

Configuration Manual

MSc Research Project
Master of Science in Data Analytics

James Cookland
Student ID: x20167814

School of Computing
National College of Ireland

Supervisor: Muhammad Zahid Iqbal

National College of Ireland
MSc Project Submission Sheet
School of Computing



Student Name: James Cookland

Student ID: X20167814

Programme: Master of Science in Data Analytics **Year:** 2022

Module: Research Project

Lecturer: Muhammad Zahid Iqbal

Submission Due Date: 01/02/2023

Project Title: Flattening Event Data Extracted From SAP For Process Discovery

Word Count: 1746 **Page Count:** 15

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature: 

Date: 01/02/2023

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	<input type="checkbox"/>
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	<input type="checkbox"/>
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	<input type="checkbox"/>

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Configuration Manual

James Cookland
x20167814

1 Introduction

This manual details the environment, software and software packages required to replicate the flattening of event data extracted from SAP for process discovery. Each section in this manual will detail where each software application and programming language can be downloaded from. Additionally, this manual will introduce how to use each of the applications in order to replicate the results presented in the research paper.

2 Environment Settings

The first step in replicating this research is to set up a remote server with the configuration settings seen in Table 1. Table 1 also displays the applications and Python packages required.

Table 1 - Configuration settings for research machine.

ITEM NAME	DESCRIPTION
Operating System	Windows 10
RAM	16GB
Processor	Intel(R) Xeon(R) Gold 6142 CPU @ 2.60GHz 2.59 GHz (2 processors) processor
SDD	249GB
Applications	Production machine must have the following applications installed: <ul style="list-style-type: none">• Python v3.9.13• QlikView 12 Application• ProM UITopia v6.11• ProM Plugin Manager (PM) v6.11
Required Python Packages	import numpy as np import pandas as pd import os from datetime import datetime import random from datetime import timedelta

3 Qlikview 12

To start, install both the Qlikview 12 application¹ and the Qlikview SAP Connector² onto the research machine. Once installed the SAP connector is instantiated within the Qlikview script as a separate tab in the initialization scripts.

Qlikview scripts are ran as an interpreted piece of code. These scripts are separated into different tabs and ran in succession from left to right as seen in Figure 1.

¹ https://www.qlik.com/us/trial/download-qlikview?_ga=2.258404362.1012470109.1670761552218316698.1670761552

² https://help.qlik.com/en-US/connectors/Subsystems/SAP_Connectors_Help/Content/Connectors_SAP/Installation/Connectorinstallation.htm

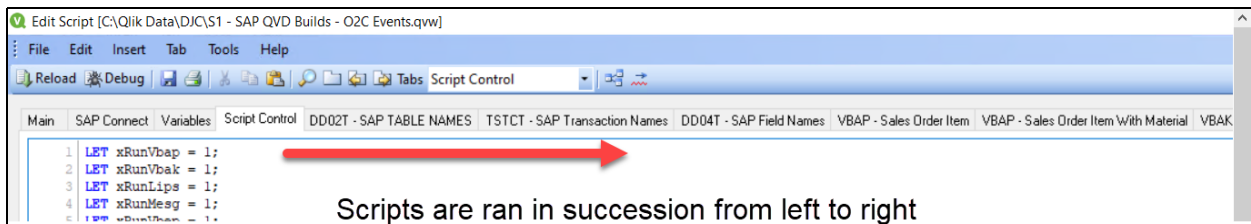


Figure 1 - Image showing which direction Qlikview scripts are interpreted.

Before running scripts, Figure 1 also displays the “Script Control” tab, this tab controls which SQL statements are ran within the script. Variables are named after the SAP tables their respective SQL statements will extract. If a variable is set to 0, the SQL script pertaining to that table will not run, alternatively, setting the variable to 1 will run the relevant SQL statement. This allows the researcher to extract specific tables if troubleshooting is required. The following steps detail how to run the Qlikview application after installation.

1. To run the Qlikview script, click the “reload” button seen in Figure 2. Alternatively pressing “CTRL + R” will run the Qlikview Script.

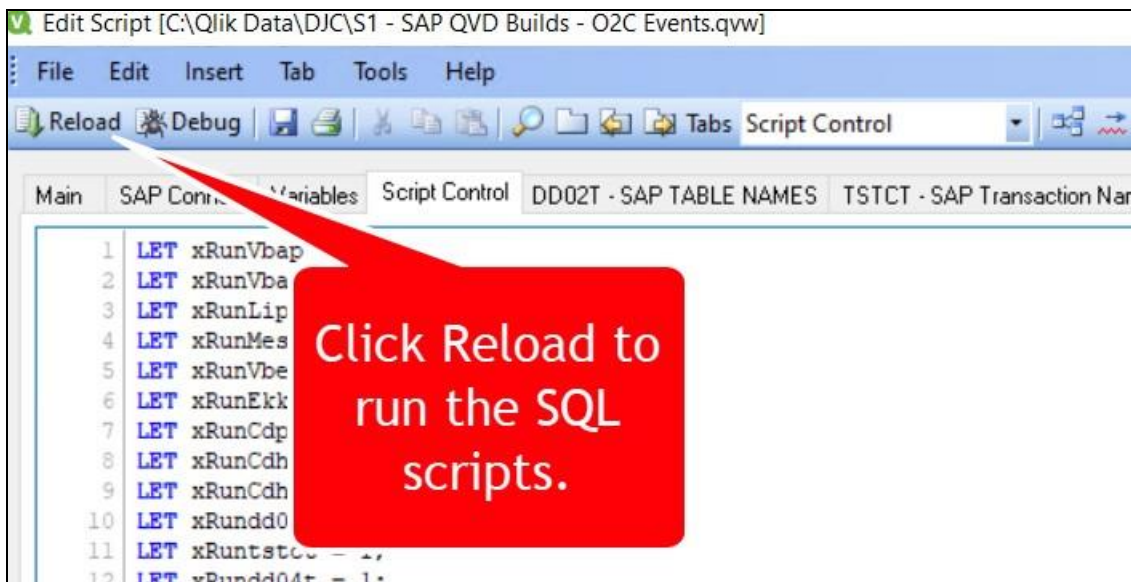


Figure 2 - Image showing how to run Qlikview Scripts

2. Once Qlikview scripts are ran they will save each of the extracted tables to the folder “C:\Qlik Data\DJC” in the local machine. Once extracted these files are then moved to the “C:\Users\{{username}}\Documents\RawData” folder which can then be encrypted while the research machine is not in use to protect the sensitive data contained in the tables as seen in Figure 3.

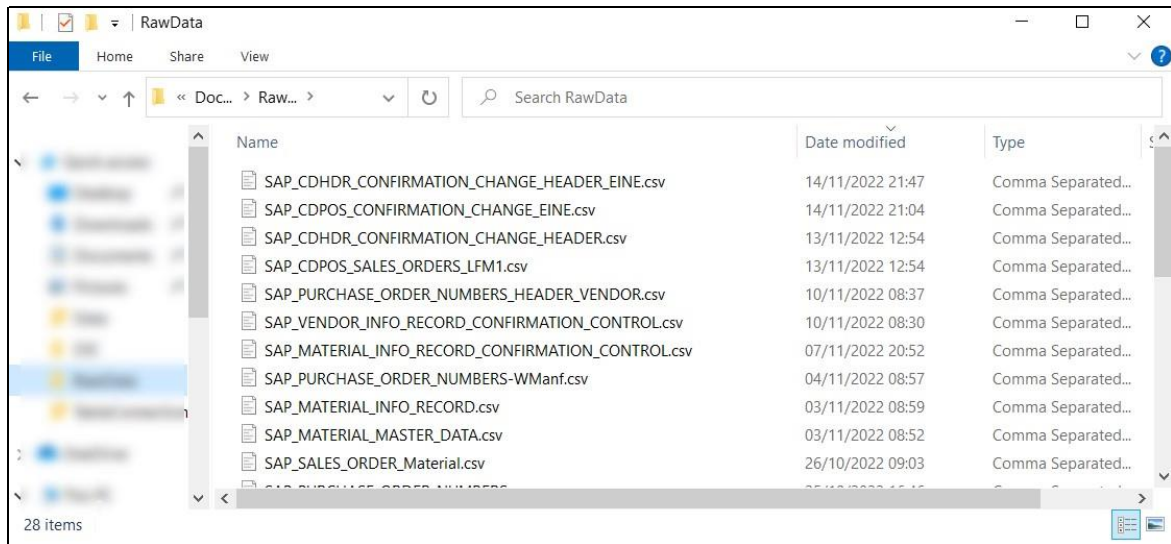


Figure 3 - Image showing the location of all tables extracted from SAP in CSV format.

