

MARKET BASKET ANALYSIS

Under : Kumar Sir
By: Revanth Kumar

Problem Statement:

Market Basket Analysis is a technique used by retailers and marketers to understand the purchasing behavior of their customers by analyzing the items that are frequently purchased together. The goal of this analysis is to identify patterns and relationships between products, which can be used to improve sales, cross-sell, and optimize inventory management.

The problem statement for Market Basket Analysis can be defined as follows: Given a dataset of transactions containing items purchased by customers, the objective is to analyze the data to identify the association between different items and to create rules that can be used to make recommendations to customers, optimize product placement, and improve sales. The analysis involves identifying frequent itemsets, association rules, and support, confidence, and lift measures to identify the relationships between different items. The analysis output is a set of rules that can be used to make recommendations to customers, optimize product placement, and improve sales.

Dataset Description:

The Dataset is acquired from Kaggle named "Groceries" which is a CSV file. The dataset contains 9835 transactions by customers shopping for groceries. The data contains 169 unique items.

Potential Use Cases:

The dataset can be used for market basket analysis, customer segmentation, and product recommendation systems. It can also be used for inventory management, as it provides information on the quantity of each product sold and the customer who purchased it. The dataset can also identify popular products and customer segments to optimize marketing campaigns and improve sales.

Methodology :

The "Apriori" algorithm is used for this project. Invented by R. Agarwal and R. Srikant. The Apriori algorithm is a popular method for market basket analysis and association rule mining. It is a rule-based machine learning technique that analyzes a large dataset of transactions to identify patterns and relationships between products that are frequently purchased together. Here are some key points to know about the Apriori algorithm:

1. Support, Confidence, and Lift: The Apriori algorithm uses three metrics to measure the strength of association between products: support, confidence, and lift. Support measures the frequency of a product or product set in the dataset. Confidence measures the conditional probability of purchasing product B given the purchase of product A. Lift measures the degree of association between two products and indicates whether their purchase is independent or correlated.

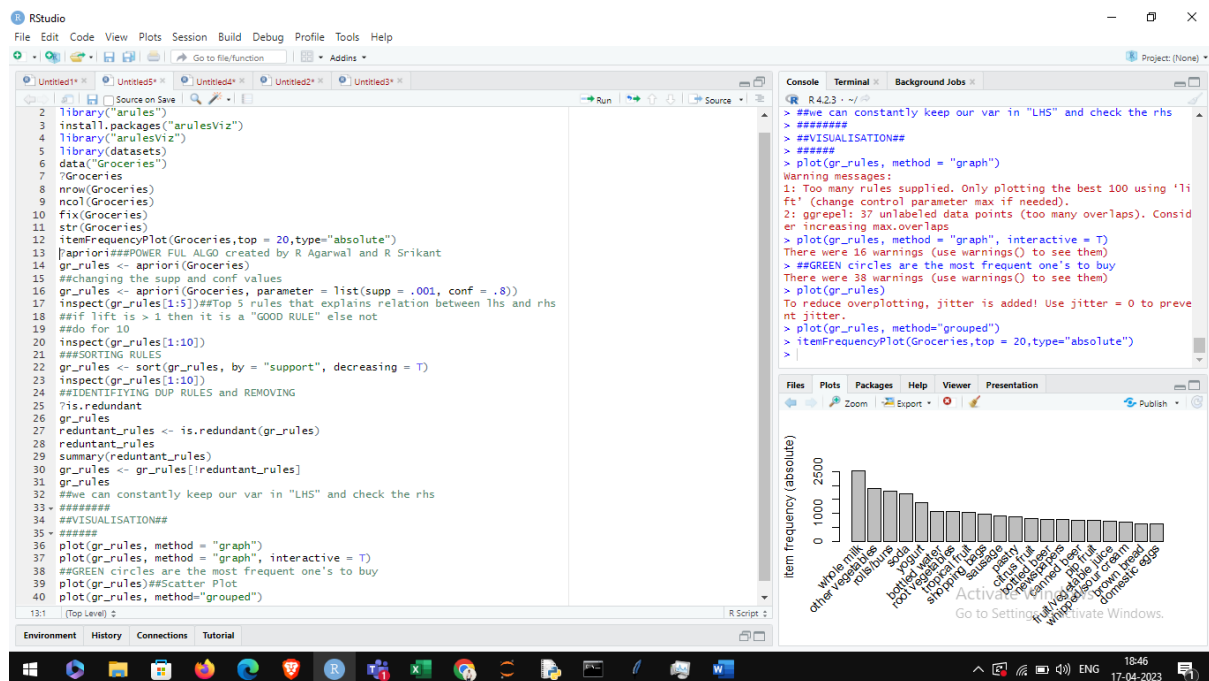
MARKET BASKET ANALYSIS

Under : Kumar Sir
By: Revanth Kumar

2. Frequent Itemsets: The Apriori algorithm uses the concept of frequent itemsets to identify sets of products that frequently occur together in transactions. The algorithm scans the dataset multiple times to identify all the frequent itemsets based on the support threshold provided by the user.
3. Apriori Principle: The Apriori algorithm uses the Apriori principle to reduce the search space of frequent item sets. The principle states that if an item set is frequent, then all its subsets must also be frequent. This reduces the number of candidate itemsets that need to be checked, which speeds up the algorithm.
4. Association Rules: The Apriori algorithm uses frequent item sets to generate association rules. An association rule is a relationship between two sets of products that satisfy a minimum support and confidence threshold. The rules are represented as "If X, then Y," where X and Y are sets of products.
5. Applications: The Apriori algorithm is widely used in market basket analysis, product recommendation systems, and customer segmentation. It can help businesses to identify customer preferences, optimize product placement, and improve sales.

