**ASSIGNMENT X**

**Project Title:**

Market Basket Analysis of Retail Market Places

**Team Leader:**

- Ejumalla Saikiran

**Team members:**

1. J. Sinduja

2. A. Bhavya Sri

3. S. Lakshanya

4. Naveen Rampa

**Submission Date:**

**07/04/2024**

**Market Basket Analysis Report**

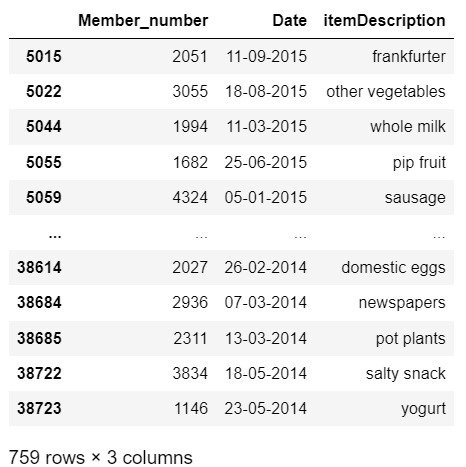
**1. Dataset**

The dataset comprises transactional records from a grocery store, containing information about purchases made by customers over a period of time. Each transaction is associated with a unique member number and includes details such as the date of purchase and the item description.

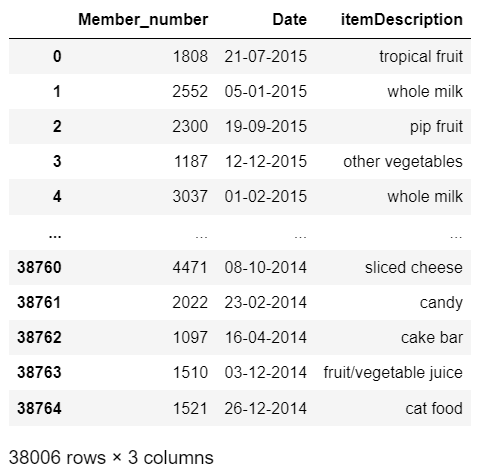


1. **Data Pre-Processing**

During the data preprocessing stage, we conducted an initial examination of the dataset to ensure data quality and consistency. While investigating the dataset, we did not identify any missing values, indicating that the dataset was complete in terms of information availability. However, we did observe the presence of duplicate entries within the dataset.