4 Python Jupyter Notebook

To run the Python code for preparing and clustering the event log data, Python 3.9.13³ was installed. Additionally, Microsoft Visual Studio was used as the programming environment.⁴

Before running the jupyter notebook the following methods need to be installed using the “pip install” command in the research machines command prompt.

3. Numpy
4. Pandas
5. Os
6. Random
7. Datetime

Each cell in the Jupyter notebook can be ran in succession by opening the development.ipynb file and clicking the “Run All” button in the VS Code application, as seen in Figure 4.

³ <https://www.python.org/downloads/release/python-3913/>

⁴ <https://code.visualstudio.com/docs/?dv=win64user>

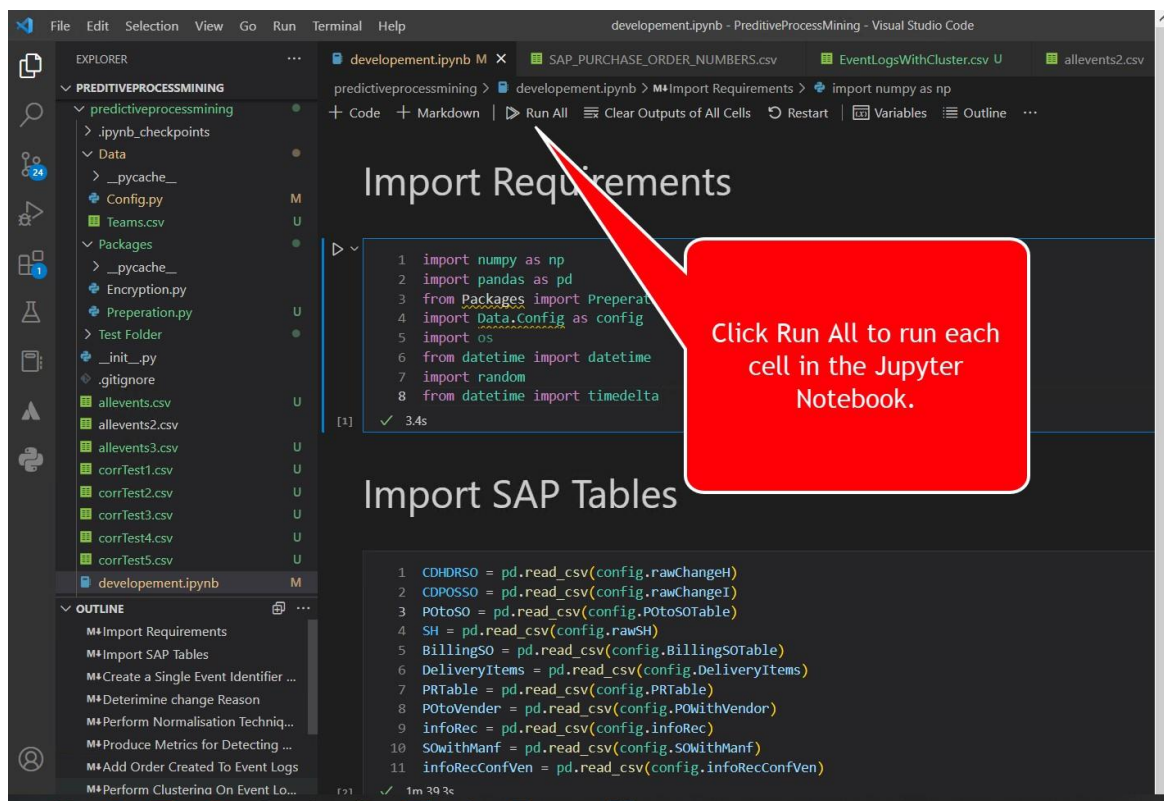


Figure 4 - Image explaining how to run every cell in the notebook.

Once this notebook is run in its entirety, the prepared event logs will be exported as a csv file with the file name “EventLogsWithCluster.csv”. This file is saved to:

“C:\Users\{{username}}\Documents\PredictiveProcessMining\predictiveprocessmining”.

From this point the event logs can be imported to the ProM software.

5 ProM Process Mining

ProM is an open-source application dedicated to the analysis of event logs and the creation of process mining models. To start using ProM, the application must be installed on the research machine.⁵ A full guide on how to install the application and required plugins can be found in ⁶. Once ProM is installed and open, the prepared event logs can be installed by clicking the import button as seen in Figure 6.

The following steps demonstrate how to use the ProM application to generate the same results found in the research report. (Leemans, Poppe and Wynn, 2019)

1. Once ProM is installed and open, the prepared event logs can be installed by clicking the import button as seen in Figure 5.

⁵ <http://promtools.org/doku.php?id=prom611>

⁶

<http://www.promtools.org/doku.php?id=gettingstarted:installation#:~:text=You%20can%20download%20ProM%20%20from%20www.promtools.org%2Fprom6%20or,archive%20to%20a%20file%20on%20your%20local%20drive.>

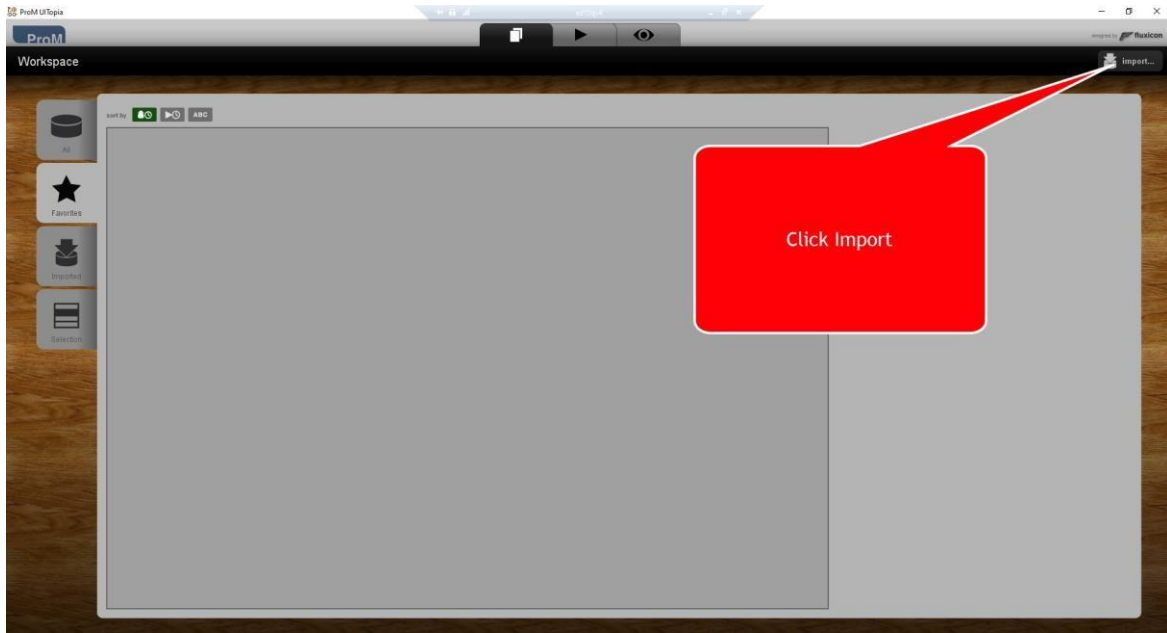


Figure 5 - Image of ProM application for importing event logs.

