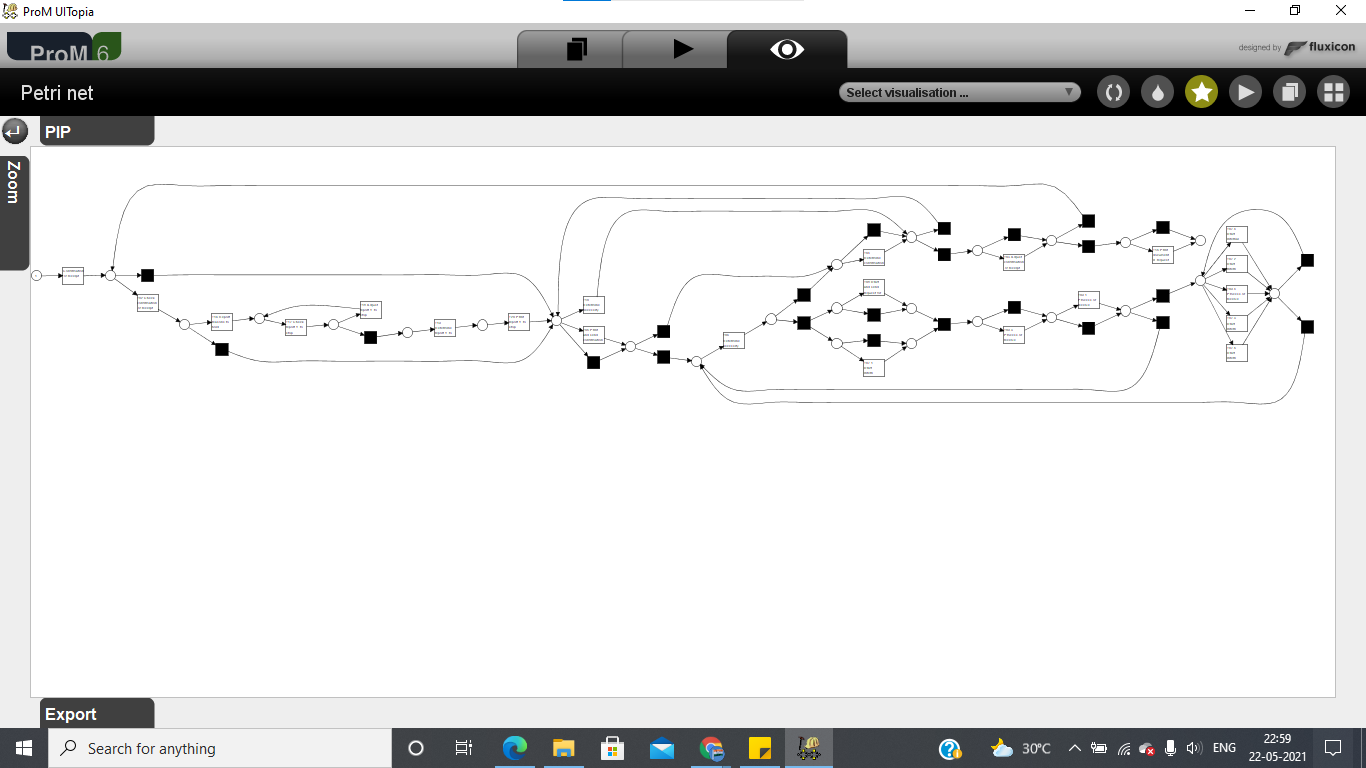
4.6) 'Confirmation of receipt and T06 Determine necessity of stop advice.

4.7) T06 Determine necessity of stop advice and T06 Determine necessity of stop advice,

4.8) T06 Determine necessity of stop advice and T10 Determine necessity to stop indication

The untagged transition would represent the activities that are not found in the petri net but if found in the event log these missing activities are 20 in total.

