**Experiment 8: Association Rule Analysis using Python**

**Objective** :To implement association rule analysis using Python.

**Time Required** : 3 hrs

**Programming Language** : Python

**Software Required** : Anaconda

**Introduction**

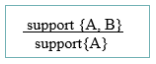
Association Rule Analysis finds interesting associations and relationships among large sets of data items. This rule shows how frequently an itemset occurs in a transaction. A typical example is Market Based Analysis. Market Based Analysis is one of the key techniques used by large relations to show associations between items. It allows retailers to identify relationships between the items that people buy together frequently.

**Apriori Algorithm:**

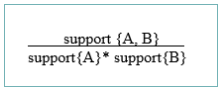
The Apriori algorithm is used for finding frequent item-sets in a dataset for Boolean association rule. Name of the algorithm is Apriori because it uses prior knowledge of frequent itemset properties. We apply an iterative approach or level-wise search where k-frequent item-sets are used to find k+1 item-sets. To improve the efficiency of level-wise generation of frequent item-sets, an important property is used called Apriori property which helps by reducing the search space. Walmart especially has made great use of the algorithm in suggesting products to its users.

The output of the apriori algorithm is the generation of association rules. This can be done by using some measures called support, confidence, and lift. Now let’s understand each term.

**Support:** It is calculated by dividing the number of transactions having the item by the total number of transactions.

**Confidence:** It is the measure of trustworthiness and can be calculated using the below formula.

Conf(A => B)=

**Lift:** It is the probability of purchasing B when A is sold. It can be calculated by using the below formula.

Lift(A => B)=

**Lift(A => B) =1 :** There is no relation between A and B.  
**Lift(A => B)> 1:** There is a positive relation between the item set . It means, when product A is bought, it is more likely that B is also bought.  
**Lift(A => B)< 1:** There is a negative relation between the items. It means, if product A is bought, it is less likely that B is also bought.

**Dataset**

**Load the data using the following link** <https://www.kaggle.com/datasets/mrmining/online-retail>

**Lab Tasks**

Write a Python program that accomplishes the following:

1. Load the transaction data from the 'Online Retail.xlsx' file into a panda DataFrame.
2. Preprocess the data by removing extra spaces in the 'Description' column, dropping rows without invoice numbers, and filtering out credit transactions.
3. Create separate transaction baskets for each country of interest (France, United Kingdom, Portugal, and Sweden) by grouping the data based on 'Country', 'InvoiceNo', and 'Description' columns. Calculate the sum of 'Quantity' for each unique combination of 'InvoiceNo' and 'Description'. Reshape the resulting DataFrame to have 'InvoiceNo' as the index and each unique 'Description' as a column, representing the quantity of the corresponding item in the transaction.
4. Apply the Apriori algorithm using the 'apriori()' function from the mlxtend library to find frequent itemsets that include the **'Cutlery Set'** in each country. Set the minimum support threshold to 0.05.
5. Generate association rules from the frequent itemsets using the 'association\_rules()' function, considering a minimum lift threshold of 1.
6. Sort the association rules based on confidence and lift values in descending order.
7. Extract and analyze the top association rules that involve the 'Cutlery Set' for each country.
8. Interpret the rules to identify patterns of the 'Cutlery Set' being purchased with other items in different countries. Look for high-confidence rules with significant lift values, which indicate strong associations between the 'Cutlery Set' and other items.
9. Print the top association rules for each country, including the antecedent (items commonly purchased before the 'Cutlery Set') and consequent (items commonly purchased after the 'Cutlery Set') of each rule, along with their confidence and lift values.
10. Provide insightful interpretations of the association rule patterns in each country, highlighting any interesting and meaningful findings related to the 'Cutlery Set' item.