2. A popup will appear displaying the machines local folders. Locate and select the “C:\Users\{{ username }}\Documents\PredictiveProcessMining\predictiveprocessmining\EventLogsWithCluster.csv” file to import as seen in Figure 6.

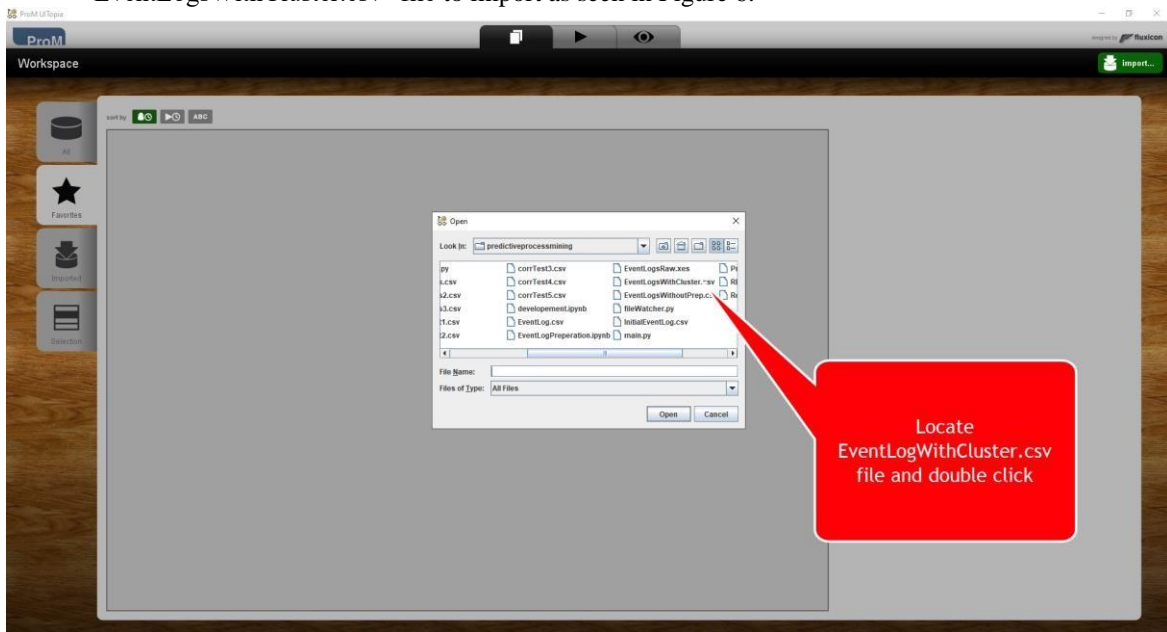


Figure 6 – Image showing how to locate prepared event log CSV file.

3. Another popup will appear asking which file type to import the data as, make sure “CSV File (XES Conversion with Log package)” (this should be the default value) is selected as seen in Figure 7.

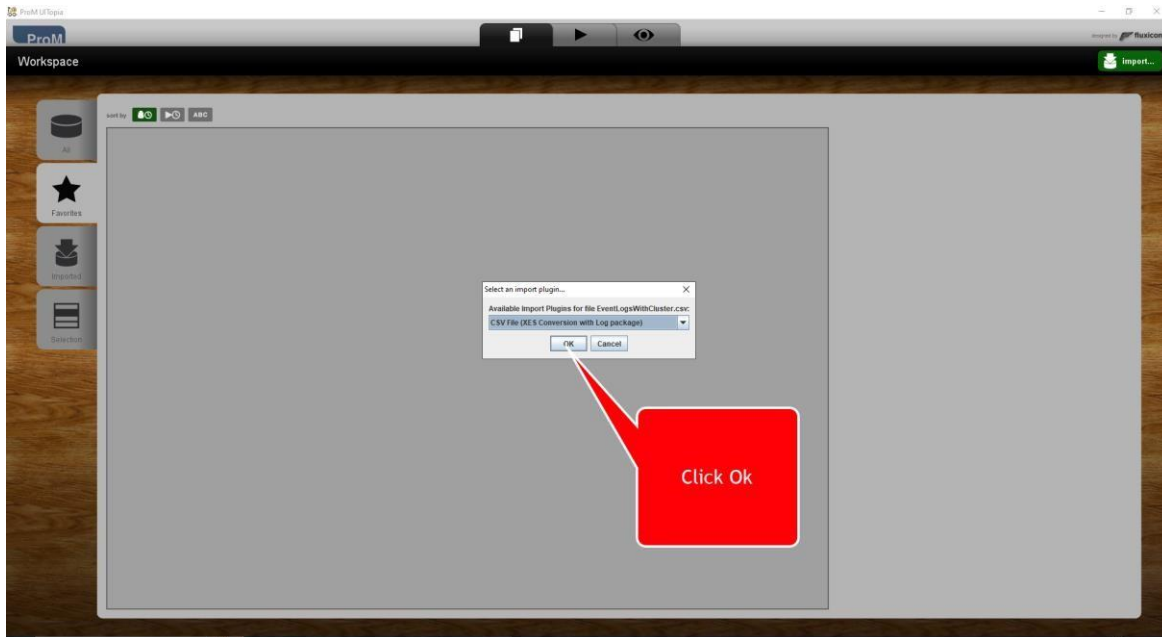


Figure 7 – Image showing how to select to import the file as a CSV file.

4. To convert the CSV file to a XES file, which ProM requires for process mining, click the play button on the Favorites screen as seen in Figure 8.

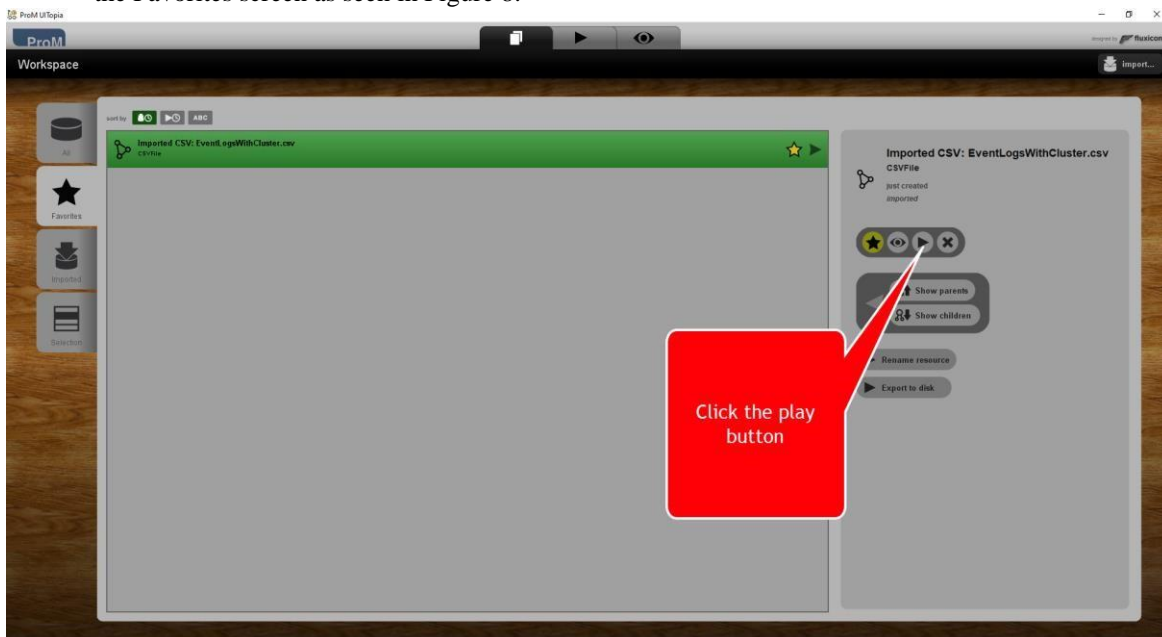


Figure 8 – Image showing how to open the file conversion window.