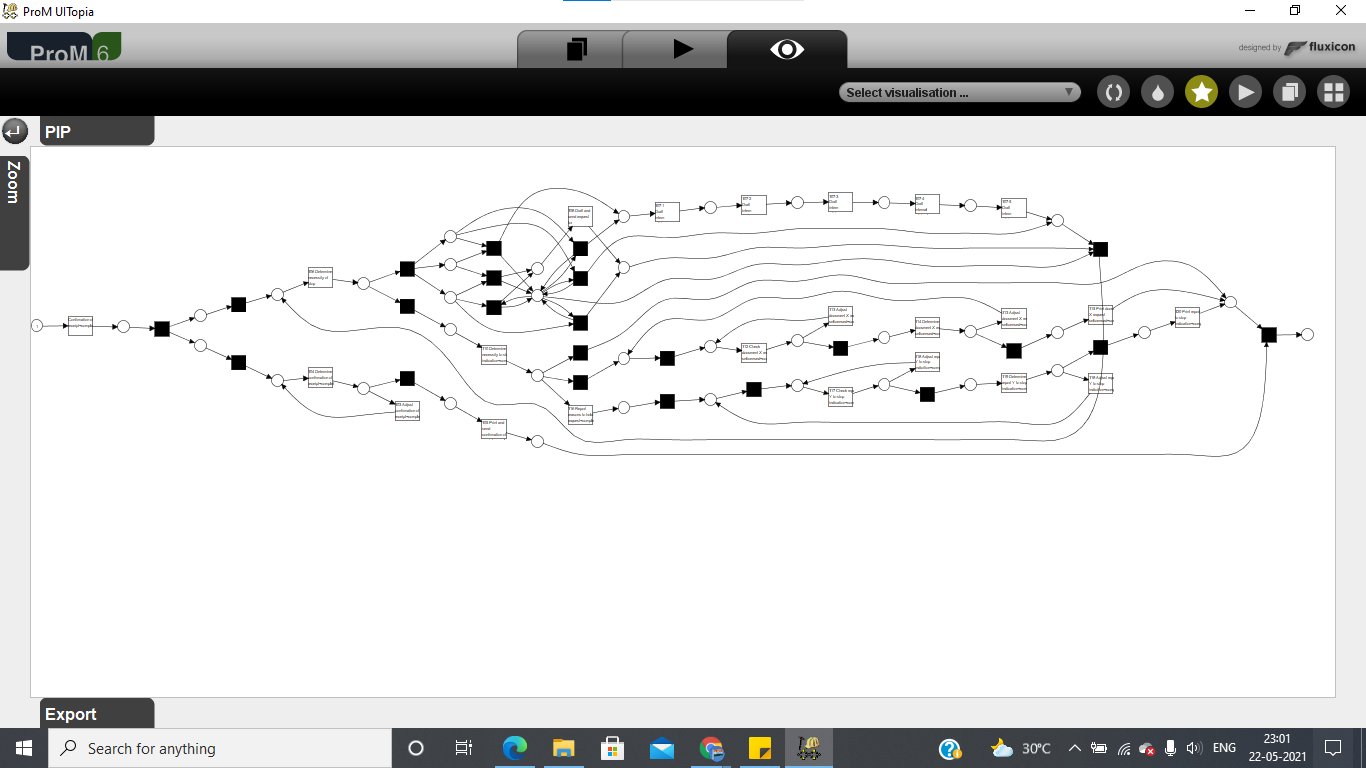




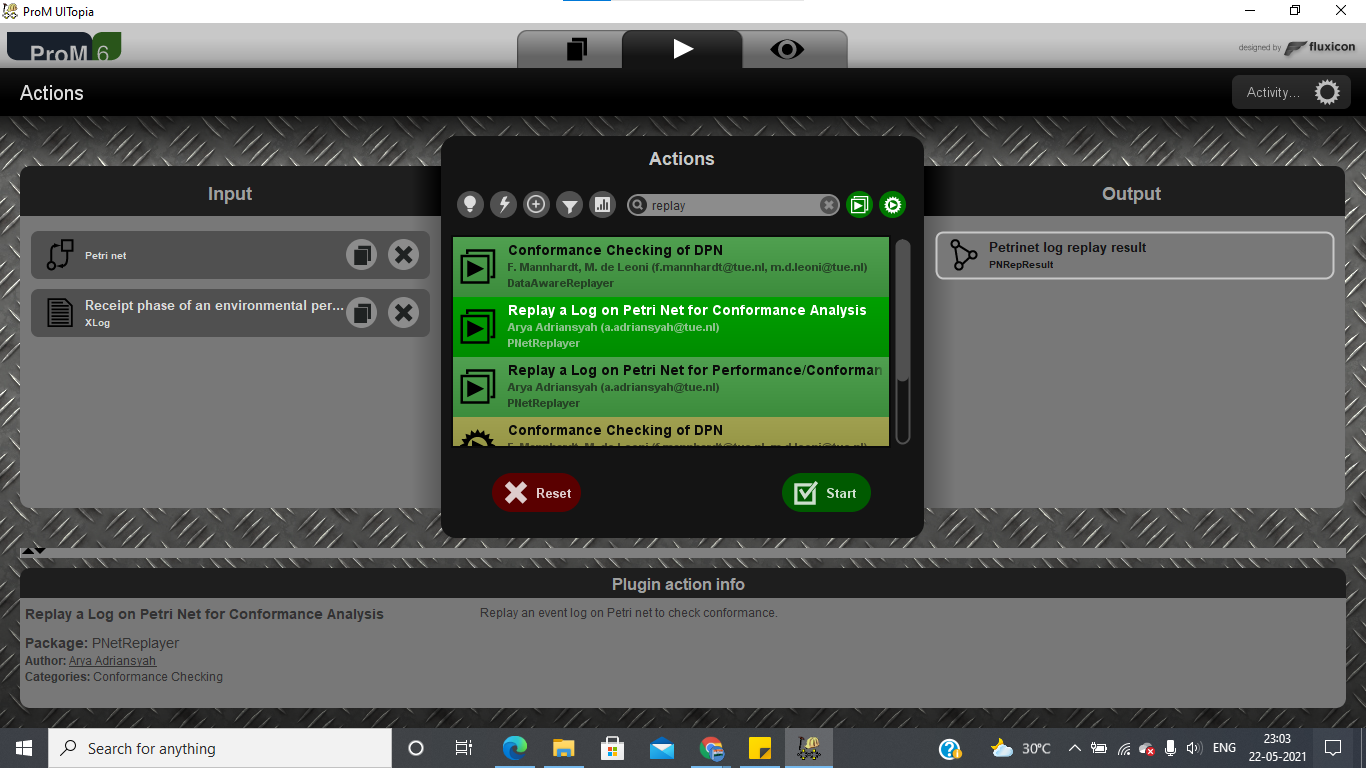
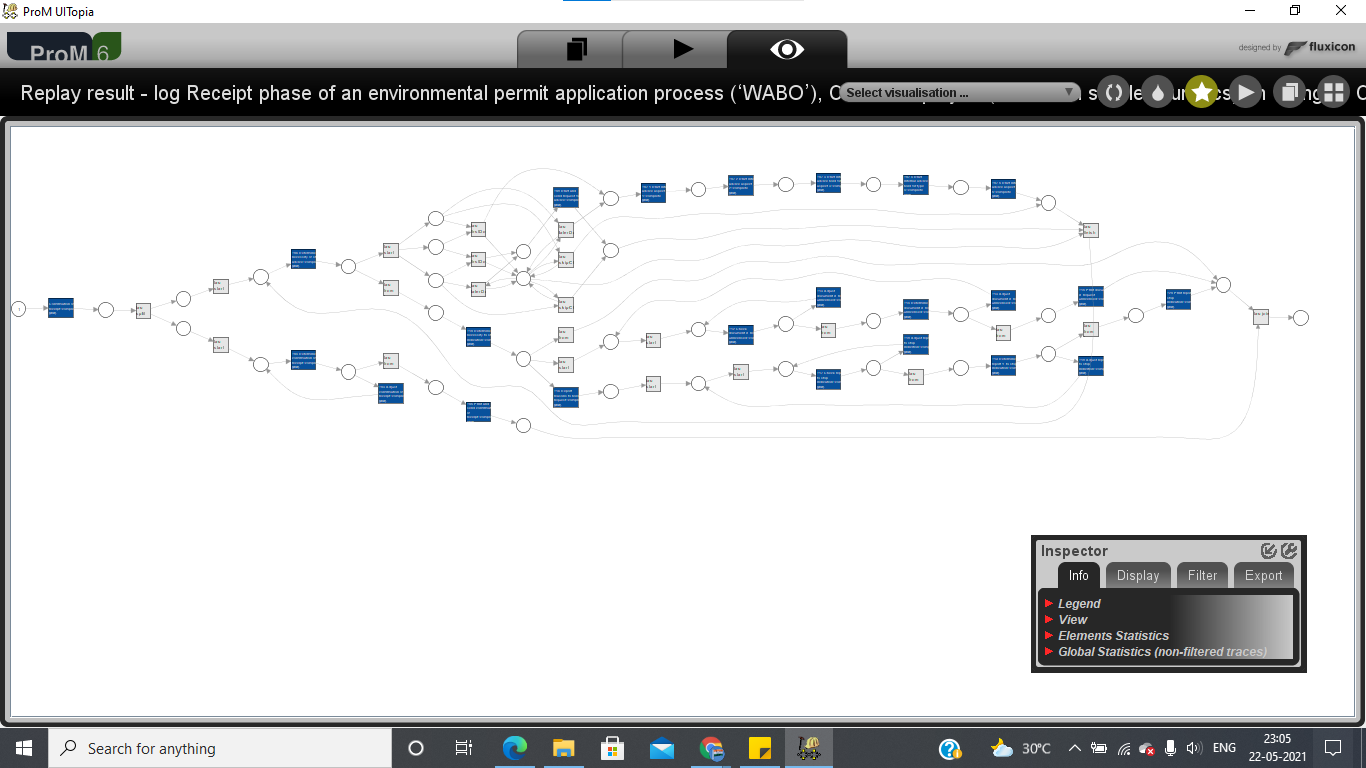
**The organization has a process model that describes the 'should be' process (i.e. a normative process model). Load the file 'normativeModel.pnml' into ProM and apply conformance checking on this process model, and on the full unfiltered original event log.**

1. **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.**
2. **What is the replay fitness (the 'trace fitness' statistic) of the event log on the normative process model?**
3. **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?**
4. **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?**

1.



