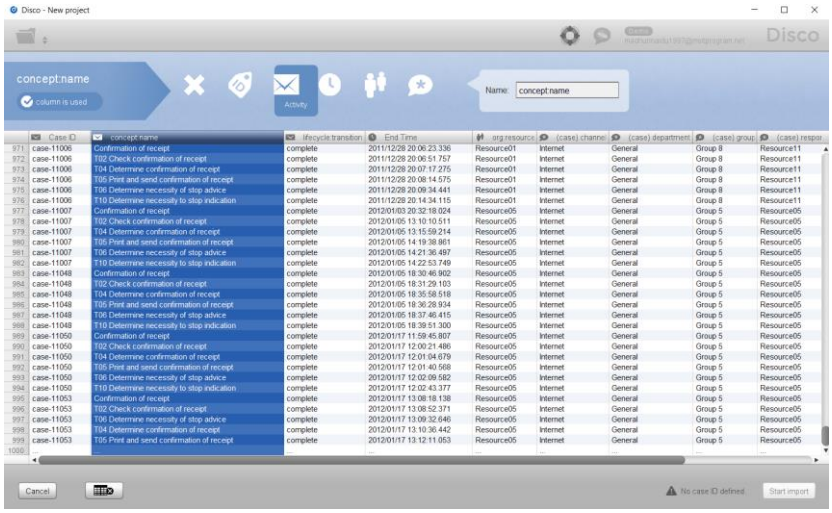
Q1: Open the event log ('Receipt phase of an environmental permit application process (_WABO_) CoSeLoG project.fbt') in Disco and switch to the 'Statistics' view.

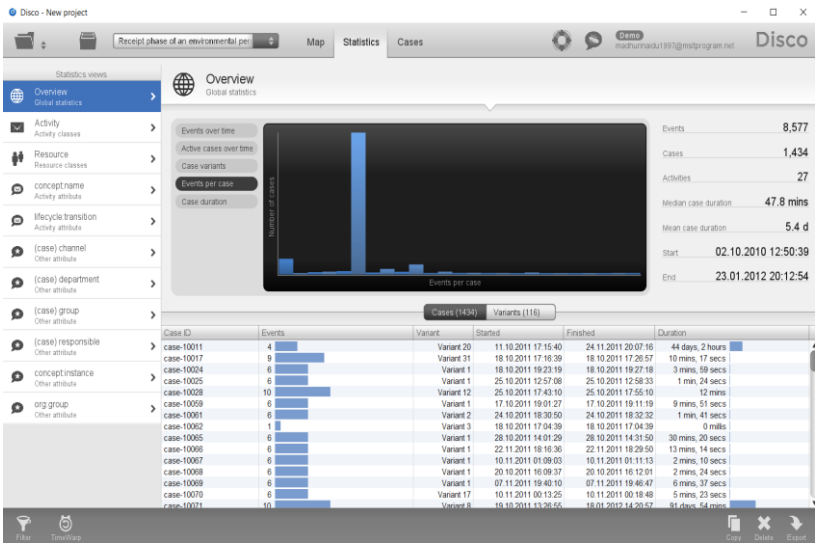
Without switching to other views, use the statistics view to answer the following three sub questions:

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?**

1. First i have loaded the file "Receipt phase of an environmental permit application process (_WABO_) CoSeLoG project" in the disco tool and clicked on statistics https://github.com/Madhuri97/Data_Science_2019501105/blob/master/Process%20Mining/1A.PNG

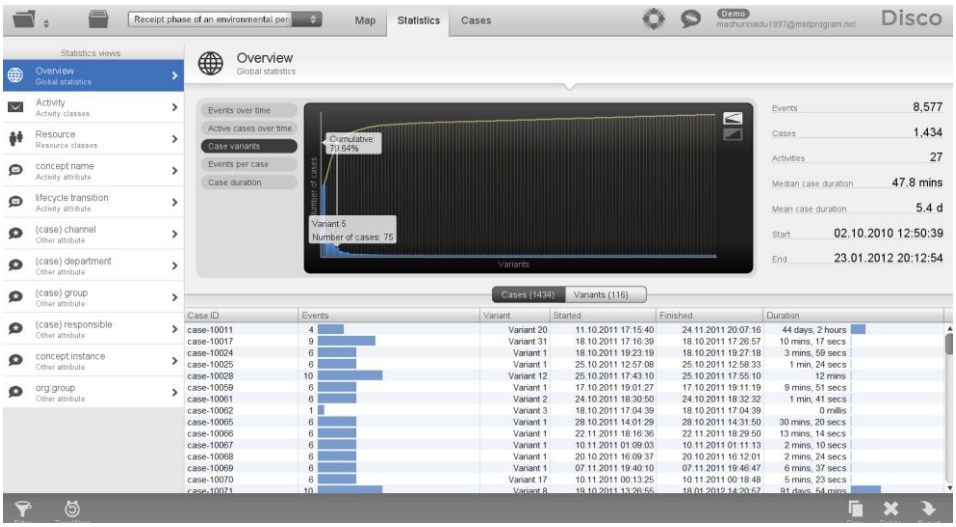


I saw a graph showing the number of cases on the Y axis and the volume events per case on the X-axis. On right side there is a resume table for process data https://github.com/Madhuri97/Data_Science_2019501105/blob/master/Process%20Mining/1B_eventpercase.PNG