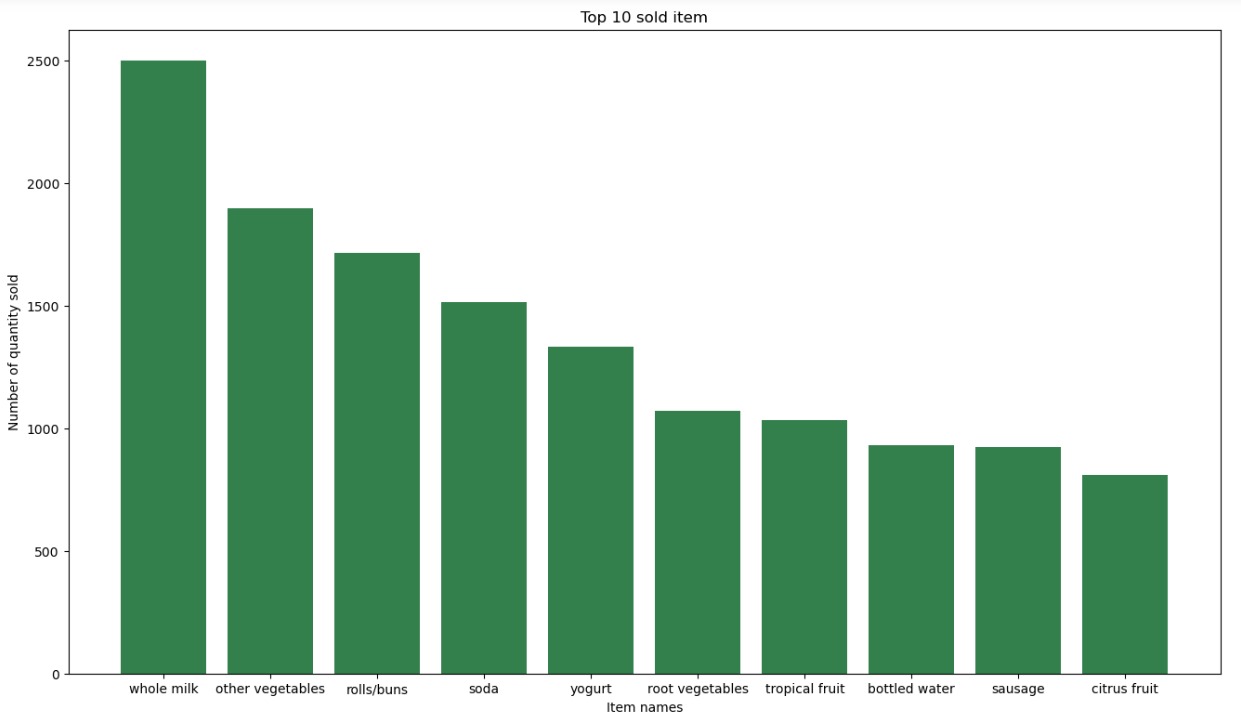


To address this issue, we performed duplicate detection and elimination procedures and acquired the following clean data set.

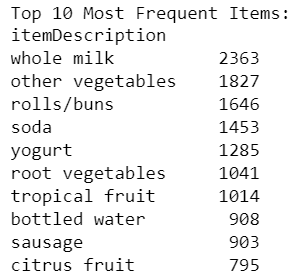


1. **Exploratory Data Analysis (EDA)**
   1. **Top 10 Most Frequently Purchased Products**

We utilized a bar graph visualization technique to identify the top 10 most frequently purchased products in the dataset**.**

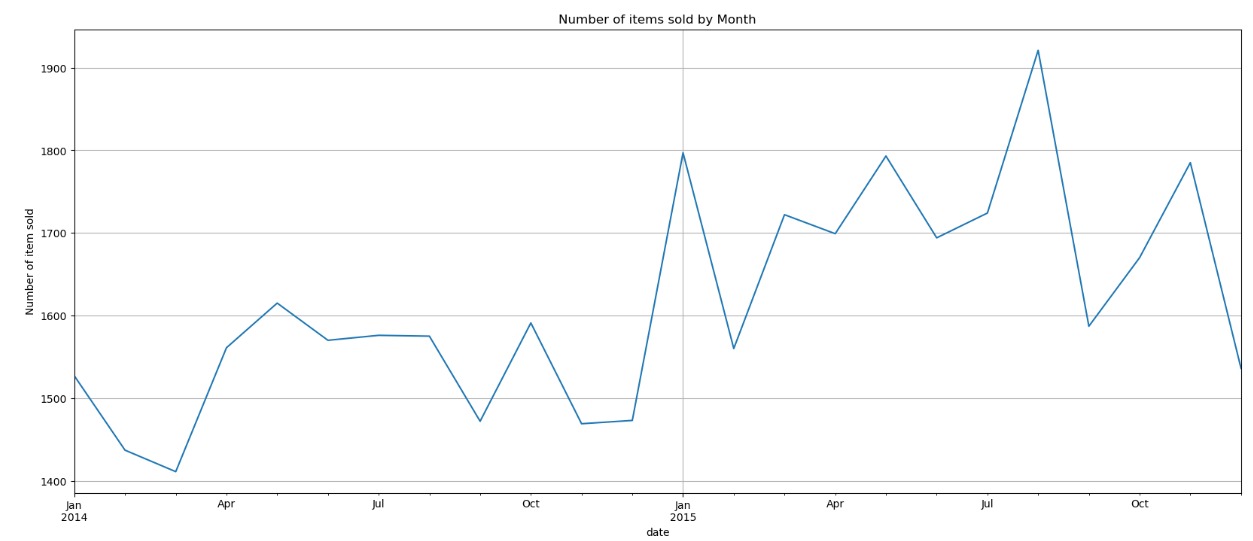


The following are the numerical purchase value of each item :



