PA 4: Association Analysis - Apriori Algorithm

Student Details

Student Name and ID: <-------only this student will upload the assignment

Team member name and ID:

NO REPORT REQUIRED

Submission Instructions

Step 1: Create a folder and name it 'lastname_firstname_yourNetID_PA4'

Step 2: Rename this submission file as 'lastname_firstname_yourNetID_PA4.ipynb' and place it inside the folder 'lastname_firstname_yourNetID_PA4'

Step 3: Rename the updated dataset file 'dataset.csv' and place it inside the folder 'lastname_firstname_yourNetID_PA4'

Step 4: Your submission folder should include ONLY the following files:

```
* apriory.py,
* lastname_firstname_yourNetID_PA4.ipynb,
* dataset.csv,
* toyDS.csv
```

Step 5: Zip this folder and submit it on Canvas. Your final submission folder name should be 'lastname_firstname_yourNetID_PA4.ZIP'

Programming Assignment Details

Before you start:

- Be familiar with the algorithm and with the dataset.
- If you use external sources make sure that you cite them, and be specific!
- Make sure that your code is running before you upload your submission file. TA will not debug your code.
- Start early!

For this assignment, you will have to use:

- Jupyter notebook,
- the 'Random Shopping cart' dataset [01],
- and the Apriori Algorithm (apriory.py)[02]. Note that the apriory.py file is modified to run with Python 3.

- SOLUTION -

In [1]:
```
%%javascript
IPython.OutputArea.prototype._should_scroll = function(lines) {
    return false;
}
```

In [2]:
```
# Import your Libraries
```

Task 1: DataSet Preprocess

Before you start you need to modify your dataset 'dataset_group.csv' to look like the toyDS.csv. Each transaction is at one line with a variable length. Discard the date attribute (1st attribute) from your dataset. For example, in your dataset transaction#4 should look like:

```
cereals,juice,lunch meat,soda,toilet paper,all-purpose
```

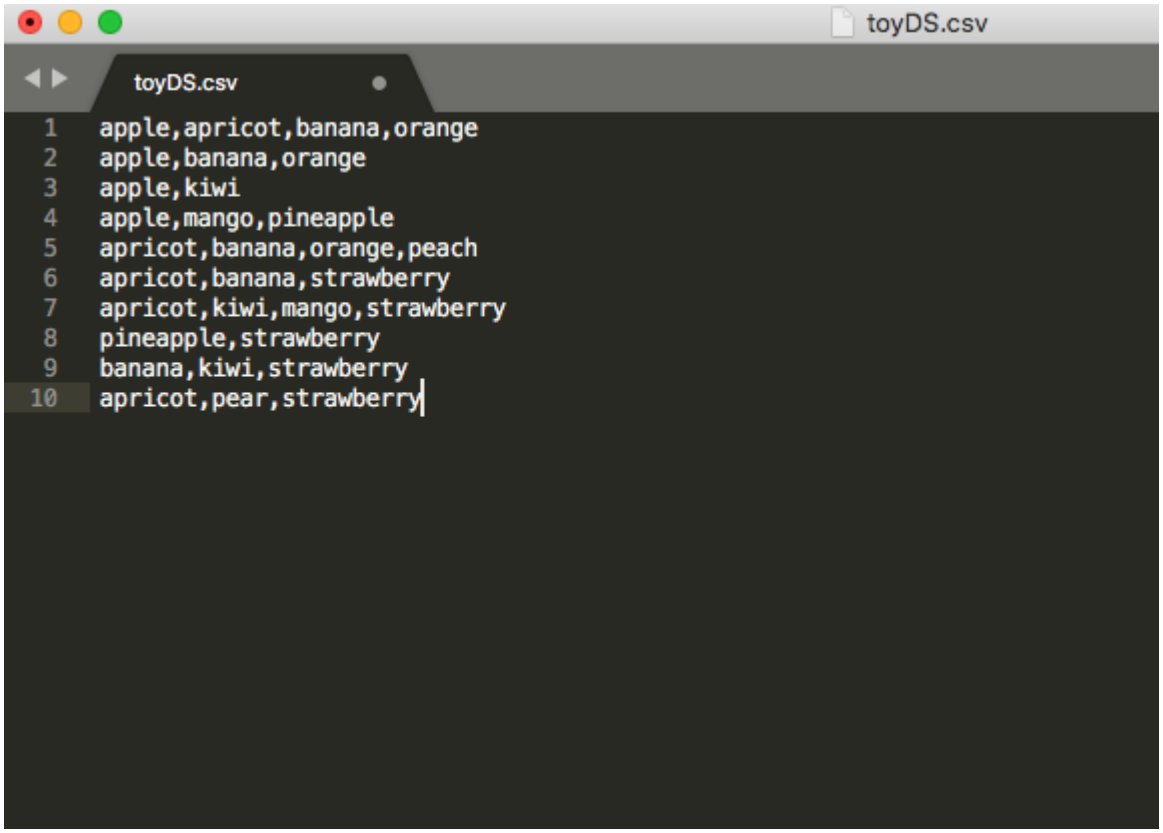
Export your modified dataset in a file named 'dataset.csv'.

