**Phase3**

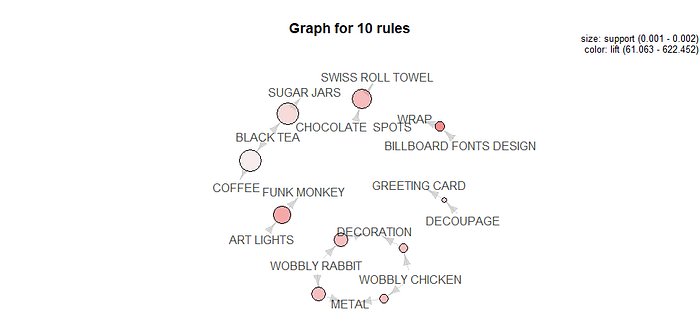
Development part-1

**Market Basket Insights**

**Introduction:**

To start a market basket insights project, you'll need to load and preprocess transaction data for association analysis. This project typically involves using techniques like Apriori or FP-growth to discover patterns and associations between items purchased together. Here's a step-by-step guide on how to load and preprocess the data:

* Market Basket Analysis is one of the key techniques used by large retailers to uncover associations between items. It works by looking for combinations of items that occur together frequently in transactions. To put it another way, it allows retailers to identify relationships between the items that people buy.
* Association Rules are widely used to analyze retail basket or transaction data, and are intended to identify strong rules discovered in transaction data using measures of interestingness, based on the concept of strong rules.



|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | |  |  |  |  | |  | |
|  |  | | | | |  | |  | |  |  |  |
|  |  | | | | |  | |  | |  |  |  |
|  |  | | | | |  | |  | |  |  |  |
| **Given** | **Data Set :** | | | | |  | |  | |  |  |  |

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | BillNo | Itemname | Quantity | Date | | Price | CustomerID | Country |  |
| 1 | 536365 | WHITE HANGING HEART T-LIGHT HOLDER | 6 | ######## | | 2.55 | 17850 | United Kingdom |
| 2 | 536365 | WHITE METAL LANTERN | 6 | ######## | | 3.39 | 17850 | United Kingdom |
| 3 | 536365 | CREAM CUPID HEARTS COAT HANGER | 8 | ######## | | 2.75 | 17850 | United Kingdom |
| 4 | 536365 | KNITTED UNION FLAG HOT WATER BOTTLE | 6 | ######## | | 3.39 | 17850 | United Kingdom |
| … | … | … | … | … | | … | … | … |
| 522061 | 581587 | PACK OF 20 SPACEBOY NAPKINS | 12 | ######## | | 0.85 | 12680 | France |
| 522062 | 581587 | CHILDREN'S APRON DOLLY GIRL | 6 | ######## | | 2.1 | 12680 | France |
| 522063 | 581587 | CHILDRENS CUTLERY DOLLY GIRL | 4 | ######## | 4.15 | | 12680 | France |
| 522064 | 581587 | CHILDRENS CUTLERY CIRCUS PARADE | 4 | ######## | | 4.15 | 12680 | France |
| 522065 | 581587 | BAKING SET 9 PIECE RETROSPOT | 3 | ######## | | 4.95 | 12680 | France |

522065 Rows x 8 columns

**Necessary Step to follow:**

**1. Data Source:**

- Obtain the transaction data. This data should contain records of items purchased in each transaction, such as a retail store's sales data or an e-commerce website's order history.

**2. Data Format:**

- Ensure that your transaction data is structured in a tabular format where each row represents a unique transaction, and each column represents an item or product purchased in that transaction.

**3. Load the Data:**

- Depending on your data format, you can load the data using Python libraries such as Pandas or any other suitable tool for your data format.

**Program:**

**import pandas as pd**

**# Load the transaction data**

**data = pd.read\_csv('transaction\_data.csv')**

**4. Data Exploration:**

- Explore the data to get a sense of its structure and contents. This step will help you understand your data better.

**Program:**

**# Display the first few rows of the data**

**print(data.head())**

**# Check the data shape**

**print(data.shape)**

**# Check for missing values**

**print(data.isnull().sum())**

**# Check data types**

**print(data.dtypes)**

**5. Data Preprocessing:**

- Preprocess the data to ensure it's suitable for association analysis. Common preprocessing steps include:

- Encoding the data into a suitable format for association rule mining. This is typically a binary format where each column represents an item, and a '1' or '0' indicates whether the item was purchased in a transaction.

- Handling missing values and ensuring all data is in the same format.

- Removing duplicate transactions.

- Scaling or normalizing data if necessary.

Here's an example of encoding the data into a suitable format using Python:

**Program:**

**# One-hot encode the data (convert it into a binary format)**

**encoded\_data = pd.get\_dummies(data)**

**# Remove duplicate transactions**

**encoded\_data = encoded\_data.drop\_duplicates()**

**# Reset the index**

**encoded\_data.reset\_index(drop=True, inplace=True)**

**6. Association Rule Mining:**

- Use an association rule mining algorithm such as Apriori or FP-growth to discover patterns and associations in the data. You can use libraries like `mlxtend` in Python for this purpose.