**3.2 Analysis of Monthly Sales Trends**

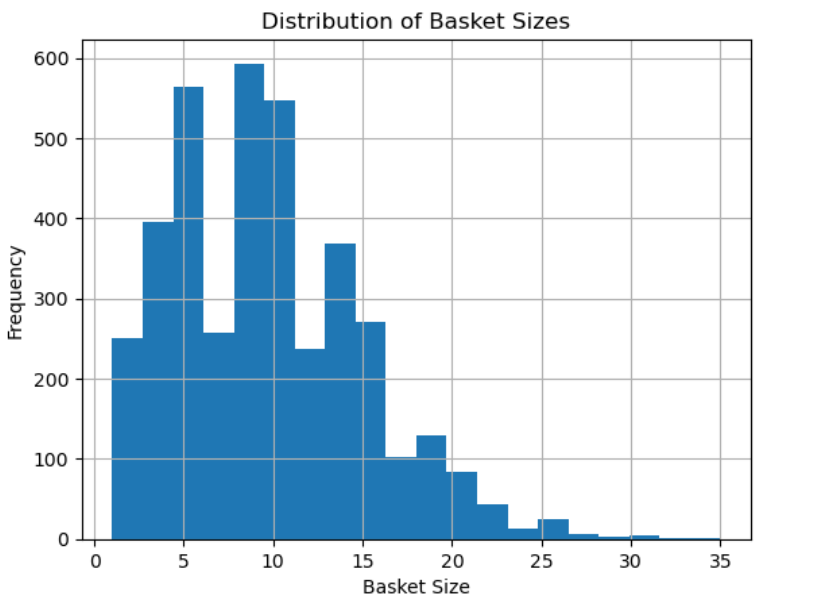
We conducted an analysis of monthly sales trends to understand the variations in the number of items sold over time.



The graph revealed that August 2015 experienced the highest peak in sales, indicating a surge in purchasing activity during that period. Following closely behind, January 2015 exhibited another significant peak, suggesting heightened sales during the beginning of the year.Conversely, March 2014 recorded the lowest drop in sales volume, indicating a decrease in purchasing activity during that month compared to others.

**3.3 Basket Analysis**

Analyze the distribution of basket sizes (number of items purchased per transaction) to understand the typical purchasing behavior of customers.

