PA 3: 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_PA3'

Step 2: Rename this submission file as 'lastname_firstname_yourNetID_PA3.ipynb' and place it inside the folder 'lastname_firstname_yourNetID_PA3'

Step 3: Rename the updated dataset file 'dataset.csv' and place it inside the folder 'lastname_firstname_yourNetID_PA3'

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

```
* lastname_firstname_yourNetID_PA3.ipynb,
* dataset.csv,
* fruits.csv
* Apriori algorithm (.py file)
```

Step 5: Zip this folder and submit it on Canvas. Your final submission folder name should be 'lastname_firstname_yourNetID_PA3.ZIP'

Programming Assignment Details

Before you start:

- Be familiar 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 'Groceries' dataset [01],
- and the Apriori Algorithm [02] (You can use the algorithm provided as reference)

Note:

- The algorithm to attached in the end, has to be reffered, and make sure to code it on your own.
- Any plagiarism detected will be subjected to additional consequences.

- SOLUTION -

In [13]:
```
%%javascript
IPython.OutputArea.prototype._should_scroll = function(lines) {
    return false;
}
```

In [14]:
```
# Import your Libraries
```

Task 1: DataSet Preprocess

Before you start you need to modify your dataset 'dataset.csv' to look like the fruits.csv. Each transaction is at one line with a variable length. Discard the date attribute from your dataset.

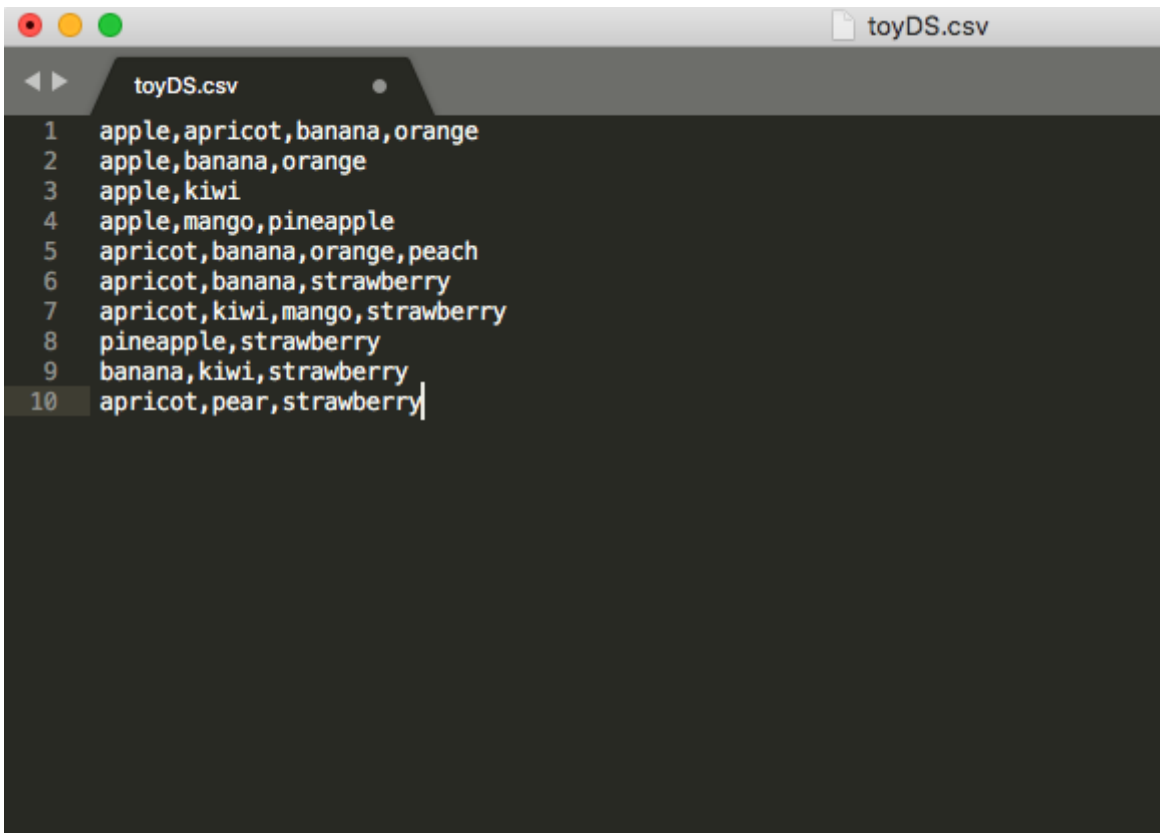
Export your modified dataset in a file named 'dataset_updated.csv'.

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

In [15]:
```
from IPython.display import Image
print ('ScreenShot of the fruits.csv')
Image("SampleScreen01.png")
```

```
ScreenShot of the fruits.csv
```

Out[15]:



In [16]:
```
# # # # # # # # #  Code for Task 1 # # # # # # # # #
```

In [17]:
```
# # #  Solution  # # #
```

Task 2: Implement apriory algorithm and Evaluate Results

In this task, you have to find how you will be able to execute and print apriory results. (*) For those that are not familiar with python and coding this could be a quite demanding task.

- Use this as a reference to use the apriori algorithm.[https://github.com/asaini/Apriori (https://github.com/asaini/Apriori)]

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

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 or observation as comments.

In [18]:
```
print ('# # # # # # # # #  Code for Task 2, Case:1 # # # # # # # # #')
print ('Case 1 (minimum support=XX and minimum confidence=YY)')
print ('Case 1 Reasoning: Put Your Reasong here')
print ('Case 1 Output:')
print ('Case 1 Visualization: Graph & Observations')
```

```
# # # # # # # # #  Code for Task 2, Case:1 # # # # # # # # #
Case 1 (minimum support=XX and minimum confidence=YY)
Case 1 Reasoning: Put Your Reasong here
Case 1 Output:
Case 1 Visualization: Graph & Observations
```

In [19]:
```
print ('# # # # # # # # #  Code for Task 2, Case:2 # # # # # # # # #')
print ('Case 2 (minimum support=XX and minimum confidence=YY)')
print ('Case 2 Reasoning: Put Your Reasong here')
print ('Case 2 Output:')
print ('Case 2 Visualization: Graph & Observations')
```

```
# # # # # # # # #  Code for Task 2, Case:2 # # # # # # # # #
Case 2 (minimum support=XX and minimum confidence=YY)
Case 2 Reasoning: Put Your Reasong here
Case 2 Output:
Case 2 Visualization: Graph & Observations
```

In [20]:
```
print ('# # # # # # # # #  Code for Task 2, Case:3 # # # # # # # # #')
print ('Case 3 (minimum support=XX and minimum confidence=YY)')
print ('Case 3 Reasoning: Put Your Reasong here')
print ('Case 3 Output:')
print ('Case 3 Visualization: Graph & Observations')
```

```
# # # # # # # # #  Code for Task 2, Case:3 # # # # # # # # #
Case 3 (minimum support=XX and minimum confidence=YY)
Case 3 Reasoning: Put Your Reasong here
Case 3 Output:
Case 3 Visualization: Graph & Observations
```

References

[01] https://www.kaggle.com/heeraldedhia/groceries-dataset (https://www.kaggle.com/heeraldedhia/groceries-dataset)

[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 apriori algorithm
- [30 points] - Evaluate Results