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?
2. What is support and Confidence. How do you calculate them?
3. What are some limitations or challenges of Association rules mining?

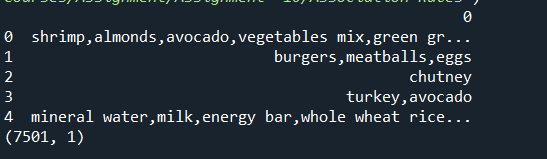
Observations:

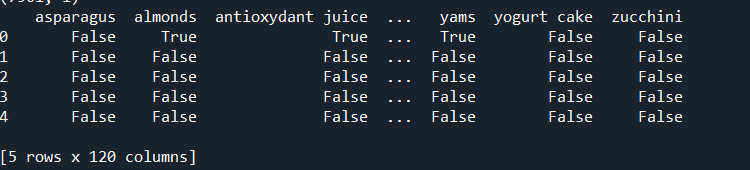
**Dataset:**

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.



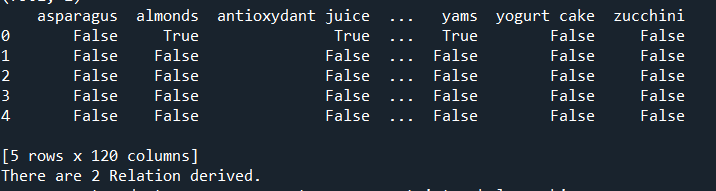


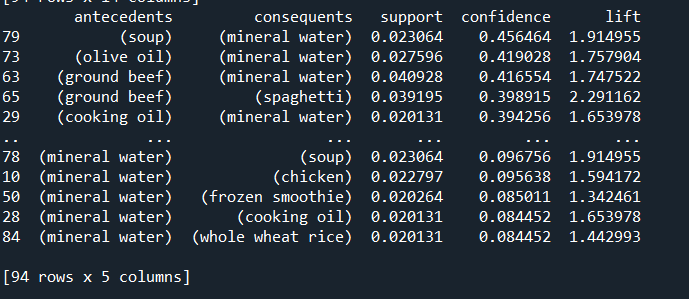
**Association Rule Mining:**

* 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.

Apriori requires data in a format where each column represents an item, and each row represents a transaction. Set thresholds for **support, confidence, and lift** to extract meaningful rules.

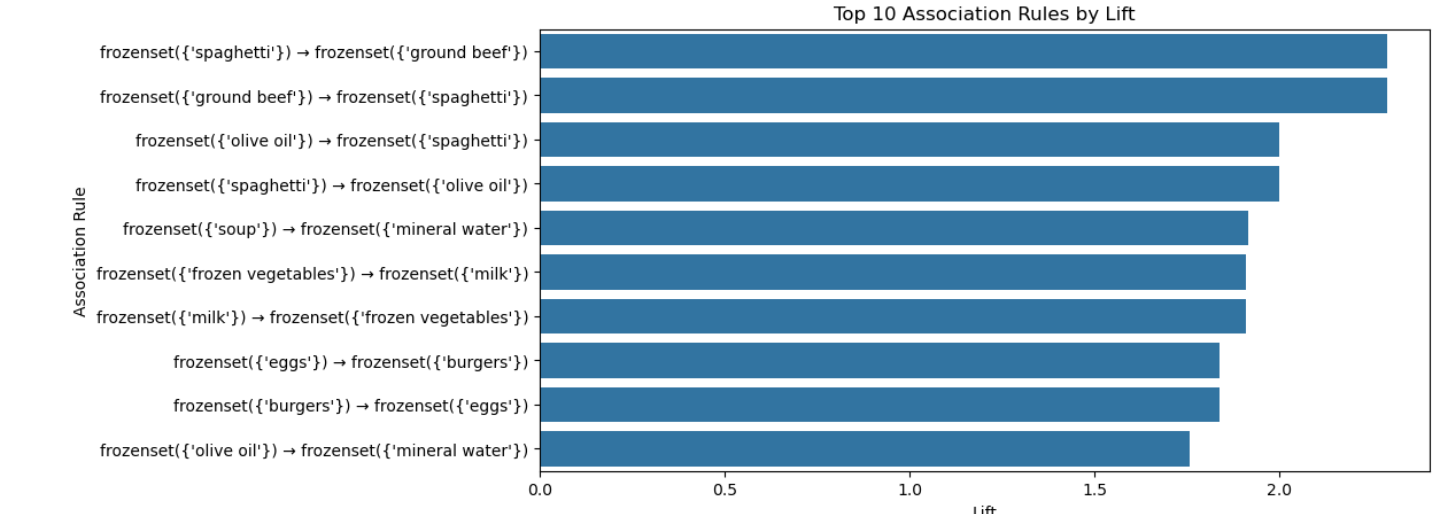
* **Support:** Set min\_support (e.g., 0.02) to filter out infrequent itemsets.
* **Confidence:** Set min\_threshold in association\_rules() to control rule strength.
* **Lift:** A **lift > 1** indicates a strong positive association.





