**CS 634 Data Mining Midterm Project**

**Name: Rahman Mhate**

**Date: 11/15/2021**

**GitHub:** <https://github.com/RahmanMhate/DataMining>

**Py Apriori Module:**

This is a basic implementation in Python of the Apriori Association Algorithm

**Requirements:**

**Software:** Python 3.7

**Libraries:** Pandas library is used to import the CSV file.

Itertools used to iterate over data structures that can be stepped over using a for-loop

NumPy used for working with arrays.

**Setup**:

Download the TAR/ZIP file from google to a location where you want to save.

Double-click the icon labelling the file python-3.7.9-amd64.exe.

A Python 3.7.9 (64-bit) Setup pop-up window will appear.

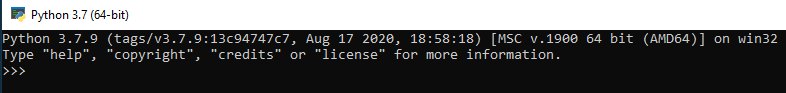
1. Ensure that the Install launcher for all users (recommended) and the Add Python 3.9 to PATH checkboxes at the bottom are checked.
2. Highlight the Install Now (or Upgrade Now) message, and then click it.
3. Click the Yes button.

During installation, it will show the various components it is installing and move the progress bar towards completion. Soon, a new Python 3.7.9 (64-bit) Setup pop-up window will appear with a Setup was successfully message.

1. Click the Close button.

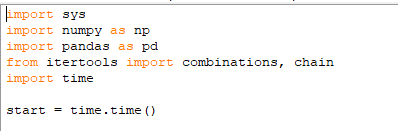
Python should now be installed.

To confirm installation, open the .exe file from the location

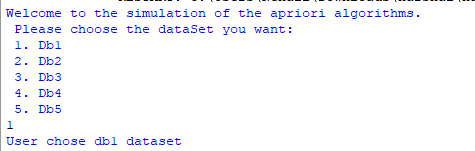


**Installation:**

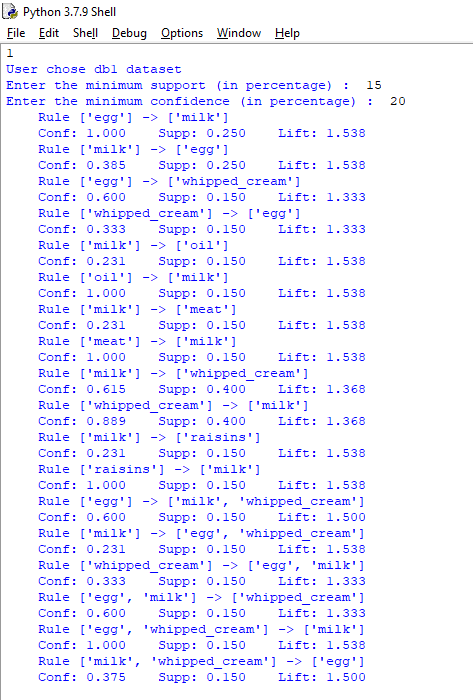
1. Install the necessary packages numpy, pandas, os, Itertools, csv, combinations
2. Once the packages are installed start with the imports



1. Run the .py file which is generated by Jupyter notebook to select the dataset and retrieve the support and confidence based on user input.



1. Once the user has selected the dataset and the minimum support and confidence is provided the following output is generated

