Association Rules

The Objective of this assignment is to introduce students to rule mining techniques, particularly focusing on market basket analysis and provide hands on experience.

**Dataset:**

Use the Online retail dataset to apply the association rules.

**Data Preprocessing:**

Pre-process the dataset to ensure it is suitable for Association rules, this may include handling missing values, removing duplicates, and converting the data to appropriate format.

**Association Rule Mining:**

* Implement an Apriori algorithm using tool like python with libraries such as Pandas and Mlxtend etc.
* Apply association rule mining techniques to the pre-processed dataset to discover interesting relationships between products purchased together.
* Set appropriate threshold for support, confidence and lift to extract meaning full rules.

**Analysis and Interpretation:**

* Analyse the generated rules to identify interesting patterns and relationships between the products.
* Interpret the results and provide insights into customer purchasing behaviour based on the discovered rules.

# Interview Questions:

1. **What is lift and why is it important in Association rules?**

Ans: Lift is used to compare the strength of the association between two items to the expected strength of the association if the items were independent.

Formula for Lift = No of times X and Y occurred / Total occurance of X \* Total Occurance of Y

1. **What is support and Confidence. How do you calculate them?**

Ans: Support :

Support is a measure of how frequently an item or itemset appears in the dataset

Formula for Support = No of occurance of X and Y/ Total No of Items

Confidence:

Confidence is a measure of the strength of the association between two items. It is calucaleted with no of items both occurred divided by total no of 1st item occured

Formula for Confidence = No of occurance of X and Y / Total occurance of X

1. **What are some limitations or challenges of Association rules mining?**

Ans: The primary disadvantages of association rule algorithms are obtaining boring rules, having a large number of discovered rules, and low algorithm performance