5. Ensure “Convert CSV to XES” is selected in the next window and click “Start” as in Figure 9.

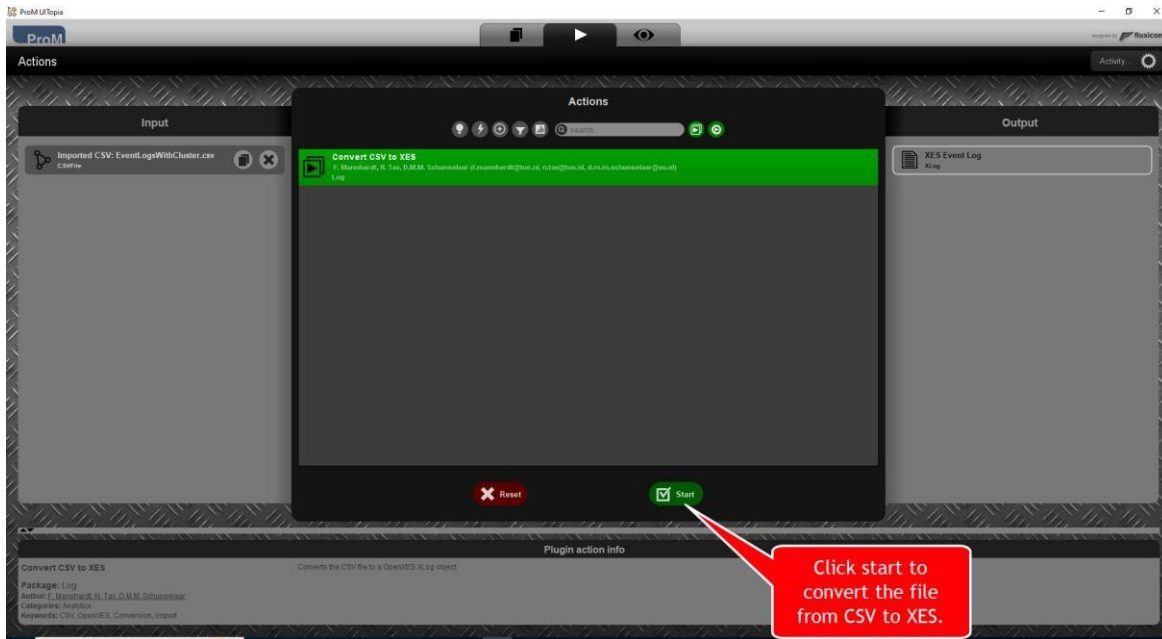


Figure 9 – Image showing how to convert CSV file to XES.

6. Click “Next” on the popup window as in Figure 10, default values are correct.

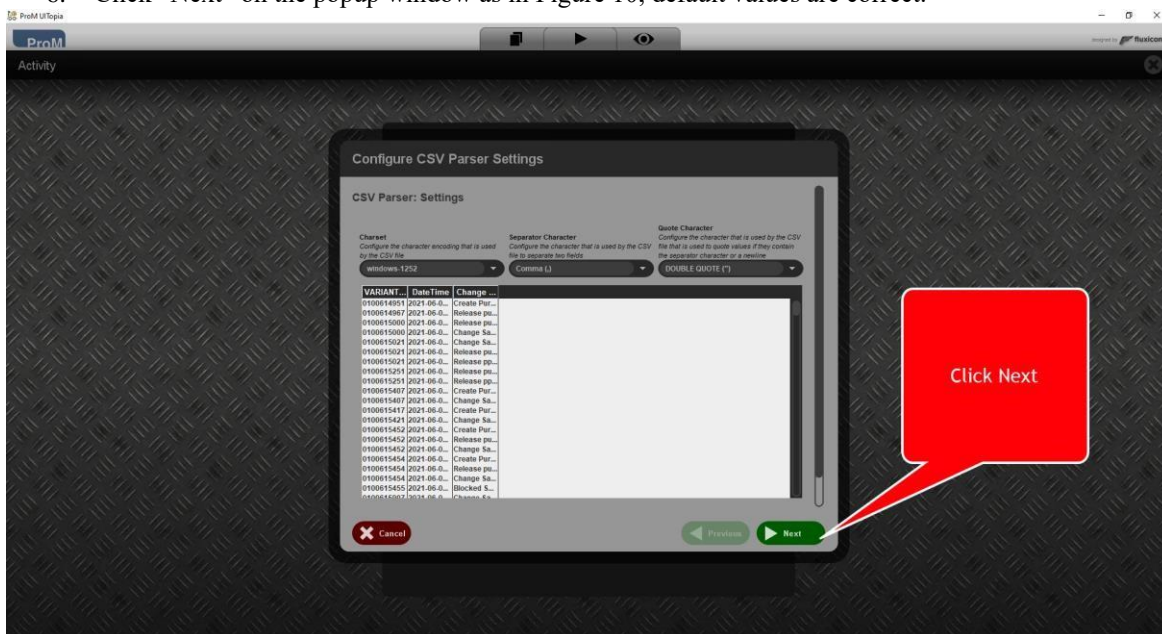


Figure 10 – Image showing how to continue with the conversion options.

7. Next, ensure that “VARIANTNUM” is selected in the Case Column field and “Change Description” is selected in the Event Column field as seen in Figure 11.



Figure 11 – Image showing how to select event log variables for the case and activity fields.

8. Click “Finish” on the next screen seen in Figure 12 to finish the conversion from CSV to XES.

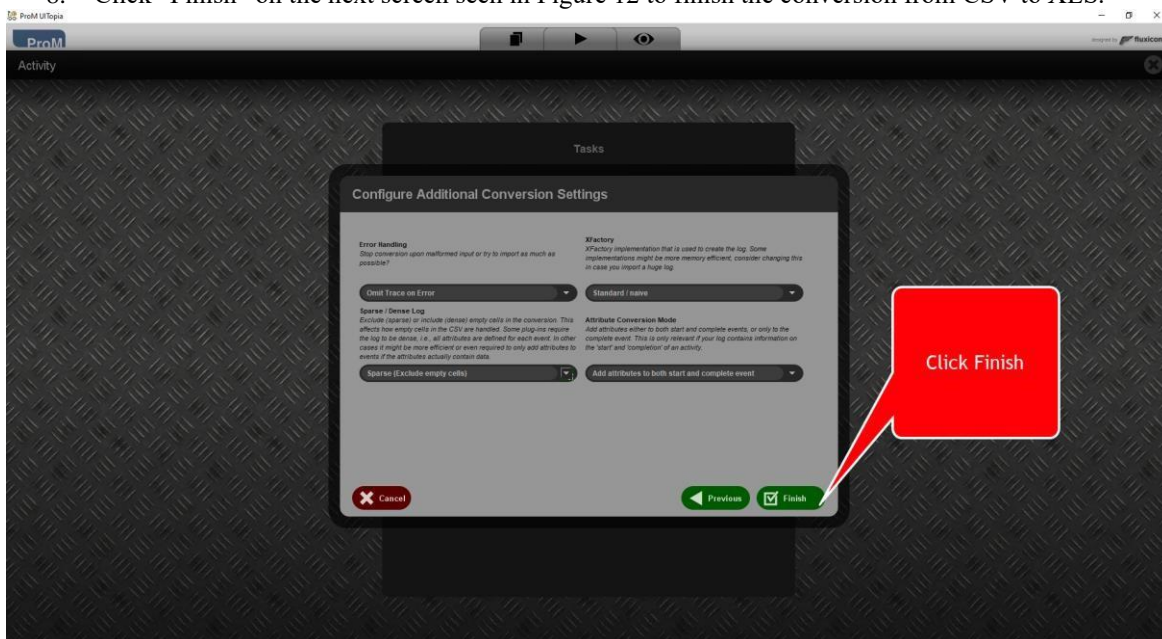


Figure 12 – Image showing how to finish conversion from CSV to XES.

