**PHASE 3 ASSIGNMENT**

**PROJECT TITLE:Preprocessing the Dataset**

**GITHUB LINK:** [**https://github.com/686969/market-basket-insights.git**](https://github.com/686969/market-basket-insights.git)

<https://github.com/686969/innovation.git>

**PROBLEM DEFINITION:** Market basket insights, or market basket analysis, is the process of discovering associations and patterns in customer transaction data. The primary goal is to uncover relationships between products or items that are frequently purchased together.

**Top of Form**

**DOCUMENT:**

**Building the project by preprocessing the data**

**DATASET LINK ON:**Market Basket Insights

[**https://www.kaggle.com/datasets/aslanahmedov/market-basket-analysis**](https://www.kaggle.com/datasets/aslanahmedov/market-basket-analysis)

### Pre-Requisites for Performing Market Basket Analysis :

Download the dataset before you start coding. Make sure you also have Jupyter Notebook installed on your device. If you are unfamiliar with the software, follow 365’s [beginner-friendly Jupyter Notebook tutorial](https://365datascience.com/tutorials/python-tutorials/jupyter-notebook-tutorial/) or [Introduction to Jupyter course](https://365datascience.com/courses/introduction-to-jupyter/) to learn about its usage and installation.

Finally, install the [pandas](https://pypi.org/project/pandas/) and [MLXtend](https://pypi.org/project/mlxtend/) libraries if you haven’t already.

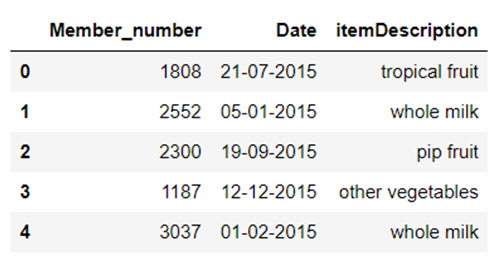
* **Reading the Dataset.**

Now, let’s read the dataset as a pandas data frame and take a look at its head:

**import** pandas **as** pd

df = pd.read\_csv('Groceries\_dataset.csv')

df.head()



### Data Preparation for Market Basket Analysis

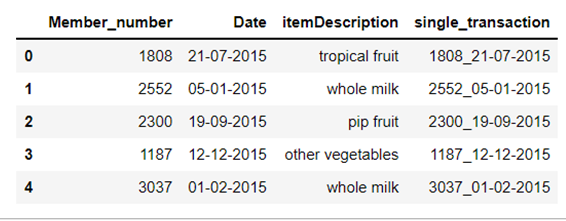
Before we perform market basket analysis, we need to convert this data into a format that can easily be ingested into the Apriori algorithm. In other words, we need to turn it into a tabular structure comprising ones and zeros, as displayed in the bread and milk example above.

To achieve this, the first group items that have the same member number and date:

df[‘single\_transaction’] = df[‘Member\_number’].astype(str)+’\_’+df[‘Date’].astype(str)

df.head()

This will provide us with a list of products purchased in the same transaction:

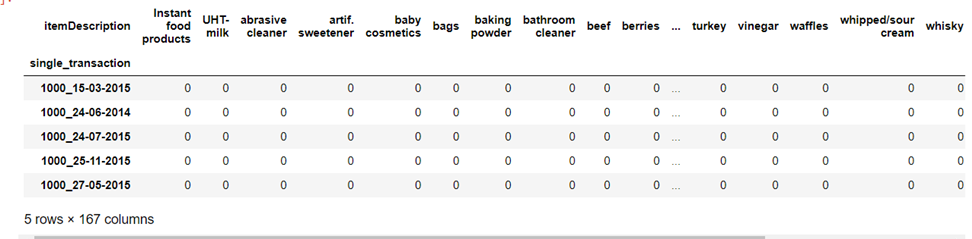


The “single\_transaction” variable combines the member number, and date, and tells us the item purchased in one receipt.Now, let’s pivot this table to convert the items into columns and the transaction into rows:

df2 = pd.crosstab(df['single\_transaction'], df['itemDescription'])

df2.head()

The resulting table tells us how many times each item has been purchased in one transaction:



There are over a hundred columns while most people only shop for 2-3 items, which is why this table is sparse and mostly comprised of zeroes.The final data pre-processing step involves encoding all values in the above data frame to 0 and 1.This means that even if there are multiples of the same item in the same transaction, the value will be encoded to 1 since market basket analysis does not take purchase frequency into consideration.

