PROJECT REPORT

<u>ASSOCIATE RULE MINING: MARKET BASKET ANALYSIS</u>
<u>NAME: RISHI RAJ SINGH</u>
<u>AI AND ML B-4</u>

Problem Statement:
Apriori is a statistical algorithm for implementing associate rule mining, that primarily relies on three components: Life, Support, and Confidence. Using this algorithm tries to find the rules that describe the relationship between each of the products that were brought by the customers as described in.

Prerequisites:
● Python 3

Libraries Used:
● Pandas
● Numpy
● Matplotlib
● Seaborn

Method :
Association Rule Mining is used when we want to find an association between different objects in a set or find frequent patterns in a transaction database or relational databases. The applications of Association Rule Mining are found in Marketing, Basket Data Analysis (or Market Basket Analysis) in retailing, clustering, and classification. It can be used to find what items do customers frequently buy together by generating a set of rules called Association Rules.

Implementation :

- Importing all the necessary libraries

```
In [1]:  ▶  import pandas as pd
             import numpy as np
             import matplotlib.pyplot as plt
             import seaborn as sns
```

- Loading Dataset

```
In [2]:  ▶  data= pd.read_csv('store_data.csv')
             data.head()
```

Out[2]:

	shrimp	almonds	avocado	vegetables mix	green grapes	whole weat flour	yams	cottage cheese	energy drink	tomato juice	low fat yogurt	green tea	honey	salad	mineral water	salmon	antioxydant juice	sn
0	burgers	meatballs	eggs	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
1	chutney	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
2	turkey	avocado	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
3	mineral water	milk	energy bar	whole wheat rice	green tea	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
4	low fat yogurt	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	

- Some Info About Data

```
In [4]:  ▶  data.shape
```

Out[4]: (7500, 20)

```
In [5]:  ▶  data.isna().sum()
```

```
Out[5]:  shrimp                 0
         almonds             1754
         avocado             3112
         vegetables mix      4156
         green grapes        4972
         whole weat flour    5637
         yams                6132
         cottage cheese      6520
         energy drink        6847
         tomato juice        7106
         low fat yogurt      7245
         green tea           7347
         honey               7414
         salad               7454
         mineral water       7476
         salmon              7493
         antioxydant juice   7497
         frozen smoothie     7497
         spinach             7498
         olive oil           7500
         dtype: int64
```

- Pre-Processing Data

```
transactions = []
for i in range(0, data.shape[0]):
    transactions.append([str(data.values[i, j]) for j in range(0, 20)])

print(transactions[0])
```

- Applying Model

```python
from apyori import apriori

rules = apriori(transactions, min_support = 0.003, min_confidence = 0.2, min_lift = 3, min_length = 2)
# Support: number of transactions containing set of times / total number of transactions
# .        --> products that are bought at least 3 times a day --> 21 / 7501 = 0.0027
# Confidence: Should not be too high, as then this wil lead to obvious rules

#Try many combinations of values to experiment with the model.

#viewing the rules
results = list(rules)
```

- Final Output

```
In [10]:  ▶| results['ordered_statistics'][0]

Out[10]: [OrderedStatistic(items_base=frozenset({'light cream'}), items_add=frozenset({'chicken'}), confidence=0.2905982905982906, li
          ft=4.843304843304844)]
```