9. Event logs should now be imported and converted to the XES file type. Events must now be filtered. To start, click on the “Select visualization” dropdown as seen in Figure 13 and select “Event Filter on Event Log (Filtered)”.



Figure 13 – Image showing how to select filter plugin.

10. Click “Click to add filter” box in the filter event logs plugin as seen in Figure 14.

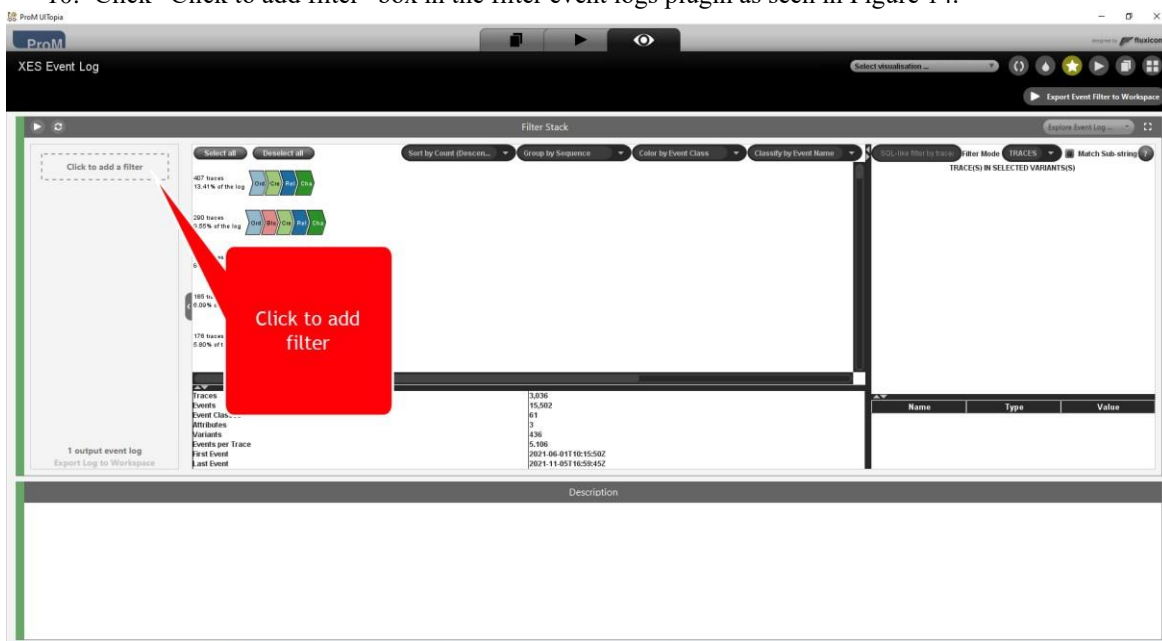


Figure 14 – Image showing how to add filter to event logs.

11. Select the “Select traces by frequency” filter in the next screen as seen in Figure 15.

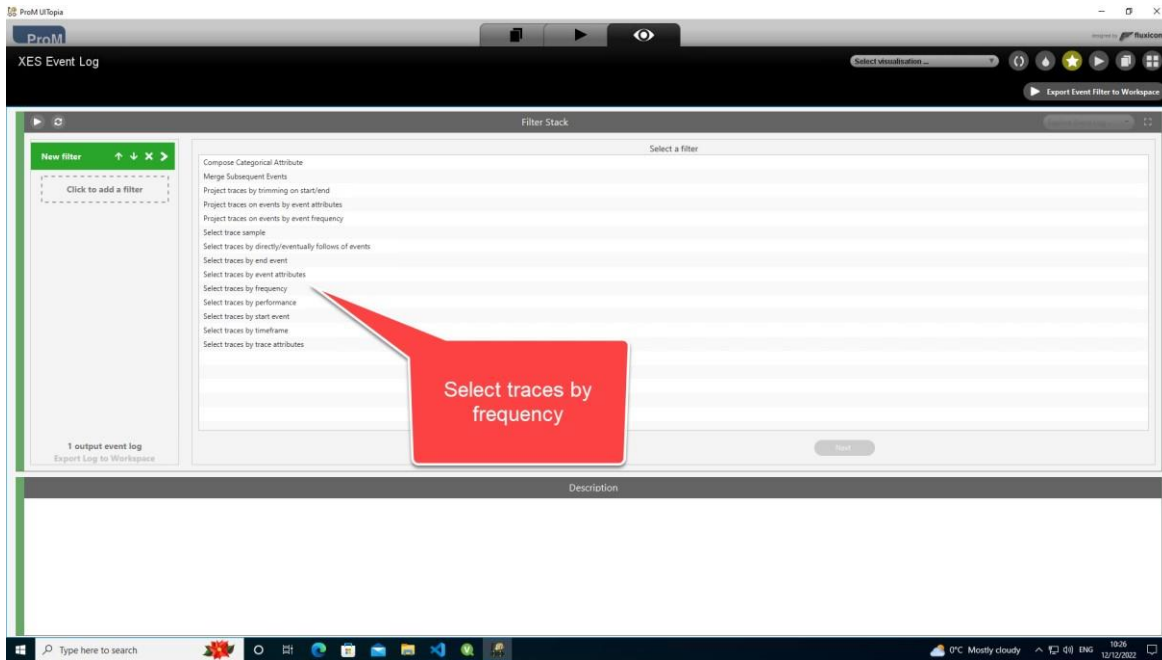


Figure 15 – Image showing how to select traces by frequency filter.

12. Next, change the Threshold type as “occurrence” instead of “frequency” as seen in Figure 16.

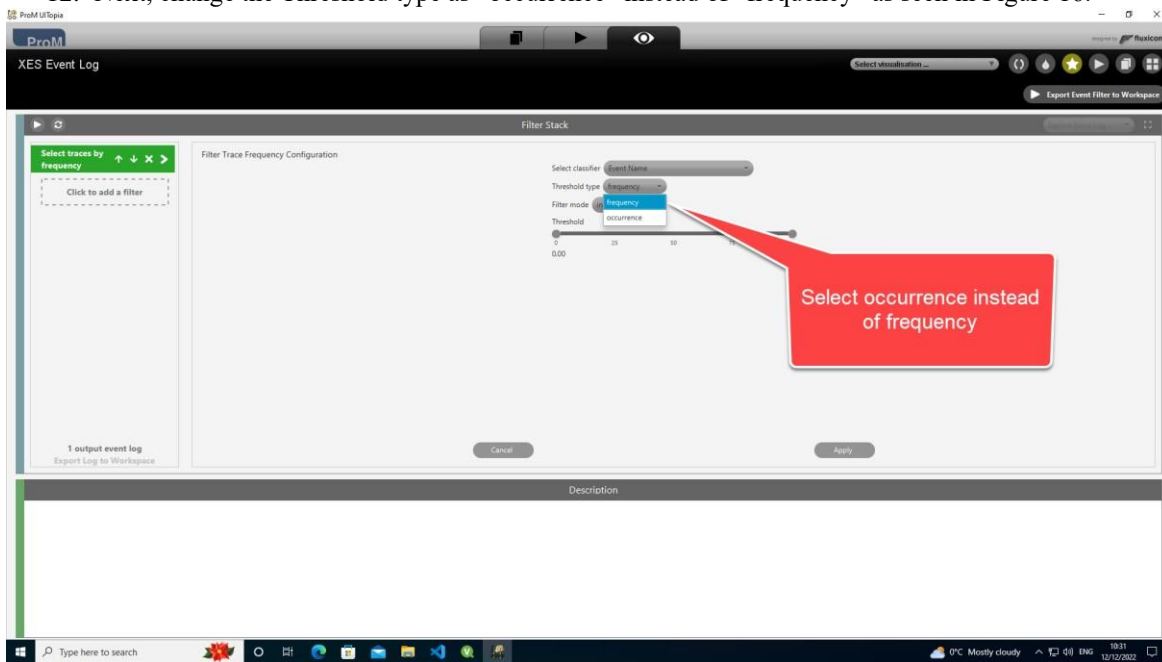


Figure 16 – Image showing how to select Frequency as Threshold type.