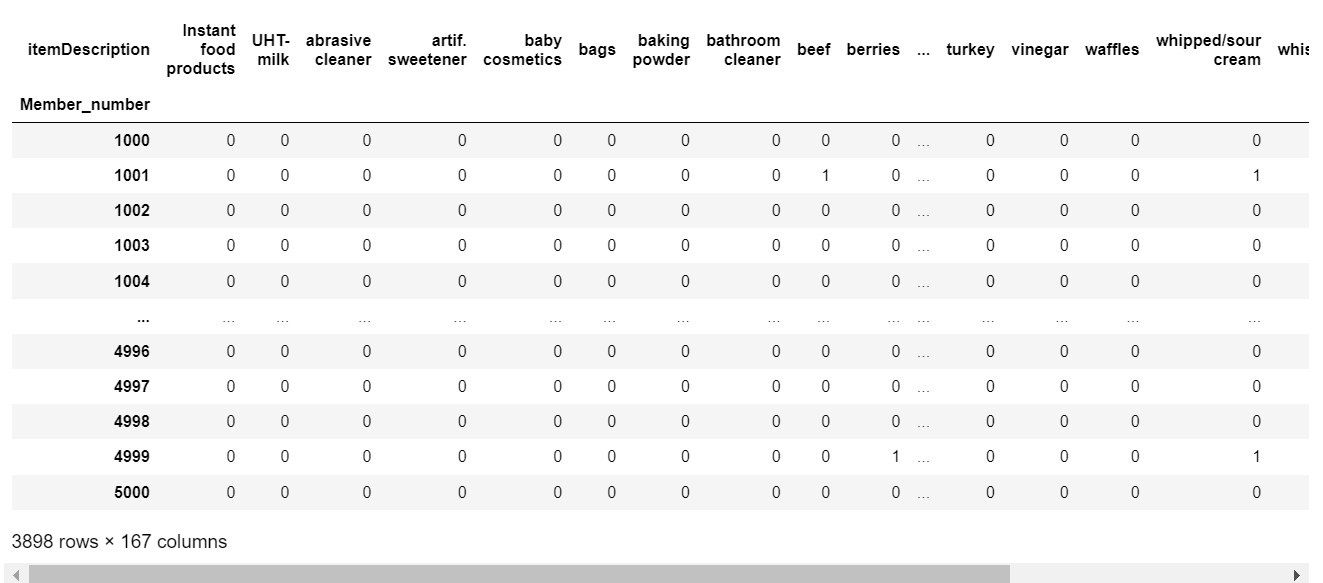


The histogram shows that most transactions have a small number of items. The most frequent basket size is 5 to 10, which means that many customers only buy 5 to 10 items at a time. The frequency of transactions then decreases as the basket size increases. There are some transactions with a large number of items, but these are less frequent.

1. **Generation of Association Rules**

The following steps demonstrate the process of generating association rules from transactional data using the Apriori algorithm and the MLxtend library in Python :

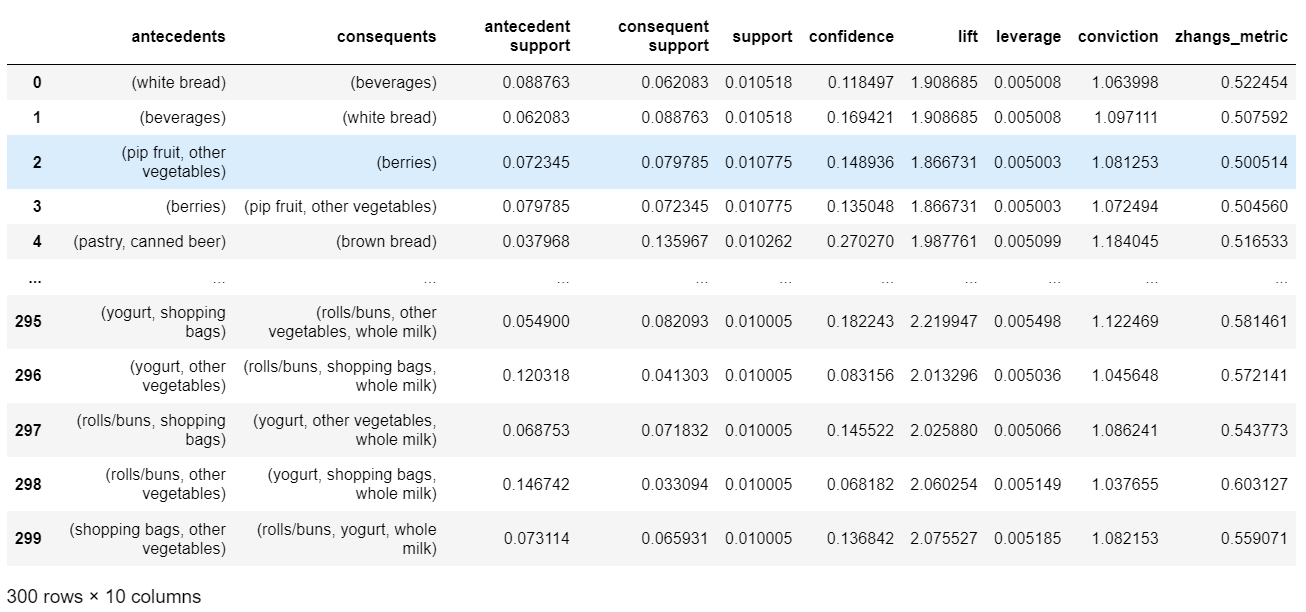
* **One-Hot Encoding:** The transactional data is transformed into a binary matrix format known as one-hot encoding. Each row represents a unique member number, and each column represents an item. The values in the matrix indicate whether a particular item was purchased by the corresponding member.