2.The 'trace fitness statistic is: 0.842543 is the best possible fitness.



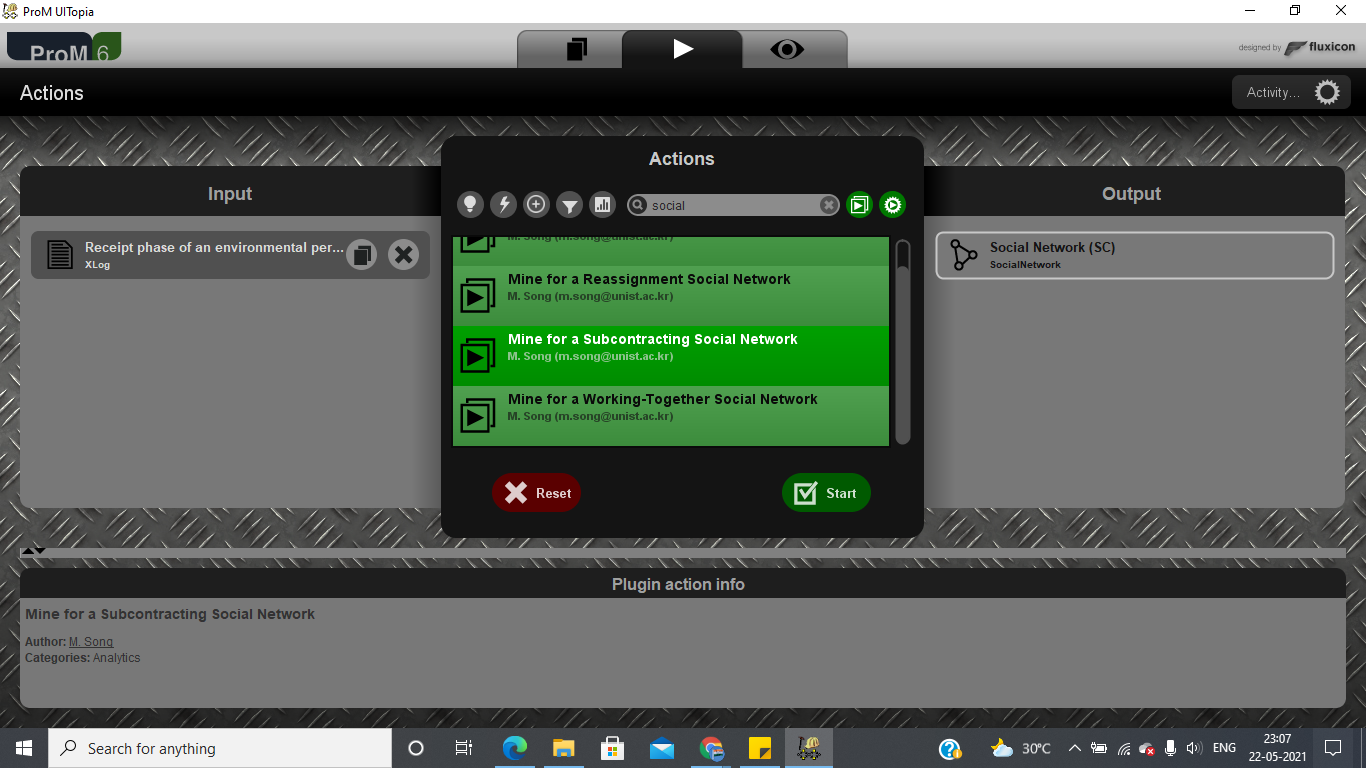
3.The transition runs 1327 times successfully and 125 times incorrectly; in detail we have log (89), model (125) and sync [log + model] (1327).