13. Now drag threshold scroller on the left from 0 to 5 as seen in figure 17.

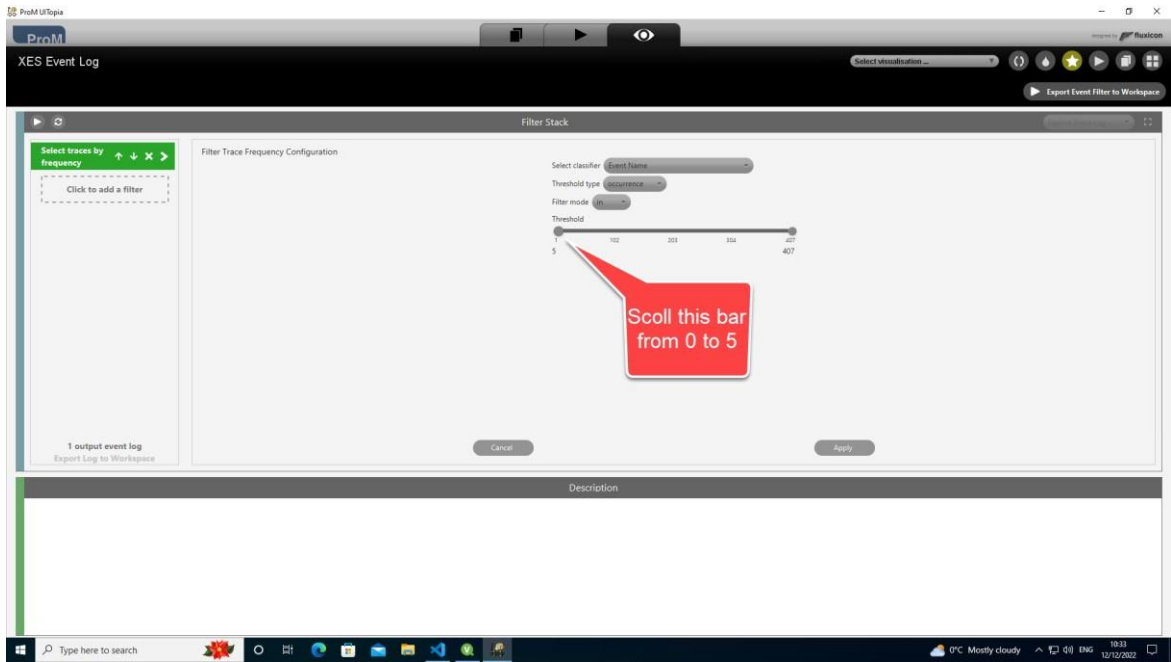


Figure 17 – Image showing how to select a minimum threshold of 5 occurrences.