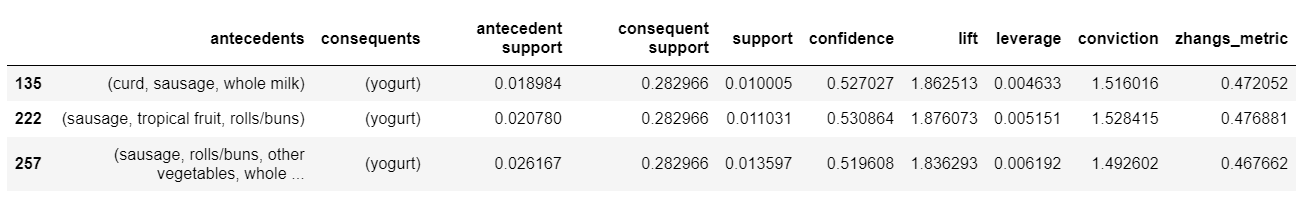
* **Frequent Itemsets Generation:** The Apriori algorithm is applied to the one-hot encoded dataset to identify frequent itemsets. These itemsets represent combinations of items that occur together frequently in transactions, surpassing a specified minimum support threshold.



* **Association Rules Generation:** Using the frequent itemsets obtained from the previous step, association rules are generated. These rules capture relationships between items, indicating which combinations of items are likely to be purchased together. The rules are evaluated based on metrics such as support, confidence, and lift.



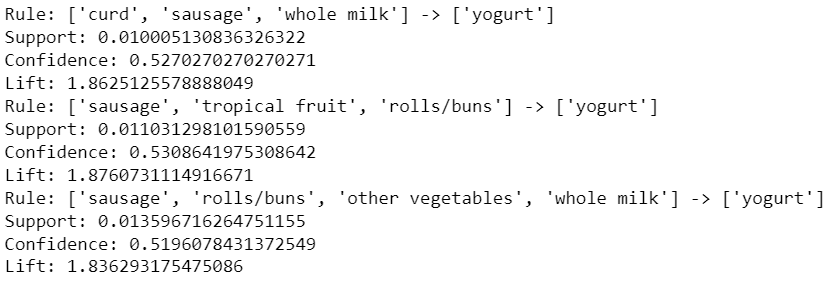
* **Filtering Rules:** The rules are filtered to include only those with a minimum lift of 1.5 and a minimum confidence of 0.5. These thresholds ensure that only rules with a significant level of association and confidence are retained.



* **Interpretation and Analysis:** Finally, the filtered association rules are interpreted and analyzed to derive insights into customer purchasing behavior. Each rule is examined to understand the antecedents (items in the basket), consequents (items likely to be purchased), support (frequency of occurrence), confidence (probability of purchasing consequents given antecedents), and lift (measure of association between antecedents and consequents).

1. **Observations for Different min\_support and min\_threshold Values**
   1. **min\_support=0.01, min\_threshold=1.8 :**

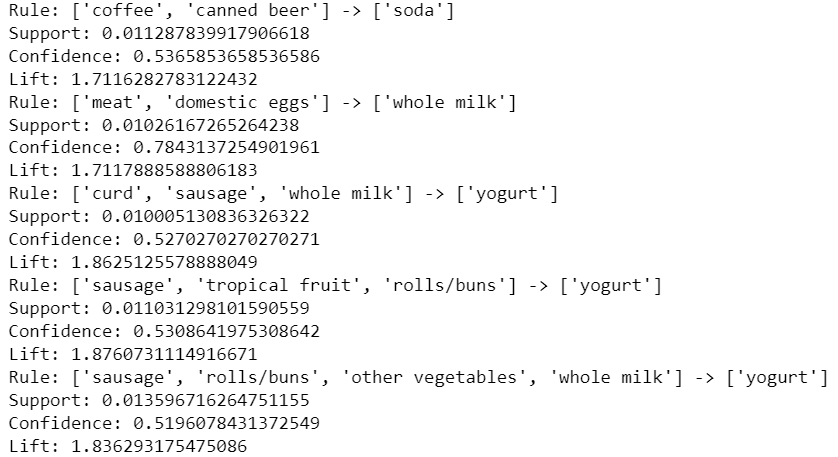
The association rules generated with a minimum support of 0.01 and a minimum lift threshold of 1.8 reveal several interesting patterns.



Observation : Notable rules include combinations such as ['curd', 'sausage', 'whole milk'] -> ['yogurt'] with a confidence of 0.527 and lift of 1.863.This combination suggests that customers who purchase curd, sausage, and whole milk are also likely to buy yogurt.