* **Support:** The frequency of a rule in the dataset.
* **Confidence:** The probability that customers who buy the antecedent also buy the consequent.
* **Lift:** How much more likely customers are to buy the consequent when they buy the antecedent (Lift > 1 indicates a strong association).

**Business Insights & Recommendations**



**Key Observations:**

**Strongest Association**

The rule **(spaghetti) → (ground beef)** has the **highest lift**.

This means customers who buy **spaghetti** are significantly more likely to also buy **ground beef**.

This makes sense, as they are commonly used together in cooking (e.g., spaghetti Bolognese).

**High Confidence Pairings**

**(ground beef) → (spaghetti)** also has a high lift, reinforcing the two-way relationship.

**(olive oil) → (spaghetti)** is another strong rule, suggesting that people who buy olive oil frequently purchase spaghetti, possibly for making pasta dishes.

**Beverage & Dairy Trends**

**(soup) → (mineral water)** indicates customers who buy **soup** often purchase **mineral water**, possibly as a meal combination.

**(frozen vegetables) → (milk)** and **(milk) → (frozen vegetables)** suggest a correlation between dairy and frozen goods, possibly for making smoothies or cooking.

**Burger & Egg Relationship**

**(eggs) → (burgers)** and **(burgers) → (eggs)** have strong lift values.

This might indicate customers buying these items together for breakfast meals or homemade burgers.

**Business Insights & Recommendations:**

**Cross-Selling Opportunities**

Place **spaghetti and ground beef** together in stores or online to boost sales.

Offer combo discounts for **olive oil and pasta products**.

**Marketing & Promotions**

Bundle **soup with mineral water** in meal deals.

Promote **milk and frozen vegetables** together, possibly targeting health-conscious buyers.

**Inventory & Stock Planning**

Ensure sufficient stock of frequently co-purchased items, especially **spaghetti, ground beef, and olive oil**.

Interview Questions:

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

Lift is a measure of the strength of the association between two items, taking into account the frequency of both items in the dataset.

It is calculated as the confidence of the association divided by the support of the second item.

lift, can be used to compare observed confidence with expected confidence, or how many times an if-then statement is expected to be found true.

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

Support is calculated by dividing the number of transactions containing an item set by the total number of transactions. Support indicates how frequently an item appears in the data.

Confidence is calculated by dividing the number of transactions containing both itemsets by the number of transactions containing the first itemset. Confidence indicates the number of times the if-then statement is found to be true.

Calculation- confidence (X ⇒ Y) = support (X ∪ Y) / support(X).

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

**1**. Computational Complexity

**2. Meaningless or Trivial Rule**

**3.** Difficulty in Setting the Right Thresholds

**4. Handling Rare Items**

**5. Difficulty in Interpreting Rules**

**6. Handling Continuous Data**

**7. Lack of Temporal Considerations**

Despite these challenges, association rule mining remains a valuable technique for **market basket analysis, recommendation systems, and customer behavior insights**. By carefully tuning parameters, using more efficient algorithms, and incorporating domain knowledge, businesses can extract meaningful patterns.