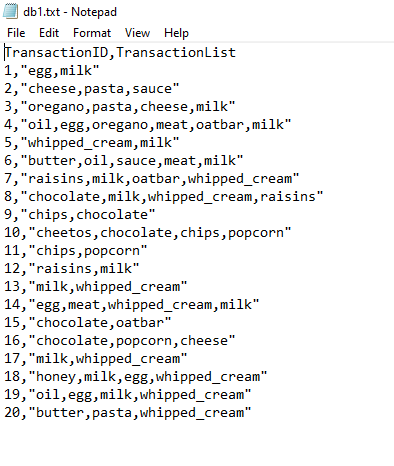


**Project Output:**

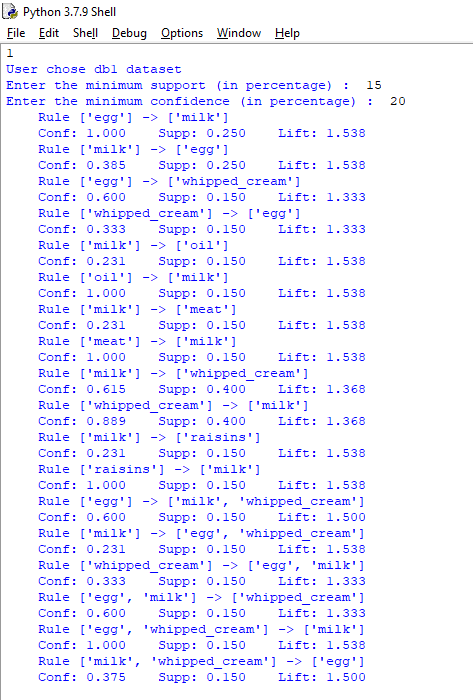
The following are the datasets and the generated output from the program.

1. **Db1**

Data File:

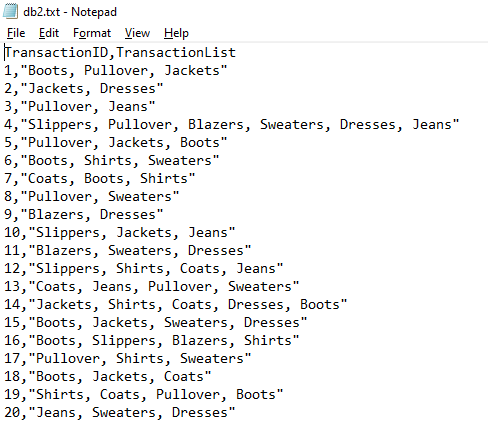


Output:

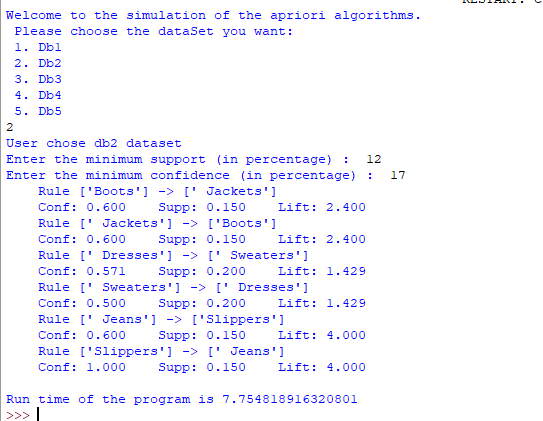


1. **Db2**

Data File:



Output:

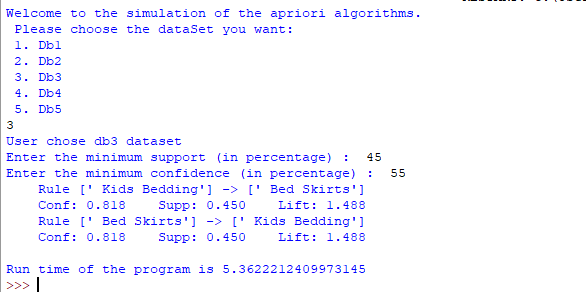


1. **Db3**

Data File:

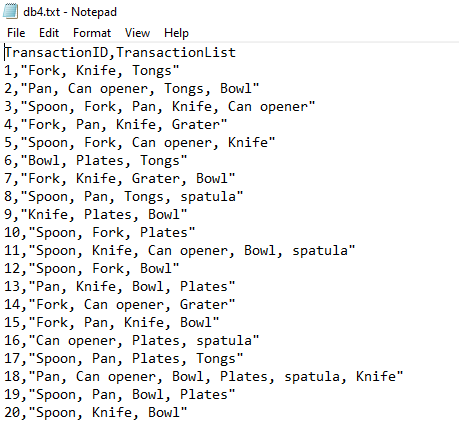


Output:

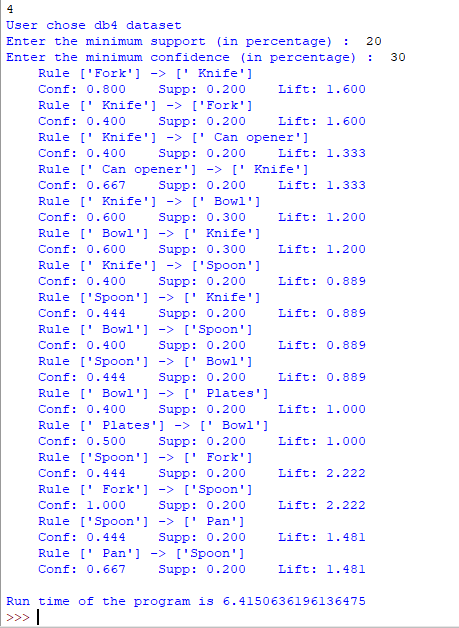


1. **Db4**

Data File:

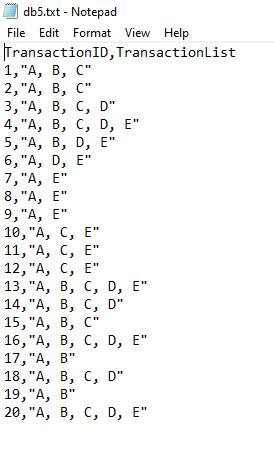


Output:

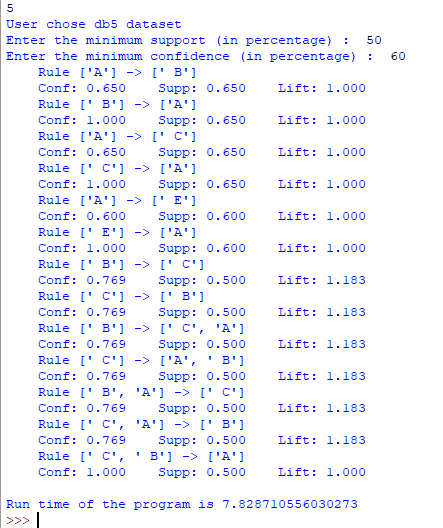


1. **Db5**

Data File:

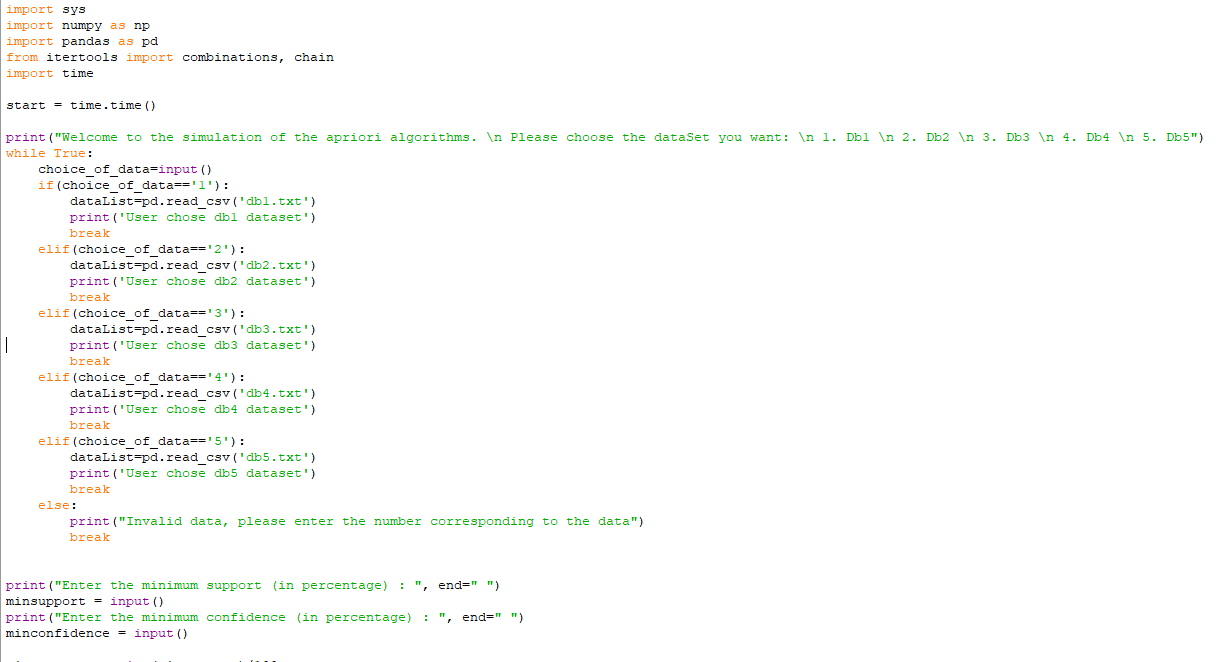


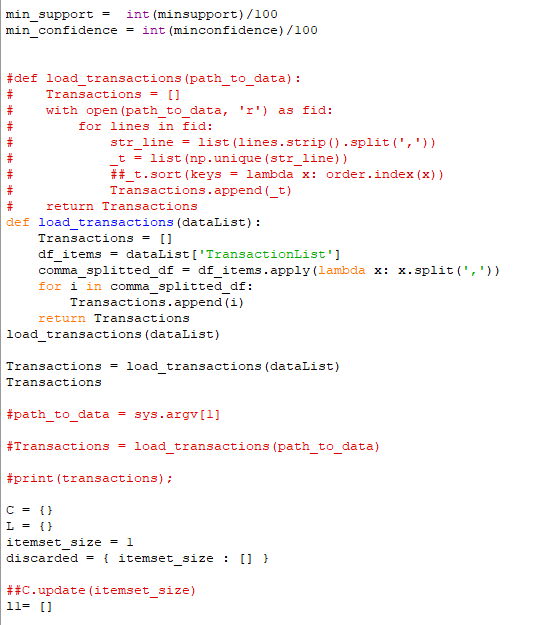
Output:



**Program Files**

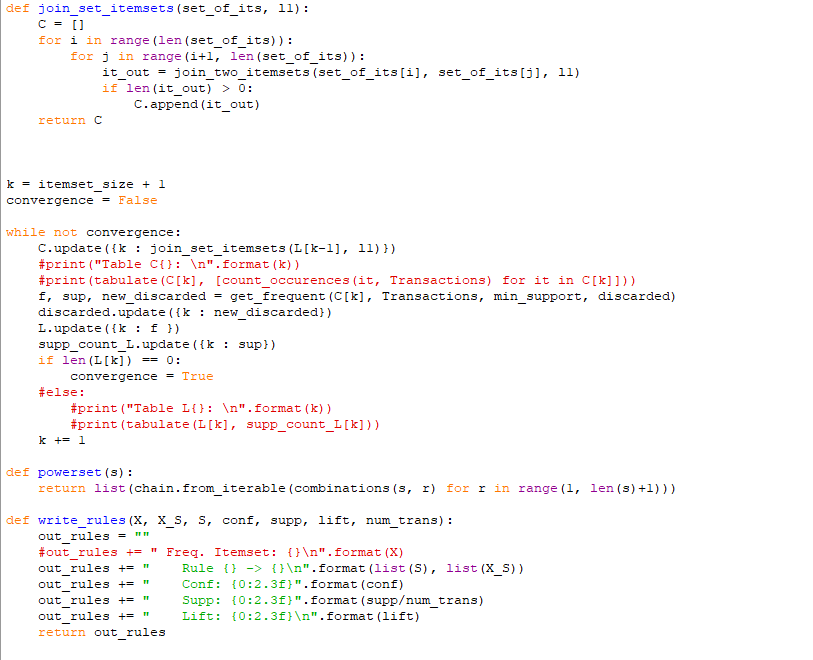
rahman.py

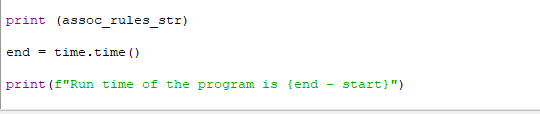












**Reference**

[**https://en.wikipedia.org/wiki/Apriori\_algorithm**](https://en.wikipedia.org/wiki/Apriori_algorithm)

[**https://opensource.com/article/18/1/step-step-guide-git**](https://opensource.com/article/18/1/step-step-guide-git)

[**https://en.wikipedia.org/wiki/Association\_rule\_learning**](https://en.wikipedia.org/wiki/Association_rule_learning)