**Program:**

**from mlxtend.frequent\_patterns import apriori**

**from mlxtend.frequent\_patterns import association\_rules**

**# Perform Apriori algorithm**

**frequent\_itemsets = apriori(encoded\_data, min\_support=0.05, use\_colnames=True)**

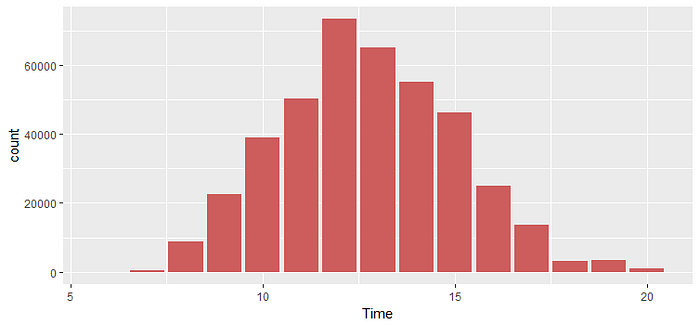
**# Generate association rules**

**rules = association\_rules(frequent\_itemsets, metric="lift", min\_threshold=1.0)**

**7. Analyze and Interpret Results:**

- Once you have generated association rules, analyze and interpret the results. You can filter the rules based on different metrics like support, confidence, and lift to identify meaningful patterns.

retail$Time <- as.factor(retail$Time)  
a <- hms(as.character(retail$Time))  
retail$Time = hour(a)retail %>%   
 ggplot(aes(x=Time)) +   
 geom\_histogram(stat="count",fill="indianred")



Shopping time distribution

**Program:**

**import pandas as pd**

**# Load the transaction data**

**data = pd.read\_csv('transaction\_data.csv')**

**# Display the first few rows of the data**

**print(data.head())**

**# Check the data shape**

**print(data.shape)**

**# Check for missing values**

**print(data.isnull().sum())**

**# Check data types**

**print(data.dtypes)**

**# One-hot encode the data (convert it into a binary format)**

**encoded\_data = pd.get\_dummies(data)**

**# Remove duplicate transactions**

**encoded\_data = encoded\_data.drop\_duplicates()**

**# Reset the index**

**encoded\_data.reset\_index(drop=True, inplace=True)**

**from mlxtend.frequent\_patterns import apriori**

**from mlxtend.frequent\_patterns import association\_rules**

**# Perform Apriori algorithm**

**frequent\_itemsets = apriori(encoded\_data, min\_support=0.05, use\_colnames=True)**

**# Generate association rules**

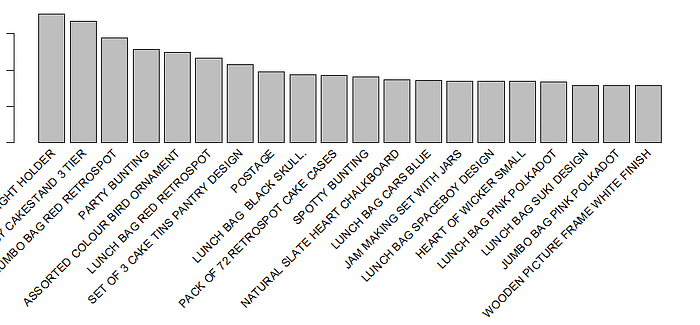
**rules = association\_rules(frequent\_itemsets, metric="lift", min\_threshold=1.0)**

**Output:**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | BillNo | Itemname | Quantity | Date | | Price | CustomerID | Country |  |
| 1 | 536365 | WHITE HANGING HEART T-LIGHT HOLDER | 6 | ######## | | 2.55 | 17850 | United Kingdom |
| 2 | 536365 | WHITE METAL LANTERN | 6 | ######## | | 3.39 | 17850 | United Kingdom |
| 3 | 536365 | CREAM CUPID HEARTS COAT HANGER | 8 | ######## | | 2.75 | 17850 | United Kingdom |
| 4 | 536365 | KNITTED UNION FLAG HOT WATER BOTTLE | 6 | ######## | | 3.39 | 17850 | United Kingdom |
| … | … | … | … | … | | … | … | … |
| 522061 | 581587 | PACK OF 20 SPACEBOY NAPKINS | 12 | ######## | | 0.85 | 12680 | France |
| 522062 | 581587 | CHILDREN'S APRON DOLLY GIRL | 6 | ######## | | 2.1 | 12680 | France |
| 522063 | 581587 | CHILDRENS CUTLERY DOLLY GIRL | 4 | ######## | 4.15 | | 12680 | France |
| 522064 | 581587 | CHILDRENS CUTLERY CIRCUS PARADE | 4 | ######## | | 4.15 | 12680 | France |
| 522065 | 581587 | BAKING SET 9 PIECE RETROSPOT | 3 | ######## | | 4.95 | 12680 | France |

**A bar plot of the support of the 20 most frequent items bought:**

itemFrequencyPlot(tr, topN=20, type='absolute')



**Conclusion:**

These insights provide a foundation for several key advantages:

* Personalized Marketing: Businesses can use market basket insights to tailor marketing efforts, creating more relevant and personalized offers for customers. This can increase customer engagement and drive sales.
* Inventory Management: By understanding which products are frequently purchased together, businesses can optimize their inventory management, ensuring that high-demand items are consistently in stock, reducing out-of-stock situations, and minimizing wastage.
* Store Layout and Design: Retailers can use market basket insights to improve the layout and design of their physical stores or websites, ensuring that complementary products are strategically placed together to encourage cross-selling.
* Pricing Strategies: Businesses can adjust pricing strategies based on market basket insights, offering discounts or bundling products that are often purchased together to increase sales and profitability.
* Product Recommendations: E-commerce platforms and online retailers can provide more accurate and relevant product recommendations to customers, increasing the chances of cross-selling and upselling.
* Customer Loyalty: By understanding customers' preferences and purchase patterns, businesses can build stronger customer loyalty and retention through targeted loyalty programs and incentives.

It's essential for businesses to continuously gather and analyze market basket data and adapt their strategies accordingly. Leveraging these insights effectively can lead to higher sales, better customer relationships, and a more competitive edge in the marketplace.

This is a high-level overview of how to start a market basket insights project by loading and preprocessing transaction data for association analysis. Depending on your specific dataset and goals, you may need to adjust these steps accordingly.