14. Click apply to set the filter as seen in Figure 18.

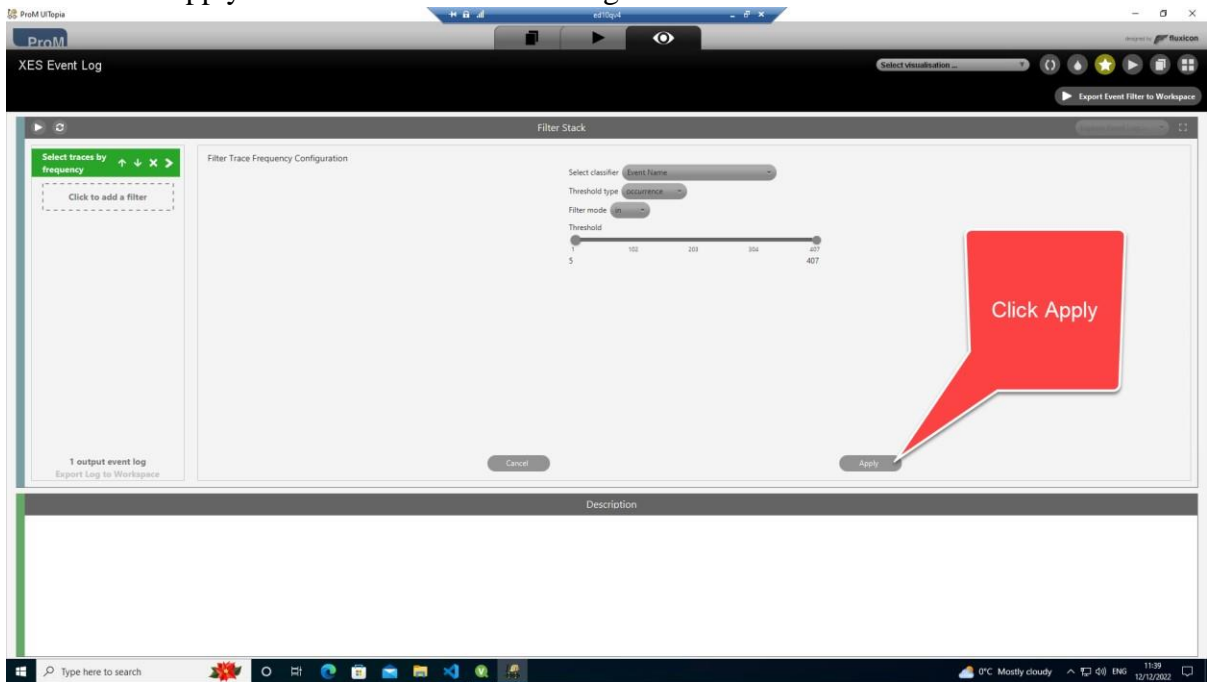


Figure 18 – Image showing how to select Apply to set the filter

15. Click the play button seen in Figure 19 to apply the filter.

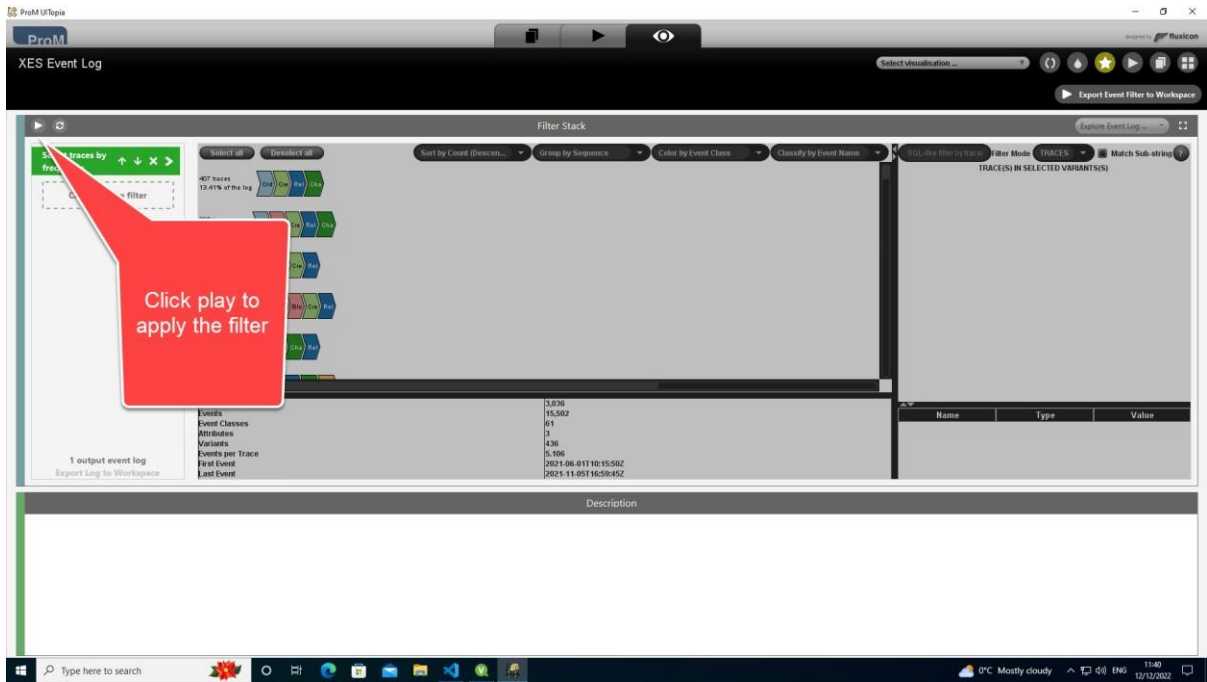


Figure 19 - Image showing how to apply the filter.

16. Click "Export Log to Workspace" as seen in Figure 20



Figure 20 - Image showing how to export the filtered event log to the ProM workspace

17. Return to the actions screen and reset the action window as seen in figure 21.

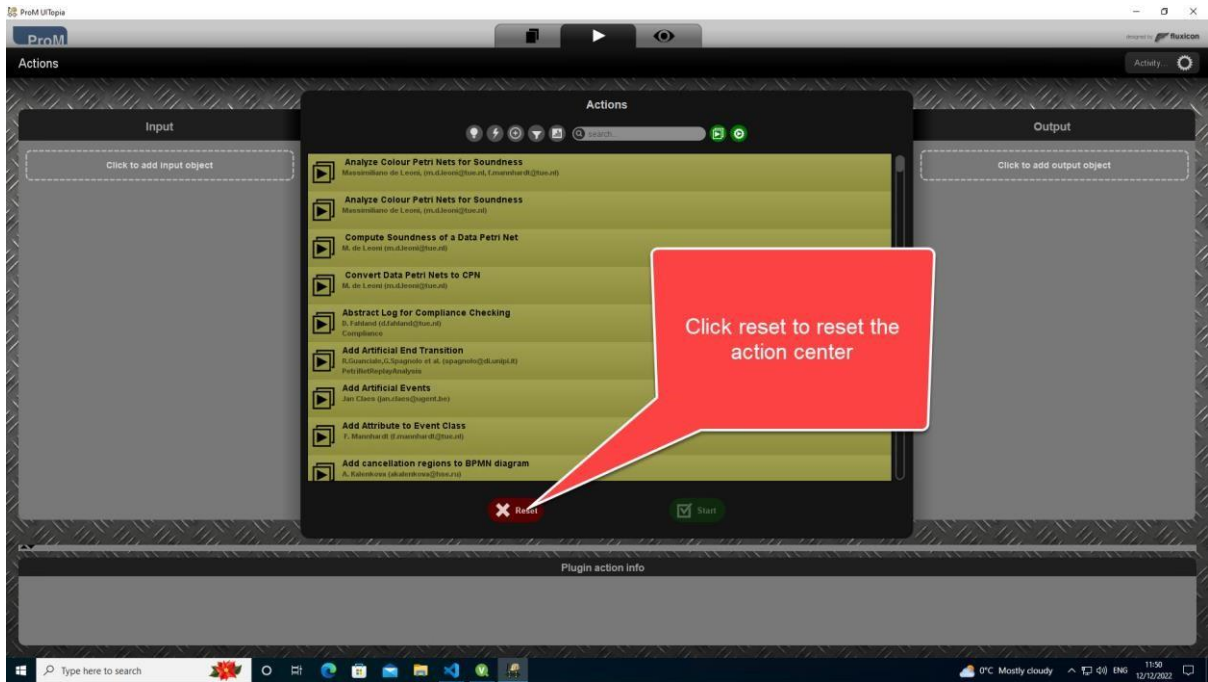


Figure 21 – Image Showing how to reset action window

18. Search for and select the “Mine with Directly Follows visual Miner” as seen in Figure 22.

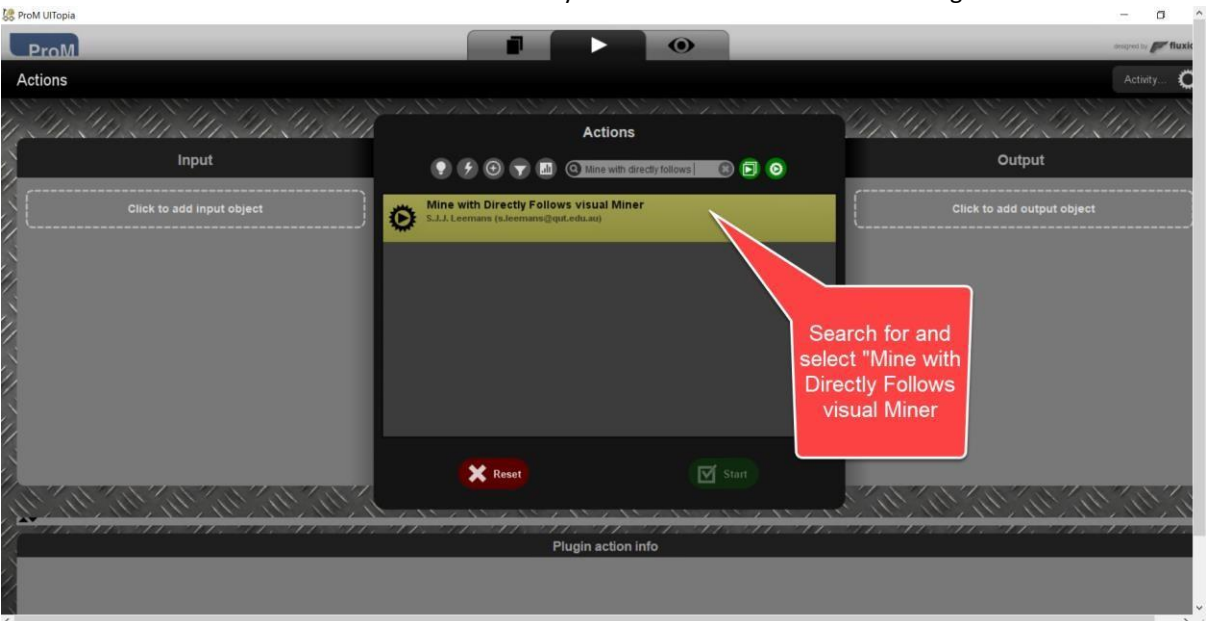


Figure 22 - Image showing the how to select the Mine with Directly Follows visual Miner plugin.

19. Click on “Click to add input data and select the filtered event logs to run the directly follows miner on as seen in Figure 23.