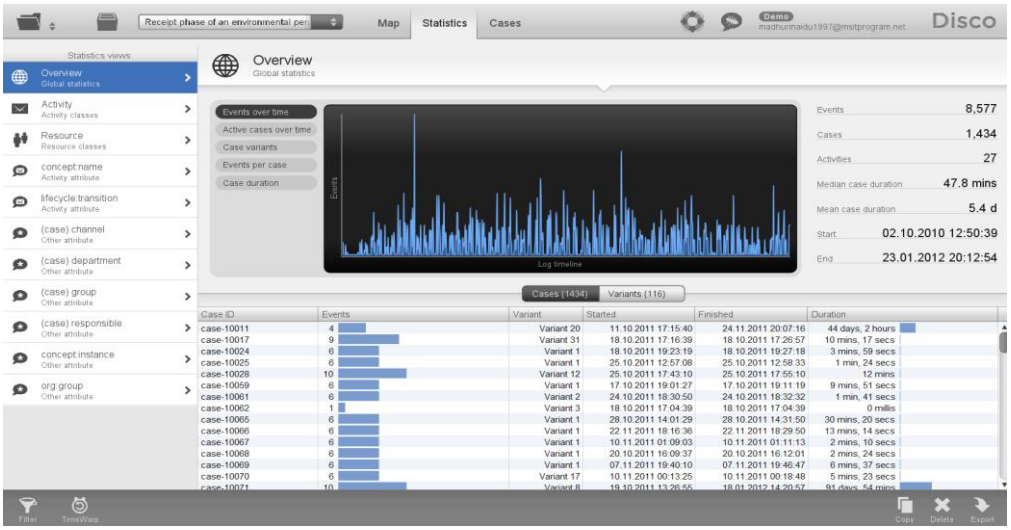


It was enough to identify 8,577 events 1,434 cases and mean case duration of 5.4 by dividing **on average 5,98**

2. I have pointed at the frequency of 5 taking the pareto's principle as a reference. In conclusion that most cases are concentrated in few variants which is 116 different ones. https://github.com/Madhuri97/Data_Science_2019501105/blob/master/Process%20Mining/case_variants.png



3. Event case time. I saw a graph composed of the events on Y-axis and the timeline on the X-axis. noticed a great variability in the duration of cases https://github.com/Madhuri97/Data_Science_2019501105/blob/master/Process%20Mining/Eventsovertime.PNG



Q2: While still in Disco, switch to the 'map' view to display a process map.

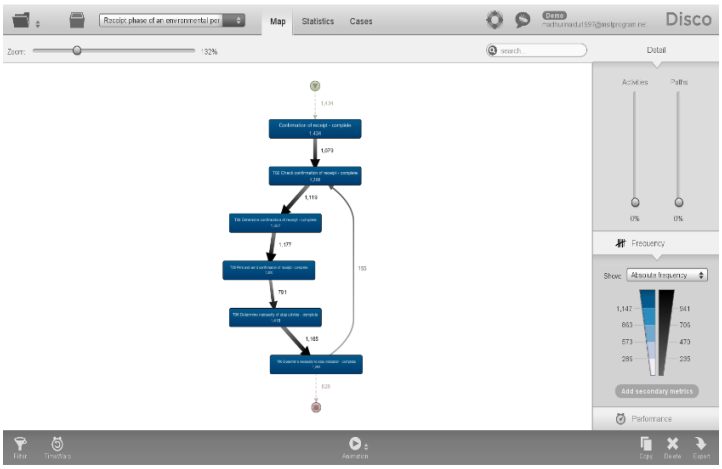
Using the map view, change the activity and path detail settings in order to create a comprehensible process map (e.g. a process map that could be printed on one A4 or letter paper or shown on a single computer screen while still being readable in full).

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.

In disco switch to map view to display process map.

1. At the top centre of the screen. I clicked at map menu, Then I have adjusted the detail activity bar to Activities equal 0% and the path equal 0% https://github.com/Madhuri97/Data_Science_2019501105/blob/master/Process%20Mining/AbsoluteFrequency0.PNG



   The process contains 6 steps which are shown in the given link above

   The main process is about **checking receipts**

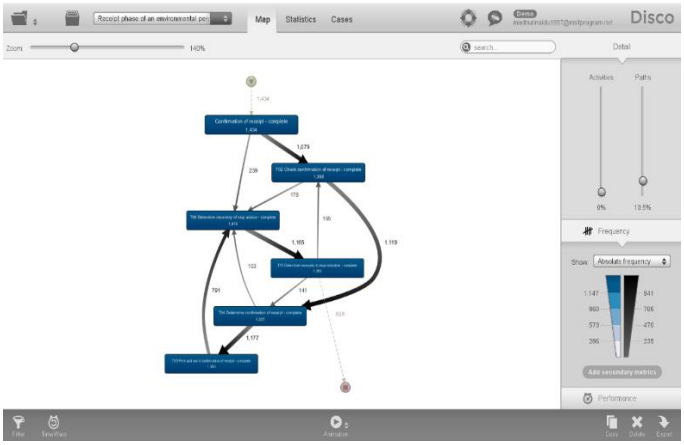
2. The most frequent sequence of activities is:

- Confirmation of receipt (1,1434),
- T02 Check Confirmation of receipt (1, 368),
- T04 Determine Confirmation of receipt (1, 307),
- T05 Print and send Confirmation of receipt (1, 300),
- T06 Determinate Necessity of stop advice (1, 416),
- T10 Determine Necessity to stop indication (1, 283)

The most frequent paths are: T04 > T05(1, 177), T06 > T10(1, 416) and T02 > T04(1, 119)

Q3: 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.

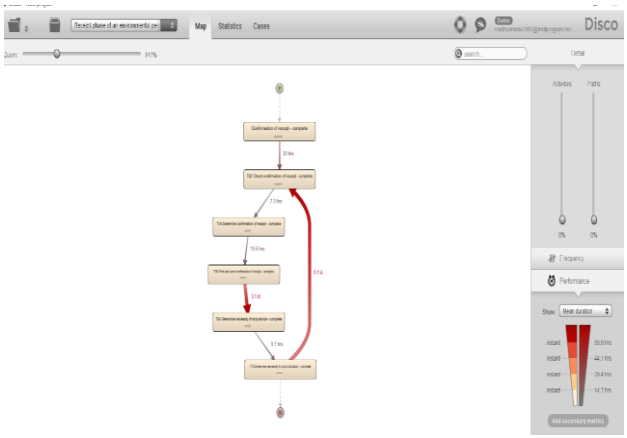
Using the same process map I set the paths bar to 10.5% in order to have a better view of the connections. After clicking the Performance button it will change the display.
https://github.com/Madhuri97/Data_Science_2019501105/blob/master/Process%20Mining/Absolute_fequency10p.PNG



