Louis Paulet

Episode Mining
Autonomous work

Question 1.

The detailed process of data transformation is described in the jupyter notebook called :

- "PMSNA_HW1.ipynb"

This version is easier to understand as the DataFrames containing the dataset transformations are outputted at each step of the processing.

And the "single function" version, with comment in place of text (but harder to understand) is named :

- "PMSNA_HW1_oneFunction.ipynb"

This version is compliant with the specification of a "Python Function".

In both cases the parameters are the ones supplied in the instructions, and the results are the same.

Question 2.

I selected the best 10 episodes based on a harmonic mean(in the spirit of the F1 score, to minimize the outliers' influence). You can find the detailed results in the included notebooks.

Here is the short version (with the columns "X" and "Y" representing relations "X -> Y") :

Louis Paulet

X	Y
Boeing=-1	Boeing=1
General_Motors=-1	Exxon_Mobil=1
Microsoft=-1	Boeing=1
Microsoft=-1	General_Motors=-1
General_Motors=-1	General_Motors=1
AT&T=-1	Exxon_Mobil=1
Verizon=-1	American_Express=-1
Boeing=1	Exxon_Mobil=-1
Verizon=-1	AT&T=-1
Intel=-1	General_Motors=-1

Top 10 Association Rules by Harmonic Mean(Confidence, Recall)