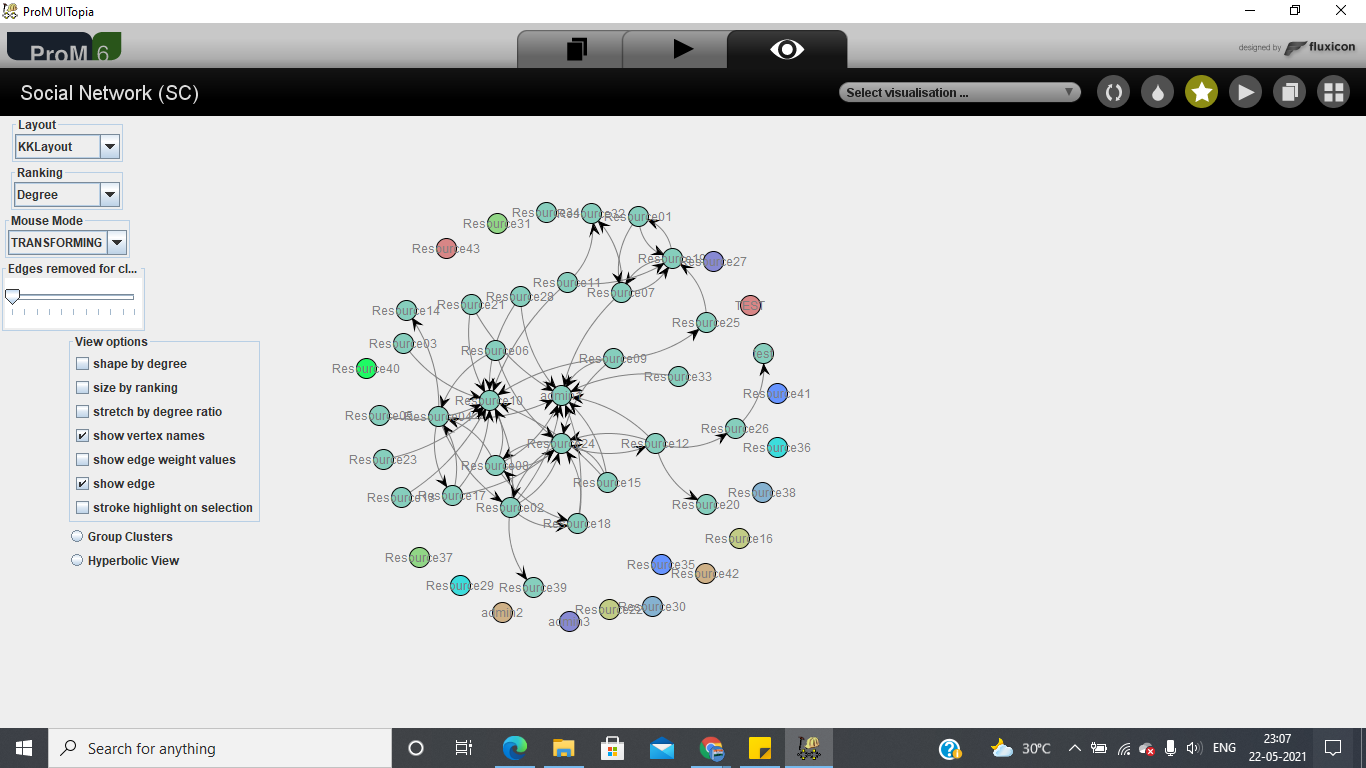
4.There are 125 traces that do not fit the model and the activity transition T06 Determine necessity of stop advice' is only directly related to unlabeled transitions through places.

**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? Explicitly discuss the settings you have used in the resulting visualization.**
2. **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:**
3. **Are all users executing activities from the start of the event log, or are some users joining later?**
4. **Are users mainly executing particular activities or are most users executing most of the activities?**

1.The Mine for a Subcontracting Social Network plug-in was loaded, and when performing the following configuration: 1.1) Layout: FRLayout 1.2) Ranking Degree 1.3) Edges removed...: a bar 1.4) View options: [x] size by ranking [x] show edge and [x] group clusters it is observed that two representative user groups are distinguished.





2.The viewer used is dotted chart view-



3.Some users join later. some of these users join from one month later to eleven months later in other cases.

4. Some users more or less sixteen of them approximately, execute particular activities that go from one to five activities, the other group of users that is two thirds of them execute most of the activities.

**To conclude this assignment, briefly discuss three observations you have made during your analysis that you would like to communicate to the business user.**

**Think for instance of possible improvement opportunities and starting points for further investigation.**

1.The most frequent activity in which the process concludes is '10 Determine necessity to stop indication and in the following activities that are executed it can be improved to reduce the time of completion of the process with a shorter duration.

2. As a starting point for an investigation, the identification of less frequent traces vs. more frequent traces is recommended for a model comparison.

3. Another starting point for an investigation and to identify the causes based on users who carried out particular activities due to the non-continuation of future activities.