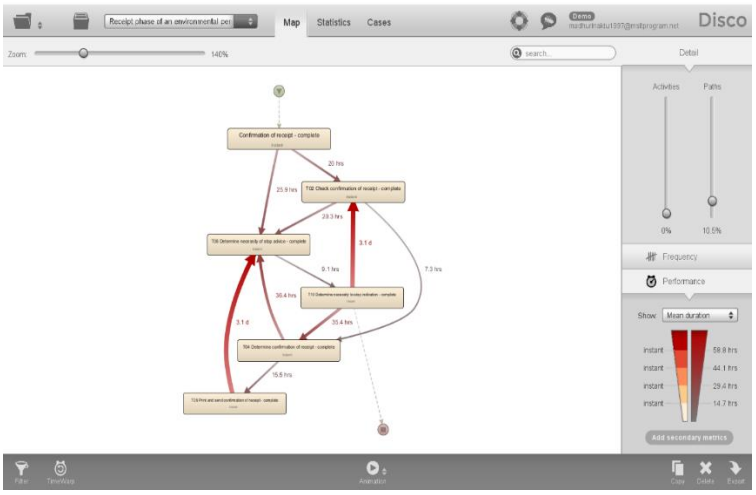
The longest duration was automatically highlighted with different thicknesses and colors. The average was chosen because it trends to normalize deviations and outliers of the process due to the large volume of cases.

The main bottlenecks are located at T05 and T10 where the average time of 3.1 days is twice as long as others

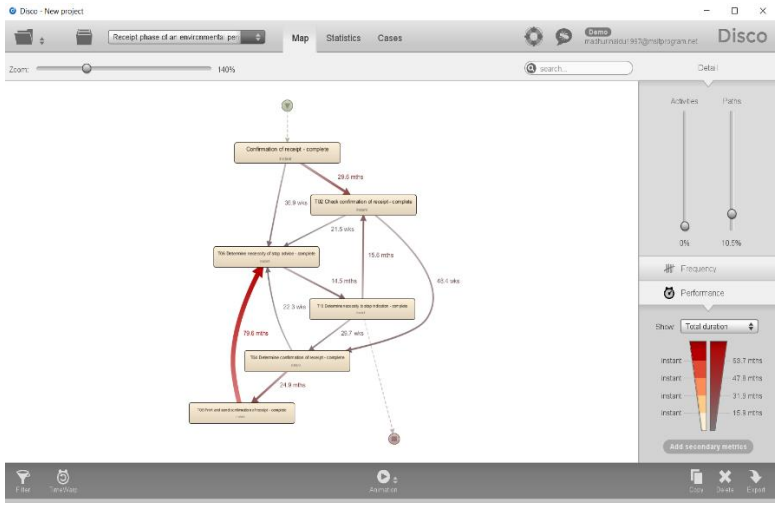
0%:
https://github.com/Madhuri97/Data_Science_2019501105/blob/master/Process%20Mining/Mean%20Duration.PNG



10%:
https://github.com/Madhuri97/Data_Science_2019501105/blob/master/Process%20Mining/MeanDuration10P.PNG



The Total Duration shows the process as a whole unfeasible, being this the main bottleneck. https://github.com/Madhuri97/Data_Science_2019501105/blob/master/Process%20Mining/TotalDuration.PNG



Q4: 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:

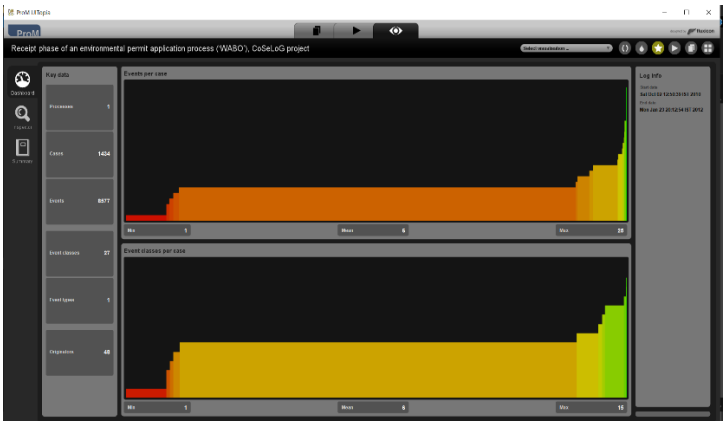
1. **Is the arrival rate of new cases constant? If not, when are there fluctuations? If yes, how can we see this from the Dotted Chart?**
2. **Can you observe a change in the global process?**

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

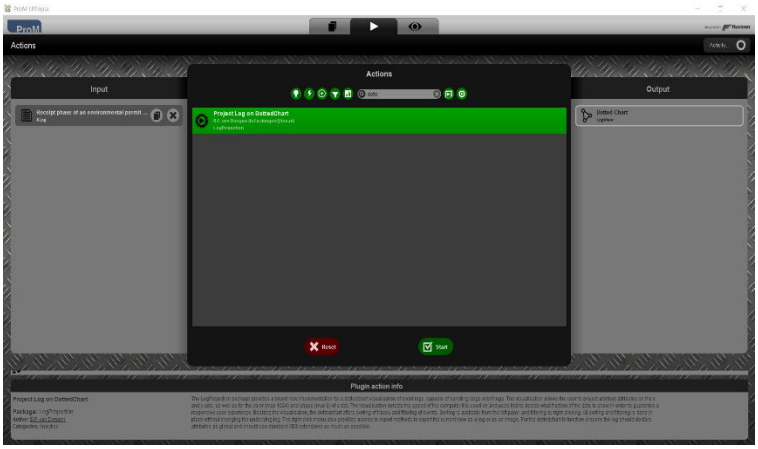
The Dotted Chart is explained in lecture 4.8: 'Exploring Event Data'.

Uncompressed the “Receipt phase of an environmental permit application process” file. Open ProM and click the “Import” button on the top-right corner. Click “Eye” button to get the visualization.

