Report: Market Basket Analysis Using the Apriori Algorithm

Objective:

The goal of this analysis is to identify relationships between products in a supermarket transaction dataset. This is achieved using the Apriori algorithm to uncover association rules, helping to understand purchasing patterns and provide actionable recommendations to enhance marketing strategies.

Steps:

1. Loading Libraries and Preparing Data

The analysis begins by loading necessary libraries, such as pandas, and reading the dataset from a CSV file containing store transactions. This enables data organization and preparation for further analysis.

2. Viewing and Inspecting Data

A subset of the data is displayed to understand the overall structure and identify main columns, such as products, store types, and cities. This step helps to understand the data layout before conducting deeper analysis.

3. Analyzing Data Distribution by Cities and Store Types

The distribution of transactions by city (City) and store type (Store_Type) is analyzed to understand how transactions vary across these attributes. This can be valuable for understanding purchasing patterns by region or store type, potentially influencing the analysis results.

4. Filtering and Preparing Data for Analysis

Only relevant columns, such as transaction ID (Transaction_ID) and product (Product), are selected. This step focuses on the actual transactions and removes unnecessary columns, simplifying the application of the algorithm.

5. Encoding Transaction Data

The data is converted into a binary format to prepare for the Apriori algorithm. This involves encoding values as 1s and 0s to indicate the presence or absence of a particular product in each transaction, which facilitates identifying common patterns.

6. Applying the Apriori Algorithm to Extract Association Rules

The mlxtend library is used to apply the Apriori algorithm to identify frequent itemsets (groups of products that frequently occur together in transactions). First, frequent itemsets

are identified to discover products that appear together consistently. Then, association rules are extracted based on the lift metric, which measures the strength of the relationship between products. This reveals whether products are likely to be purchased together more than expected.

7. Analyzing and Displaying Results by Lift and Support

The extracted rules are ranked by their lift and support values, displaying the top 10 rules based on these metrics. High lift values indicate a strong positive association between products, helping to identify items that are frequently bought together. Support values help measure the popularity of the rule.

Summary:

This notebook provides a comprehensive approach to market basket analysis using the Apriori algorithm. It covers data preparation, encoding, extracting association rules, and identifying key rules using lift and support metrics. Retailers can leverage these insights to pinpoint related products and enhance marketing strategies, such as offering bundled promotions or arranging related products near each other in stores, ultimately driving sales and improving the customer experience.