Use pandas to Read and Print the first 7 transactions of the 'dataset.csv'.

In [3]:
```
from IPython.display import Image
print ('ScreenShot of the toyDS.csv')
Image("SampleScreen01.png")
```

```
ScreenShot of the toyDS.csv
```

Out[3]:



In [4]:
```
# # # # # # # # #  Code for Task 1 # # # # # # # # #
```

In [5]:
```
# # #  Solution  # # #
```

Task 2: Run apriory.py and Evaluate Results

In this task, you have to find how you will be able to execute and print apriory results by making use only the apriory.py. In other words "DO NOT USE ANY OTHER LIBRARY FOR TASK 2!!!".

(*) For those that are not familiar with python and coding this could be a quite demanding task.

You will have to execute apriori algorithm "3" times for different combinations of support and confidence. Print the results of apriory for 'dataset.csv' by making use ONLY the provided methods.

Do not forget to add your reasoning (explain the result outcome) at the top of each case in a nice and readable way.

You are allowed to use the python print method to print your results. DO NOT add your reasoning as comments.

In [6]:
```
print ('# # # # # # # # #  Code for Task 2, Case:1 # # # # # # # # #')
print ('Case 1 (minimum support=XX and minimum confidence=YY)')
print ('Case 1 Reasoning:')
print ('Case 1 Output:')
```

```
# # # # # # # # #  Code for Task 2, Case:1 # # # # # # # # #
Case 1 (minimum support=XX and minimum confidence=YY)
Case 1 Reasoning:
Case 1 Output:
```

In [7]:
```
print ('# # # # # # # # #  Code for Task 2, Case:2 # # # # # # # # #')
print ('Case 2 (minimum support=XX and minimum confidence=YY)')
print ('Case 2 Reasoning:')
print ('Case 2 Output:')
```

```
# # # # # # # # #  Code for Task 2, Case:2 # # # # # # # # #
Case 2 (minimum support=XX and minimum confidence=YY)
Case 2 Reasoning:
Case 2 Output:
```

In [8]:
```
print ('# # # # # # # # #  Code for Task 2, Case:3 # # # # # # # # #')
print ('Case 3 (minimum support=XX and minimum confidence=YY)')
print ('Case 3 Reasoning:')
print ('Case 3 Output:')
```

```
# # # # # # # # #  Code for Task 2, Case:3 # # # # # # # # #
Case 3 (minimum support=XX and minimum confidence=YY)
Case 3 Reasoning:
Case 3 Output:
```

References

[01] https://www.kaggle.com/acostasg/random-shopping-cart (https://www.kaggle.com/acostasg/random-shopping-cart)

[02] https://github.com/asaini/Apriori (https://github.com/asaini/Apriori)

Rubric

- [02 points] - Student Details
- [08 points] - Comply with submission instructions
- [30 points] - DataSet Preprocess
- [30 points] - Run apriory.py
- [30 points] - Evaluate Results