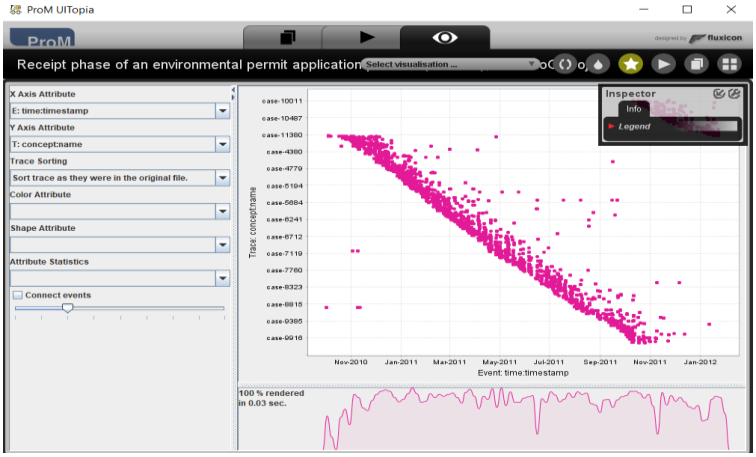
https://github.com/Madhuri97/Data_Science_2019501105/blob/master/Process%20Mining/ProM/4_Viewing1.PNG



Then click “play button” to add project log on dotted chart. https://github.com/Madhuri97/Data_Science_2019501105/blob/master/Process%20Mining/ProM/4_Dottedgraphforxels.PNG

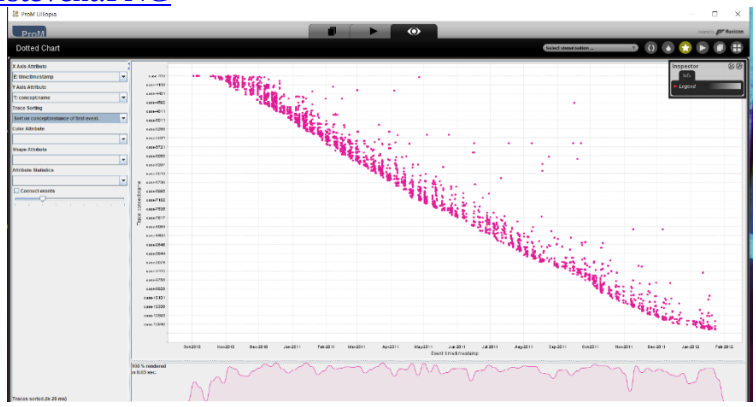


Then we get dotted graph https://github.com/Madhuri97/Data_Science_2019501105/blob/master/Process%20Mining/ProM/4_Dotted%20Chart.PNG



to get the rhythm at which new cases arrive we go to “trace sorting” after clicking sort of first event we get:

https://github.com/Madhuri97/Data_Science_2019501105/blob/master/Process%20Mining/ProM/4_Dottedoffirstevent.PNG



1. **Yes**, the rate of arriving new cases is indeed constant since we have sorted the traces according to the instance of the first event, and we cannot find abrupt jumps between them. We can see the lower flow of cases at the beginning. A little variation in receipt cases. There is small change in behaviour of the process has been observed as well as small interruptions caused by weekends in which there is no data insertion.

2. Regarding the global process **Yes,** the concentration of the points at the beginning of the series and their dilution over time denote a change in the behaviour of the process.

Q5: 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.

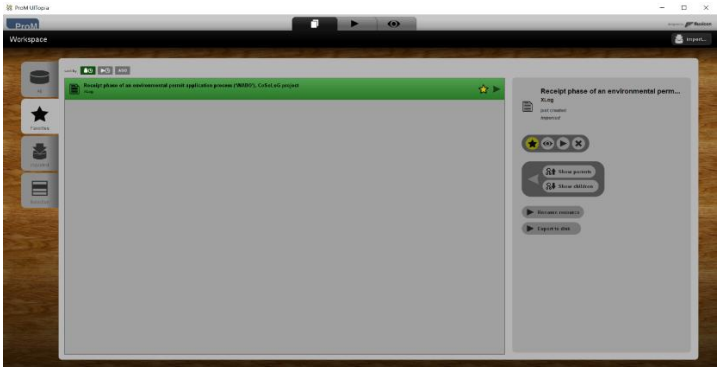
1. **Clearly indicate which settings you have used for the 'Filter log using simple heuristics' plug-in.**
2. **Explicitly motivate the filtering settings chosen, why did you pick this percentage or selection of activities?**
3. **Discuss and argue which plug-in (or chain of plug-ins) you have used to 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'.

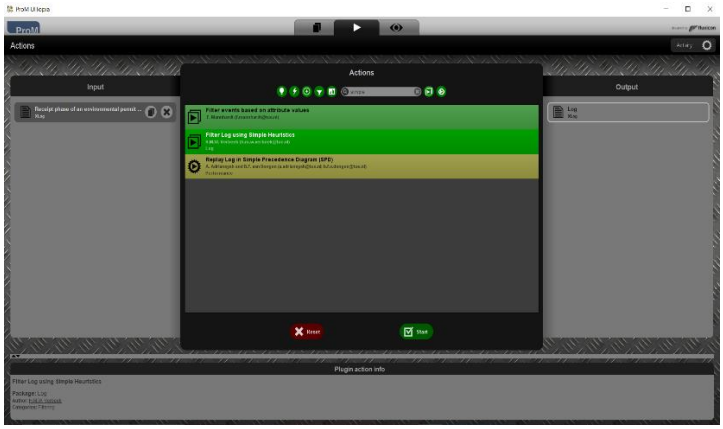
Suggested list of plug-ins or plug-in chains to produce a Petri net:

- **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. After selecting the receipt file then play there we can search as “Filter log using simple heuristics” plug-in. I used settings: Percentage: 80%, Start activity: Confirmation of receipt complete, End activities: T05 and T10, Activities: Confirmation of receipt complete, T02, T04, T05, T06, T10
https://github.com/Madhuri97/Data_Science_2019501105/blob/master/Process%20Mining/ProM/5_ImportRecieptFile.PNG