* 1. **min\_support=0.01, min\_threshold=1.7 :**

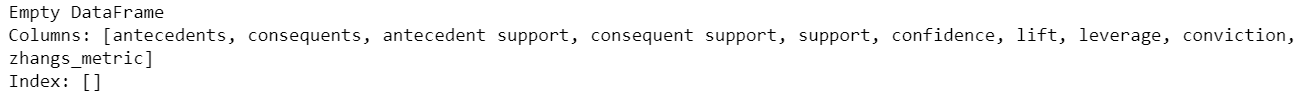
With a minimum support of 0.01 and a minimum lift threshold of 1.7, the association rules generated provide insights into purchasing patterns among different items.



Observations : Rule ['coffee', 'canned beer'] -> ['soda'] suggests that customers who buy coffee and canned beer are also likely to purchase soda, with a moderate level of confidence. Rule ['meat', 'domestic eggs'] -> ['whole milk'] suggests that customers purchasing meat and domestic eggs are highly likely to buy whole milk, as indicated by the high confidence level.

**5.3 min\_support=0.02, min\_threshold=1.6:**

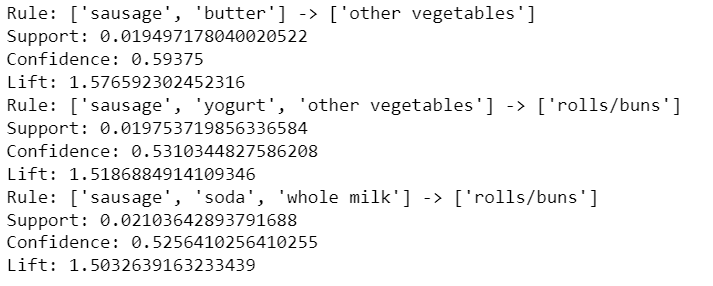
Again, no association rules are generated when both the support threshold is set to 0.02 and the lift threshold to 1.6.



This indicates that the higher support requirement further reduces the number of association rules discovered.

**5.4 min\_support=0.019, min\_threshold=1.5 :**

With a slightly lower minimum support threshold of 0.019, additional association rules are generated.



Observations : Notable rules include ['sausage', 'butter'] -> ['other vegetables'] with a confidence of 0.594 and lift of 1.577, suggesting a link between purchasing sausage and butter with other vegetables. Another rule, ['sausage', 'yogurt', 'other vegetables'] -> ['rolls/buns'], indicates a relationship between purchasing sausage, yogurt, other vegetables, and rolls/buns.

1. **Conclusion**

In our analysis of association rules for the five most frequently purchased products, we identified several meaningful patterns indicating potential relationships between items. These rules provide valuable insights into customer behavior and can inform strategic decisions for product placement, promotions, and cross-selling strategies. Here are the key findings:

1. Rule: ['meat', 'domestic eggs'] -> ['whole milk']

* Customers who purchase meat and domestic eggs are highly likely to buy whole milk, with a confidence of 78.4%.

1. Rule: ['sausage', 'butter'] -> ['other vegetables']

* Customers buying sausage and butter are also likely to purchase other vegetables, with a confidence of 59.4%.

3. Rule: ['sausage', 'yogurt', 'other vegetables'] -> ['rolls/buns']

* Customers purchasing sausage, yogurt, and other vegetables have a moderate probability of also buying rolls/buns.

4. Rule: ['coffee', 'canned beer'] -> ['soda']

* Customers who buy coffee and canned beer are likely to purchase soda as well.

1. Rule: ['curd', 'sausage', 'whole milk'] -> ['yogurt']

* This rule suggests that customers who buy curd, sausage, and whole milk are moderately likely to also purchase yogurt with the confidence level of 52.7% .

1. Rule: ['sausage', 'tropical fruit', 'rolls/buns'] -> ['yogurt']

* Customers purchasing sausage, tropical fruit, and rolls/buns have a moderate probability of also buying yogurt with the confidence level of 53.1% .

1. Rule: ['sausage', 'rolls/buns', 'other vegetables', 'whole milk'] -> ['yogurt']

* This rule indicates that customers who buy sausage, rolls/buns, other vegetables, and whole milk are likely to purchase yogurt as well with the confidence level of 52.0%.

These association rules highlight potential cross-selling opportunities and can guide retailers in optimizing their product offerings and marketing strategies to enhance customer satisfaction and drive sales.