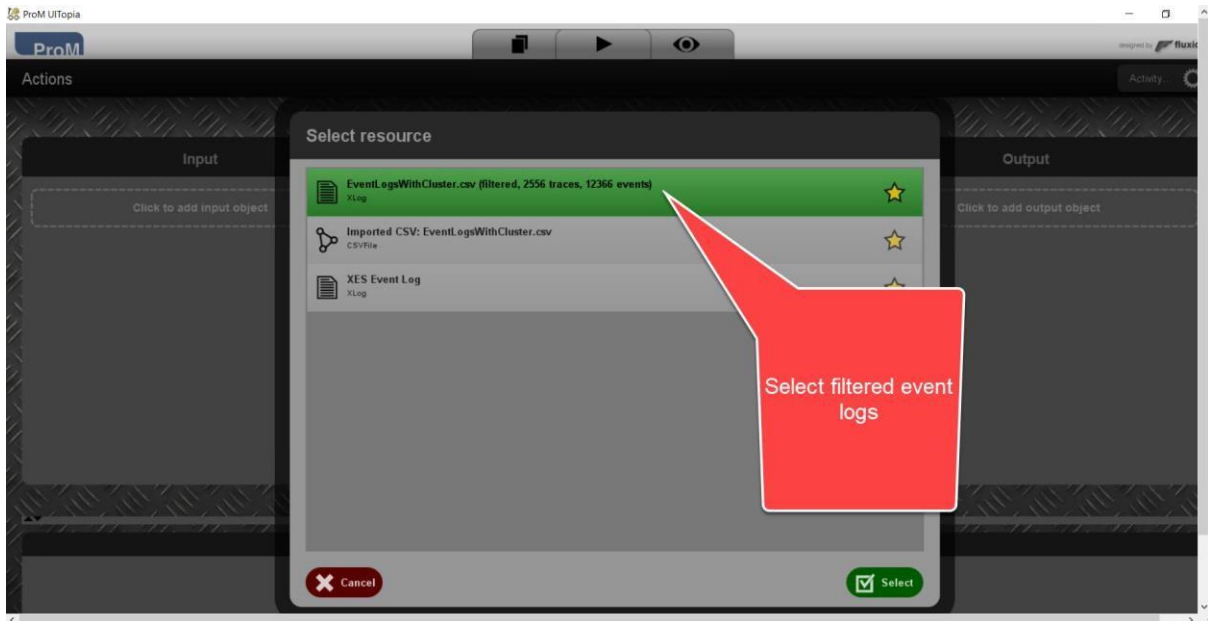


Figure 23 - Image showing how to select the input data for the directly follows visual miner.

20. Click “Start” to run the directly follows miner on the filtered data as seen in Figure 24.

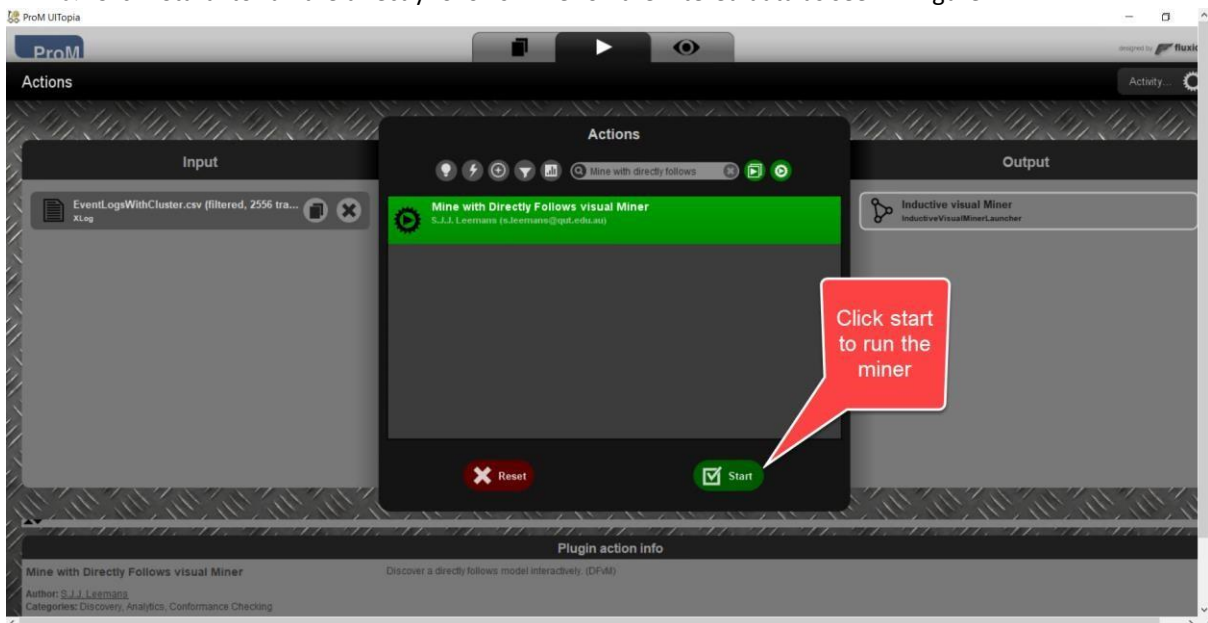


Figure 24 - Image showing how to run the directly follows visual miner.

21. As seen in Figure 25, by selecting “paths and deviations”, the conformance of the resulting model can be visualized. The resulting model can now be visualized.

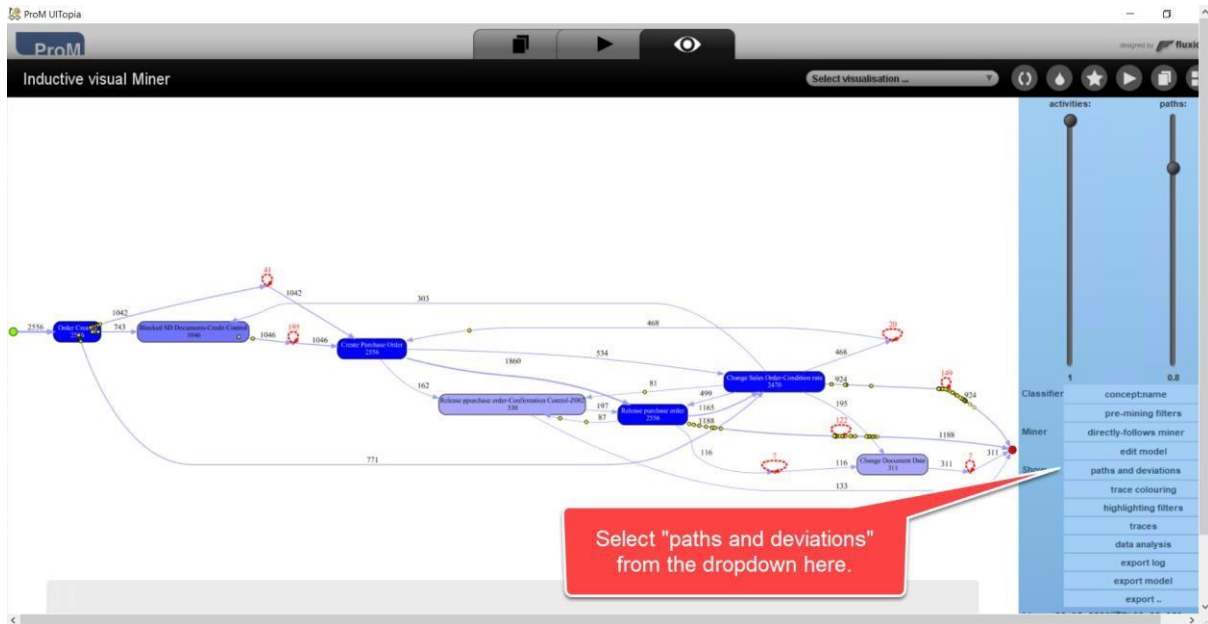


Figure 25 - Image showing how the resulting process mining model and how to display deviations in the model.

Images were generated using Snaggit.⁶

References

Leemans, S. J. J., Poppe, E. and Wynn, M. T. (2019) 'Directly Follows-Based Process Mining: Exploration & a Case Study', in *2019 International Conference on Process Mining (ICPM)*, pp. 25–32. doi: 10.1109/ICPM.2019.00015.

⁶ <https://www.techsmith.com/screen-capture.html>