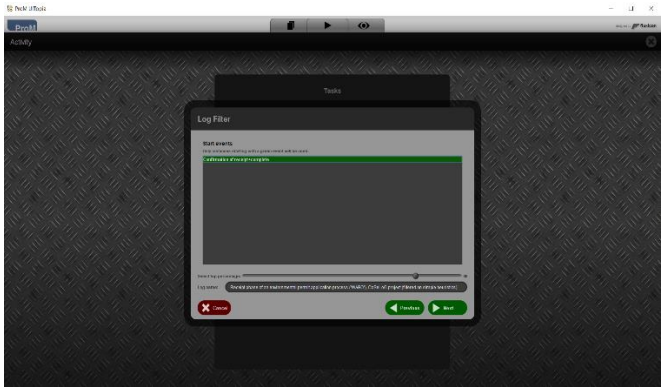


https://github.com/Madhuri97/Data_Science_2019501105/blob/master/Process%20Mining/ProM/5_filterLogSimpleHeuristic.PNG



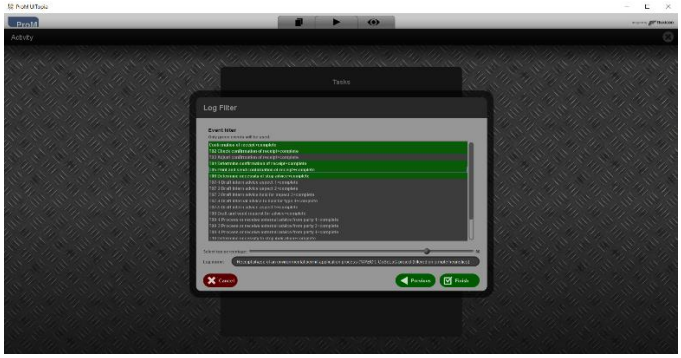
https://github.com/Madhuri97/Data_Science_2019501105/blob/master/Process%20Mining/ProM/5_LogFilterComplete.PNG

https://github.com/Madhuri97/Data_Science_2019501105/blob/master/Process%20Mining/ProM/5_LogFilterCompleteCofirm.PNG

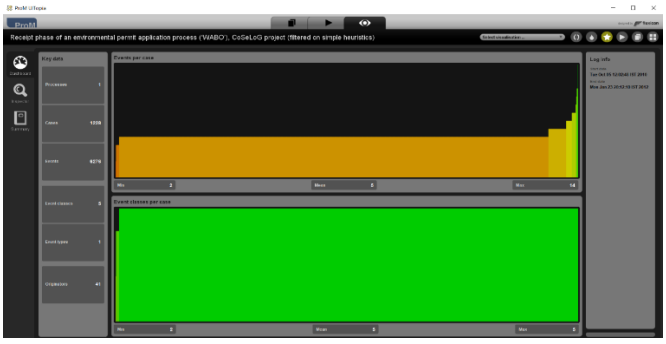


https://github.com/Madhuri97/Data_Science_2019501105/blob/master/Process%20Mining/ProM/5_LogFilterTasks.PNG

https://github.com/Madhuri97/Data_Science_2019501105/blob/master/Process%20Mining/ProM/5_LogFilterTasksEvent.PNG



https://github.com/Madhuri97/Data_Science_2019501105/blob/master/Process%20Mining/ProM/5_VisualizationAfterFilter.PNG

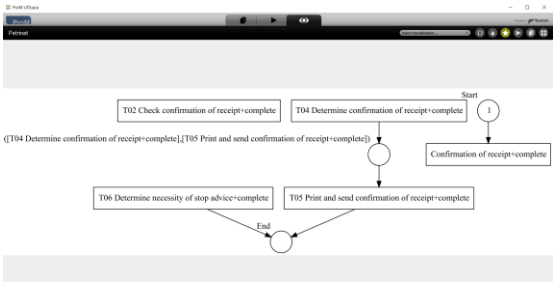


2. Motivation according to the percentages mentioned above one can consider the resulting sample to be representative of the most common process described by the filtered event log
https://github.com/Madhuri97/Data_Science_2019501105/blob/master/Process%20Mining/ProM/5_FilteredLogSummary.PNG

3. In order to discover the Petri net. I have tried the following plug-ins with the described results in each case
    - Alpha Miner is not able to discover all the places clearly resulting in a misfitting model
    https://github.com/Madhuri97/Data_Science_2019501105/blob/master/Process%20Mining/ProM/5_ActionAlphaMiner.PNG

    https://github.com/Madhuri97/Data_Science_2019501105/blob/master/Process%20Mining/ProM/5_AlphaActivityClassifier.PNG

https://github.com/Madhuri97/Data_Science_2019501105/blob/master/Process%20Mining/ProM/5_AlphaVisualizer.PNG



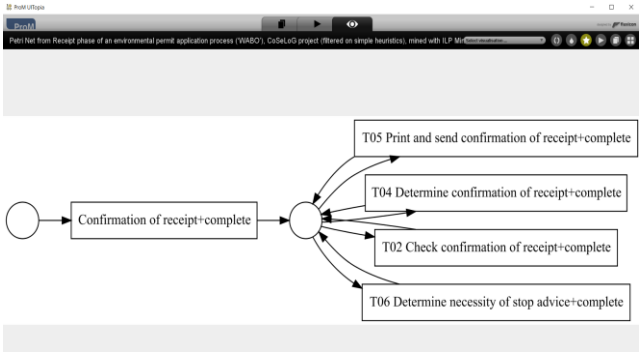
- ILP Based Process Discovery is also not able to discover all the places. clearly resulting in a misfitting model https://github.com/Madhuri97/Data_Science_2019501105/blob/master/Process%20Mining/ProM/5_ILPBasedProcessDiscovery.PNG

https://github.com/Madhuri97/Data_Science_2019501105/blob/master/Process%20Mining/ProM/5_ILPBasedActivityClassifier.PNG

https://github.com/Madhuri97/Data_Science_2019501105/blob/master/Process%20Mining/ProM/5_ILPTasks.PNG

