IBM PHASE-1 PROJECT SUBMISSION

NAME	: Sameera Banu M
COLLEGE CODE	: 4225
COLLEGE NAME	:University College of Engineering Villupuram
DOMAIN	: Artificial Intelligence (AI)
TITLE	: Market Basket Insights

Problem Definition:

The problem is to perform market basket analysis on a provided dataset to unveil hidden patterns and associations between products. The goal is to understand customer purchasing behaviour and identify potential cross-selling opportunities for a retail business. This project involves using association analysis techniques, such as Apriori algorithm, to find frequently co-occurring products and generate insights for business optimization.

Design Thinking:

Data Source:

Choose a dataset containing transaction data, including lists of purchased products.

Dataset Link: https://www.kaggle.com/datasets/aslanahmedov/market-basket-analysis

Dataset Description:

- File name: Assignment-1_Data

- List name: retaildata

- File format: . xlsx

- Number of Row: 522065

- Number of Attributes: 7

- Bill No: 6-digit number assigned to each transaction. Nominal.
- Item name: Product name. Nominal.
- Quantity: The quantities of each product per transaction. Numeric.

  Date: The day and time when each transaction was generated.

  Numeric Price: Product price. Numeric.

  Customer ID: 5-digit number assigned to each customer. Nominal.

  Country: Name of the country where each customer resides. Nominal.

	A	B	C	D	E	F	G
1	BillNo	Itemname	Quantity	Date	Price	CustomerID	Country
2	536365	WHITE HANGING HEART T-LIGHT HOLDER	6	01.12.2010 08:26	2,55	17850	United Kingdom
3	536365	WHITE METAL LANTERN	6	01.12.2010 08:26	3,39	17850	United Kingdom
4	536365	CREAM CUPID HEARTS COAT HANGER	8	01.12.2010 08:26	2,75	17850	United Kingdom
5	536365	KNITTED UNION FLAG HOT WATER BOTTLE	6	01.12.2010 08:26	3,39	17850	United Kingdom
6	536365	RED WOOLLY HOTTIE WHITE HEART.	6	01.12.2010 08:26	3,39	17850	United Kingdom

Data Preprocessing:

Prepare the transaction data by transforming it into a suitable format for association analysis. It involves cleaning, transforming, and organizing raw data into a suitable format for analysis. It includes tasks like handling missing values, encoding categorical variables, scaling numerical features, and splitting data into training and testing sets for machine learning.

Association Analysis:

Utilize the Apriori algorithm to identify frequent itemsets and generate association rules.

Apriori is a popular algorithm in data mining and association analysis. It's used to find frequent itemsets and discover association rules in a dataset. The algorithm works by iteratively generating candidate itemsets and pruning those that don't meet a minimum support threshold. This helps identify items that often appear together in transactions, which can be valuable for various applications like market basket analysis in retail.

Insights Generation:

Interpret the association rules to understand customer behavior and cross-selling opportunities.

It involves extracting valuable and meaningful information from data to inform decision-making or provide a deeper understanding of a topic. It often includes data analysis, pattern recognition, and drawing conclusions from available information.

Visualization:

Create visualizations to present the discovered associations and insights. It is the process of representing data or information visually using charts, graphs, or other graphical elements to make it easier to understand and analyse.

Business Recommendations:

Provide actionable recommendations for the retail business based on the insights. A business recommendation is a concise suggestion or advice provided to a company or individual on how to improve their operations, make strategic decisions, or address specific challenges to achieve their goals.

Sample Program:

```python
# Import the necessary libraries
from mlxtend.frequent_patterns import apriori
from mlxtend.frequent_patterns import association_rules

# Create a sample transaction dataset
dataset = [['milk', 'bread', 'nuts'],
        ['milk', 'bread'],
        ['milk', 'nuts'],
        ['bread', 'nuts'],
        ['milk', 'bread', 'nuts', 'apples']]

# Convert the dataset into a one-hot encoded format
from mlxtend.preprocessing import TransactionEncoder
te = TransactionEncoder()
te_ary = te.fit(dataset).transform(dataset)
df = pd.DataFrame(te_ary, columns=te.columns_)

# Generate frequent itemsets using Apriori
frequent_itemsets = apriori(df, min_support=0.5, use_colnames=True)

# Generate association rules
association_rules = association_rules(frequent_itemsets, metric="lift", min_threshold=1.0)
```

```python
# Print the frequent itemsets and association rules
print("Frequent Itemsets:\n", frequent_itemsets)
print("\nAssociation Rules:\n", association_rules)
```

Output:

Frequent Itemsets:
```
   support      itemsets
0    0.8        (bread)
1    0.8         (milk)
2    0.8         (nuts)
3    0.6  (bread, milk)
4    0.6  (bread, nuts)
5    0.6   (nuts, milk)
```

Association Rules:
 Empty DataFrame
Columns: [antecedents, consequents, antecedent support, consequent support, support, confidence, lift, leverage, conviction, zhangs_metric]
Index: []