The Apriori algorithm is a powerful tool for analyzing transaction data and identifying patterns and relationships between products. It is relatively easy to implement and can generate useful insights into customer behavior and preferences.

CODE USED :

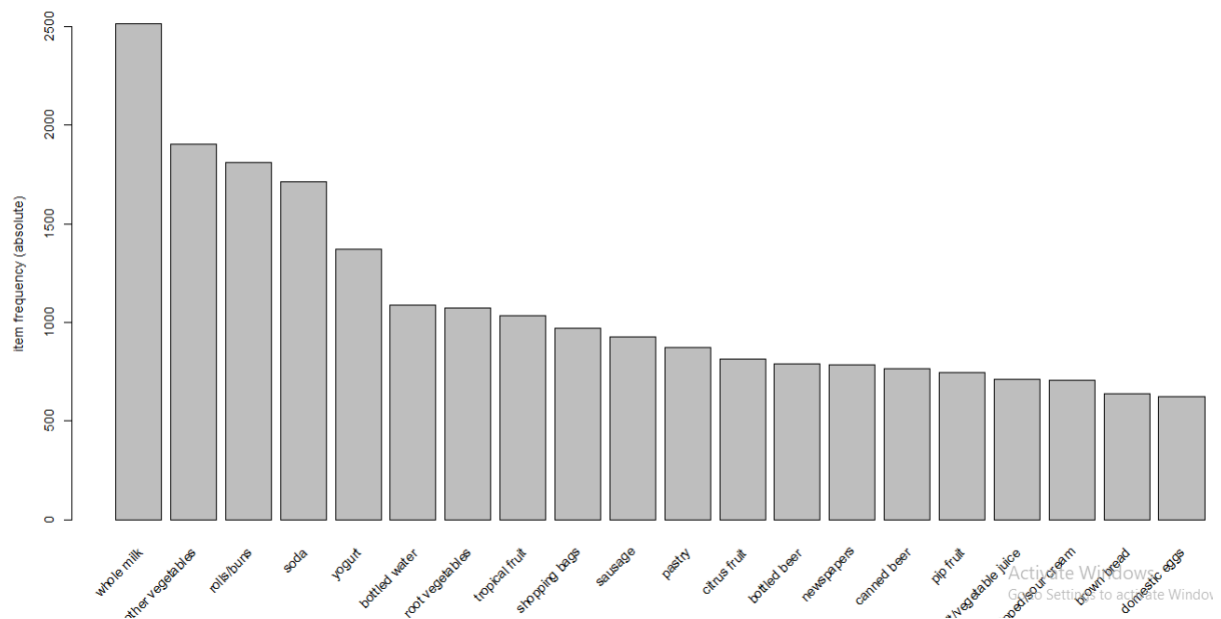


MARKET BASKET ANALYSIS

Under : Kumar Sir
By: Revanth Kumar

VISUALISATION :

- Top 20 products that are Frequently bought by the Customers



- Output for Apriori Algorithm for 5 products :

```
> inspect(gr_rules[1:5])##Top 5 rules that explains relation between lhs and rhs
```

	lhs	rhs	support	confidence	coverage	lift	count
[1]	{citrus fruit, tropical fruit, root vegetables, whole milk}	=> {other vegetables}	0.003152008	0.8857143	0.003558719	4.577509	31
[2]	{other vegetables, curd, domestic eggs}	=> {whole milk}	0.002846975	0.8235294	0.003457041	3.223005	28
[3]	{hamburger meat, curd}	=> {whole milk}	0.002541942	0.8064516	0.003152008	3.156169	25
[4]	{herbs, rolls/buns}	=> {whole milk}	0.002440264	0.8000000	0.003050330	3.130919	24
[5]	{tropical fruit, herbs}	=> {whole milk}	0.002338587	0.8214286	0.002846975	3.214783	23

Under : Kumar Sir
By: Revanth Kumar

-
- Items in LHS Groups
- 1 rules {lemon, red, blue, wine}
- 5 rules {oil, soda, +8 items}
- 2 rules {ham, grapes, +5 items}
- 3 rules {sliced cheese, white bread, +7 items}
- 11 rules {margarine, pastry, +15 items}
- 2 rules {white bread, butter, +1 items}
- 8 rules {ham, newspapers, +14 items}
- 2 rules {cream cheese, butter, +1 items}
- 8 rules {soft cheese, soda, +13 items}
- 17 rules {dessert, hard cheese, +20 items}
- 7 rules {ham, frozen vegetables, +10 items}
- 28 rules {meat, newspapers, +30 items}
- 44 rules {turkey, grapes, +40 items}
- 14 rules {frozen meals, frankfurter, +11 items}
- 18 rules {canned fish, beer, +22 items}
- 14 rules {soft cheese, beef, +12 items}
- 61 rules {soup, salad spreads, +38 items}
- 58 rules {mustard, baking powder, +39 items}
- 47 rules {jam, detergent, +39 items}
- 42 rules {hamburger meat, specialty cheese, +37 items}
- (bottled beer)
- (root vegetables)
- (tropical fruit)
- (yogurt)
- (other vegetables)
- (whole milk)
- lift
- 5 7 9 11
- support
- 0.00125 0.00150 0.00175
- Activate windows
Go to Settings to activate Windows.

- [illegible]

MARKET BASKET ANALYSIS

Under : Kumar Sir

By: Revanth Kumar

Conclusion:

In conclusion, the Market Basket Analysis project using the Apriori algorithm can provide valuable insights into customer behavior and preferences, which can help businesses to optimize product placement, increase sales, and improve customer satisfaction. Effective data preprocessing is critical for the accuracy and reliability of the analysis.

By applying the Apriori algorithm to transaction data and generating association rules, businesses can better understand which products are often purchased together, allowing them to make more informed decisions about marketing, product placement, and inventory management.