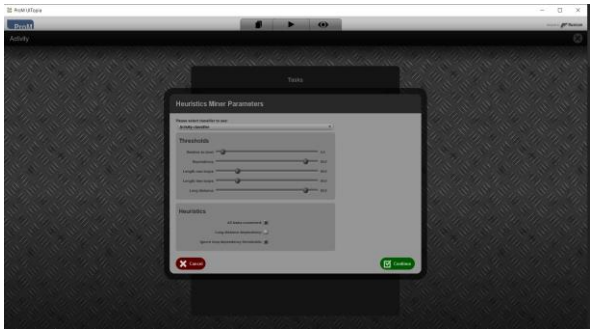
https://github.com/Madhuri97/Data_Science_2019501105/blob/master/Process%20Mining/ProM/5_ILPConfig.PNG

https://github.com/Madhuri97/Data_Science_2019501105/blob/master/Process%20Mining/ProM/5_ILPPetriNet.PNG

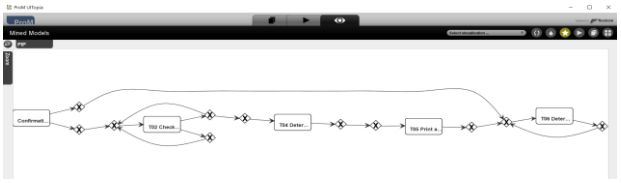


- Mine for a Heuristics Net using Heuristics Miner followed by Convert Heuristics net into Petri net able to discover a working Petri net but it lacks of simplicity at first glance and also it seems to have fitness problems as one of the events identified as End events in the Activity classifier was not discovered as an end event
https://github.com/Madhuri97/Data_Science_2019501105/blob/master/Process%20Mining/ProM/5_MineHeuristic.PNG

https://github.com/Madhuri97/Data_Science_2019501105/blob/master/Process%20Mining/ProM/5_heuristicsParams.PNG



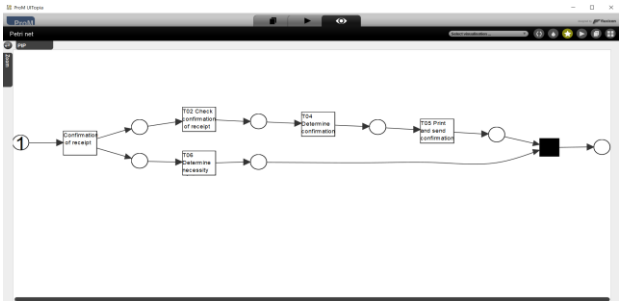
https://github.com/Madhuri97/Data_Science_2019501105/blob/master/Process%20Mining/ProM/5_HeuristicVisualization.PNG



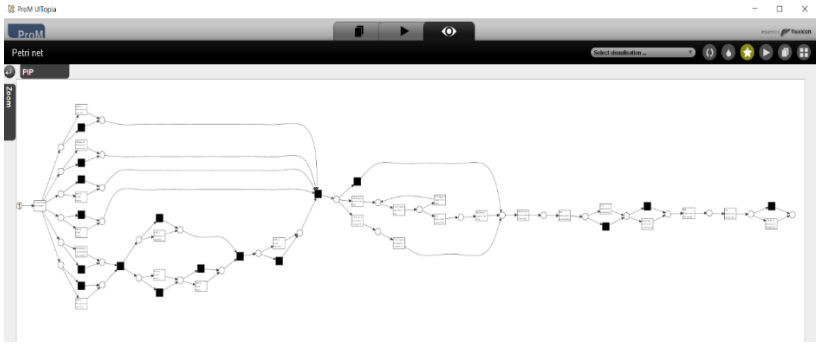
- Inductive Miner Looks like the best option as it is able to discover a working Petri net that looks very simple at a glance and does not seems to have fitness problems
  https://github.com/Madhuri97/Data_Science_2019501105/blob/master/Process%20Mining/ProM/5_MineInductiveMiner.PNG

https://github.com/Madhuri97/Data_Science_2019501105/blob/master/Process%20Mining/ProM/5_MineInductiveMinerTasks.PNG

https://github.com/Madhuri97/Data_Science_2019501105/blob/master/Process%20Mining/ProM/5_InductivePetriNet.PNG



4. My choice is the Petr net discovered by the inductive miner. The process starts with the activity Confirmation of receipt-complete then it splits in two paths: https://github.com/Madhuri97/Data_Science_2019501105/blob/master/Process%20Mining/ProM/5_MinePetrinetUsingInductive.PNG



The first of the paths follows with the activity T02 Check confirmation of receipt complete then there is the option to execute activity Tod Determine confirmation of receipt or complete which could take you to a loop back to the activity T02 Check confirmation of receipt complete or skip it to the activity T05 Print and send confirmation or receipt complete which is the end activity in this path. There is the option to skip this path all together to the end as well

The second path follows with the activity T06 Determine necessity of stop advice complete then there is the option to loop back to itself or continue with activity T10 Determine necessity to stop indication complete which also can be shipped to the end. There is also the option to skip this path all together to the end as well.

Q6: 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?**

Instructions to align the process model with the event log:

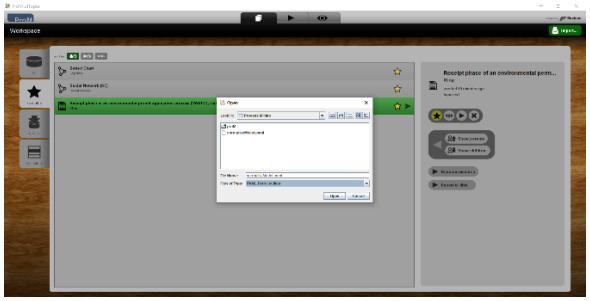
1. **Import the normative model using the 'PNML Petri net files' importer.**
2. **Select the imported normative Petri net and the event log, start the plug-in called 'Replay a Log on Petri Net for Conformance Analysis' (not the variant with performance!), and click 'yes' in the 'No Final Marking' pop-up.**
3. **Select the 'sink' place on the left (note: do not select '0-sink' etc.) and click the button 'Add Place >>' to add the place 'sink' to the candidate final marking list. Now click 'Finish'.**
4. **Click 'Finish' in the mapping wizard.**
5. **Click 'No, I've mapped all necessary event classes' to indicate that some events are not present in the normative model.**
6. **Now click 'Next' and 'Finish'. The normative process model is shown with conformance information projected onto it.**

If you followed these instructions exactly you do not need to mention these steps in your answer.

More information regarding this conformance technique is provided in lecture 4.7: 'Aligning observed and modeled behaviour' (and to a lesser extend in the lectures 4.3 through 4.6).

Instructions:

I. https://github.com/Madhuri97/Data_Science_2019501105/blob/master/Process%20Mining/ProM/6_Importingpnml.PNG



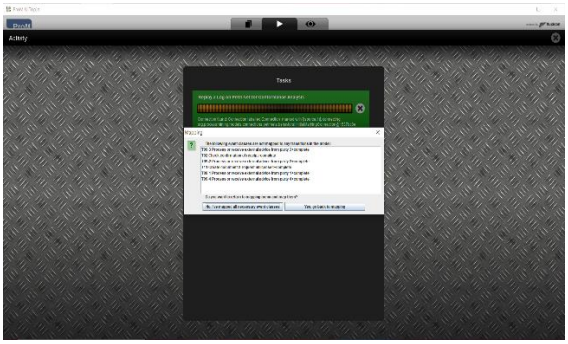
II. https://github.com/Madhuri97/Data_Science_2019501105/blob/master/Process%20Mining/ProM/6_ReplayConfromanceanalysis.PNG

III. https://github.com/Madhuri97/Data_Science_2019501105/blob/master/Process%20Mining/ProM/6_selectedsink.PNG

,

https://github.com/Madhuri97/Data_Science_2019501105/blob/master/Process%20Mining/ProM/6_Mappinglog.PNG

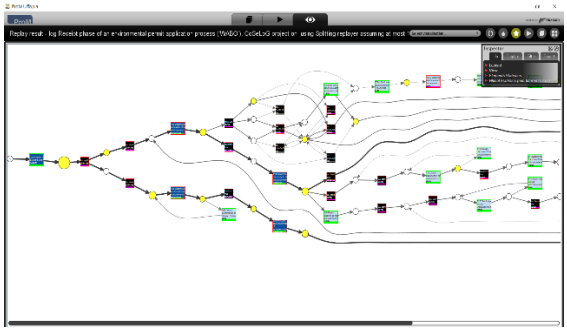
IV. https://github.com/Madhuri97/Data_Science_2019501105/blob/master/Process%20Mining/ProM/6_Nomappedall.PNG



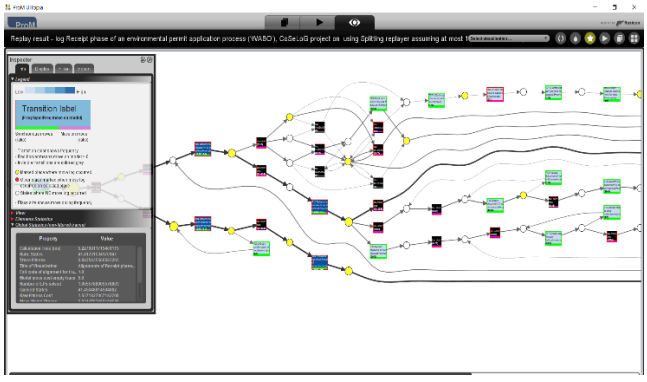
V. https://github.com/Madhuri97/Data_Science_2019501105/blob/master/Process%20Mining/ProM/6_replaywizard.PNG

VI. https://github.com/Madhuri97/Data_Science_2019501105/blob/master/Process%20Mining/ProM/6_replaysetparams.PNG

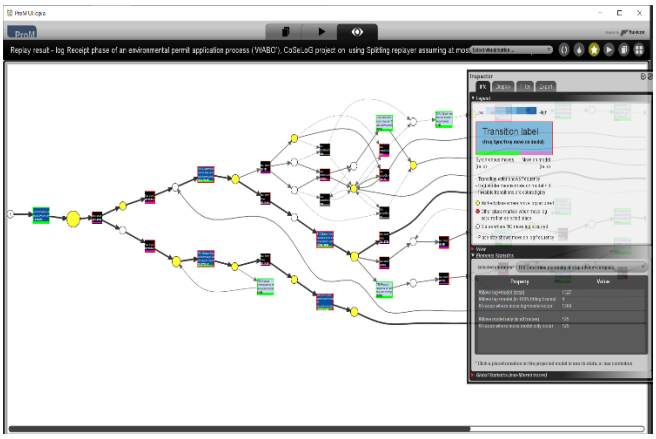
1. followed the instructions to align the process model with the event log in order to show the normative model with the respective conformance information. I saw that highest deviations occur with respect to activities **T04' or 'T06** but also to a lower degree for '**T05' and T10'**. This is shown in the visualization https://github.com/Madhuri97/Data_Science_2019501105/blob/master/Process%20Mining/ProM/6_Replayresult.PNG



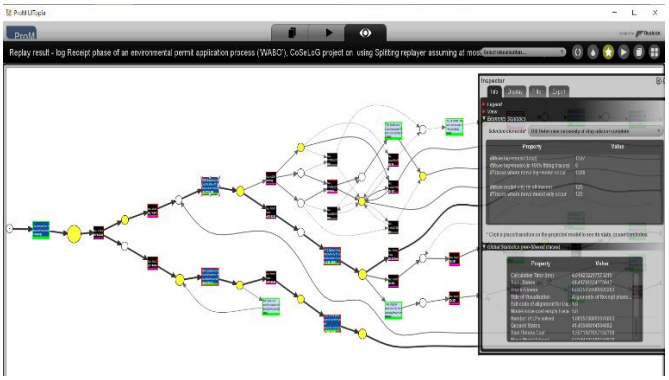
2. At the inspector menu, I've clicked at global statistics and find the Trace Fitness value. What I saw a couple of statistics values. **Trace fitness: 0.8425**, This means that the model and the observed features present problems that must be analysed in order to adjust the process. https://github.com/Madhuri97/Data_Science_2019501105/blob/master/Process%20Mining/ProM/6_Infoofreplay.PNG



3. The element statistics of T06 Determine necessity of stop advice complete can be found by clicking on activity T06' and 'Element Statistic' in the Inspector which executed **1327 times, 1309 of those are correct**, leaving 18 incorrectly executed. Move model only happens **125 times**. That adds up to 18 + 125 = 143 errors in this transition https://github.com/Madhuri97/Data_Science_2019501105/blob/master/Process%20Mining/ProM/6_t06TransitionandElementStatistics.PNG



4. Even though there are some error as described above, most of the executions are **correct (around 80 - 90%)**, so depending on situation and demands on accuracy, it is still a decent result.
https://github.com/Madhuri97/Data_Science_2019501105/blob/master/Process%20Mining/ProM/6_T06info.PNG



Q7: 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.

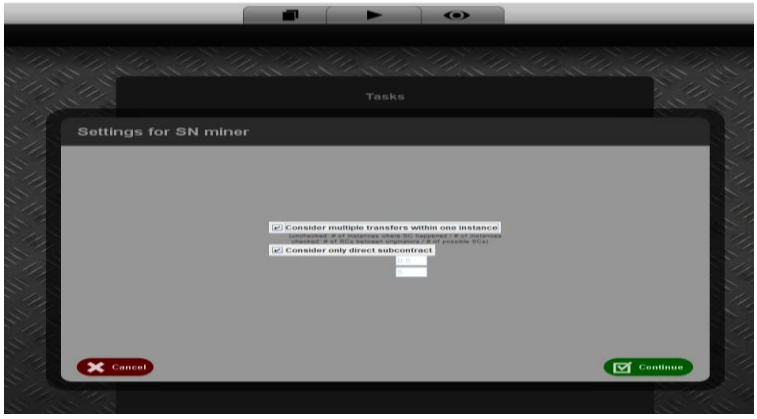
1. **Use the plug-in 'Mine for a Subcontracting Social Network'. Note that subcontracting means that if individual *j* frequently executed an activity in-between two activities executed by individual *i*, then individual *i* subcontracted work to individual *j*. Answer the following question using this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

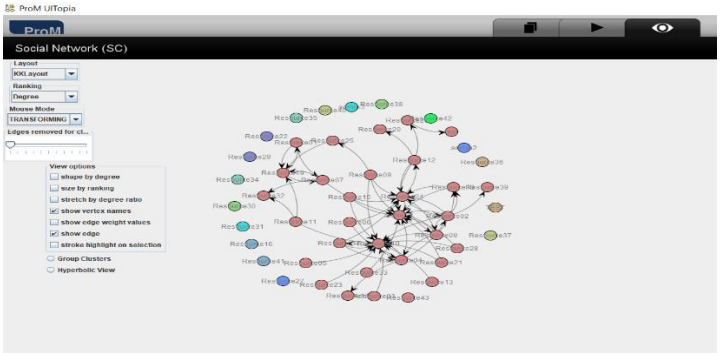
3. **Are all users executing activities from the start of the event log, or are some users joining later?**
4. **Are users mainly executing particular activities or are most users executing most of the activities?**

1. I clicked on the play icon at the top centre of the screen. Then I selected the **"Mine for a Subcontracting Social Network"** plug included the event log in the input section and pressed start. https://github.com/Madhuri97/Data_Science_2019501105/blob/master/Process%20Mining/ProM/7_MineForSubcontractingSocialNetwork.PNG

,https://github.com/Madhuri97/Data_Science_2019501105/blob/master/Process%20Mining/ProM/7_Msocialoption.PNG



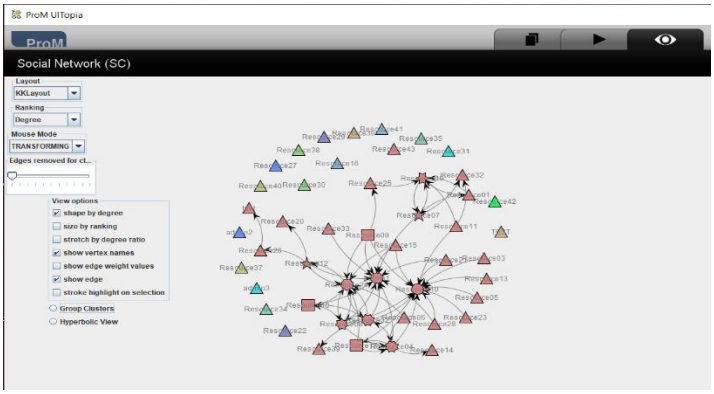
I saw a circular diagram composed of several colored circles representing the relationships between the different resources. https://github.com/Madhuri97/Data_Science_2019501105/blob/master/Process%20Mining/ProM/7_Socialgraphfinal.PNG



Yes, it is easy to distinguish several groups of users marked by different colors. You can click at Cluster groups" in order to have a better view. Many users are in the subcontracting network, while others are not

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 ' Resource classifier and the color attribute to 'C: Activity Classifier'. Answer the following two questions using this view I just followed the Instructions given in the question https://github.com/Madhuri97/Data_Science_2019501105/blob/master/Process%20Mining/ProM/7_Dottedgraph%20for%20activityclassifier.PNG

3. Don’t need new configurations. I saw a graph made up of colored dots aligned with their respective resources and classified according to their activities, Through the analysis of the graph we can conclude that some resources start their activities later. For instance, see resources 30, 38 https://github.com/Madhuri97/Data_Science_2019501105/blob/master/Process%20Mining/ProM/7_Shapebydegree.PNG



4. We can conclude that some users are executing a wide variety of activities, while others are focused on some specific types of activities

Q8: 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. There are some process deviations that should be analysed in order to understand if it is necessary to improve the reference process model.
2. The majority of the work is going through a small group of resources, possibly sousing bottlenecks.
3. Although the process has many different variants the event log shows that 80% of the cases go through 5 of all of them, there is a possibility of reducing unnecessary complexity in the process model that should be analysed