**def** **encode**(item\_freq):

res = 0

**if** item\_freq > 0:

res = 1

**return** res

basket\_input = df2.applymap(encode)

### Build the Apriori Algorithm for Market Basket Analysis.

Now, let’s import the Apriori algorithm from the MLXtend Python package and use it to discover frequently-bought-together item combinations:

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

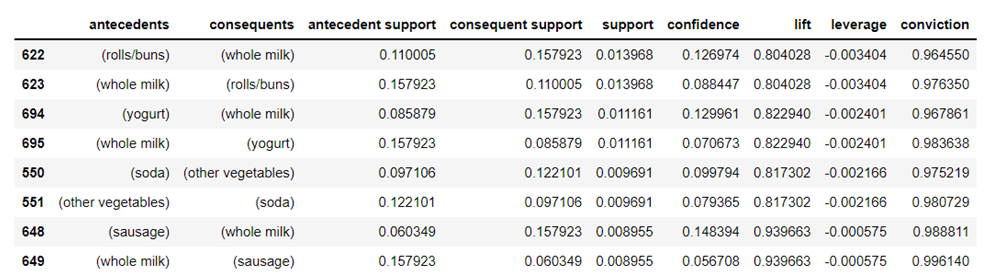
**from** mlxtend.frequent\_patterns **import** association\_rules

frequent\_itemsets = apriori(basket\_input, min\_support=0.001, use\_colnames=True)

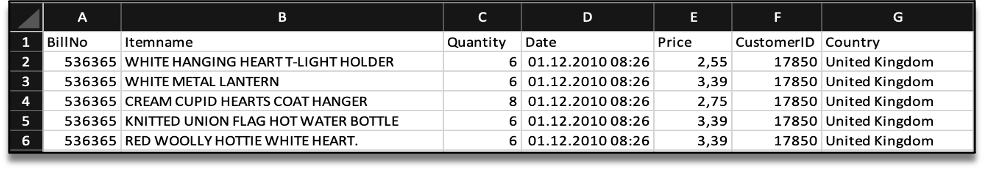
rules = association\_rules(frequent\_itemsets, metric="lift")

rules.head()

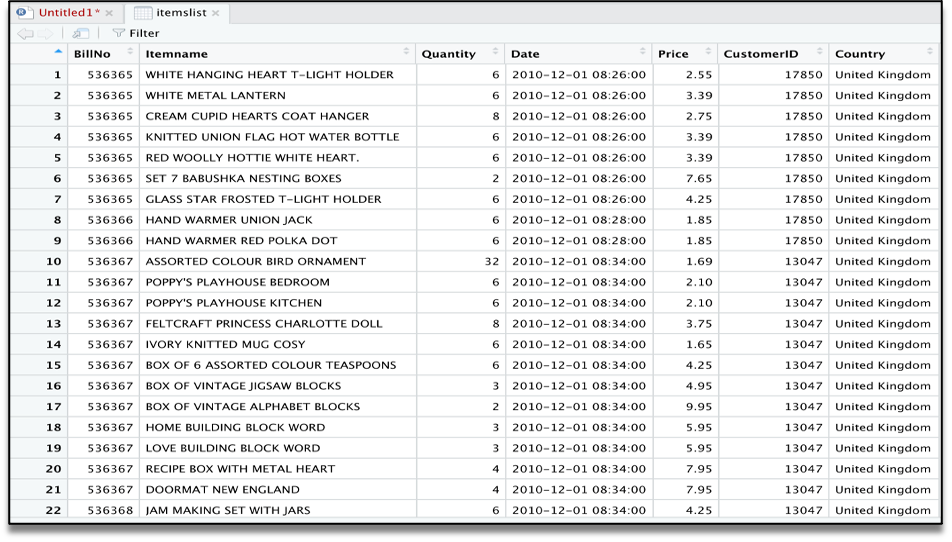
rules.sort\_values(["support", "confidence","lift"],axis = 0, ascending = False).head(8)

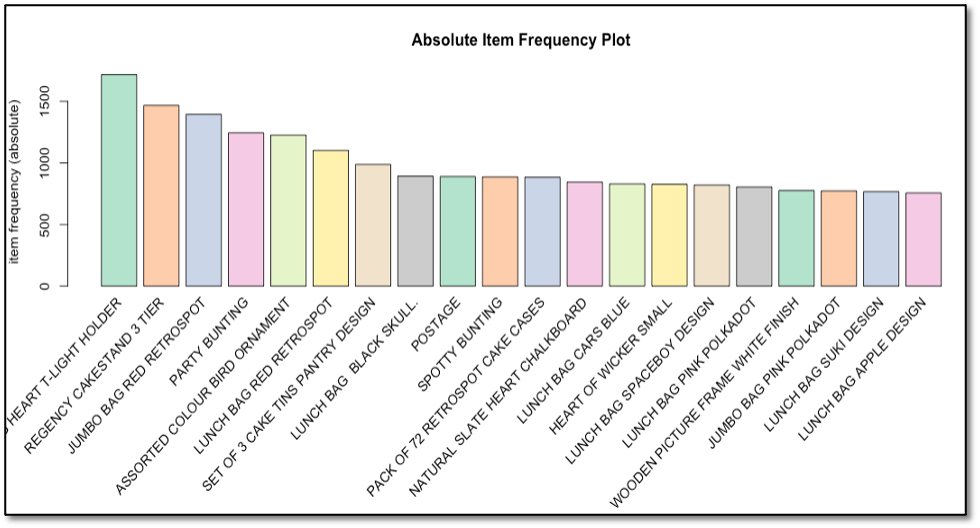


* **Example Dataset:**

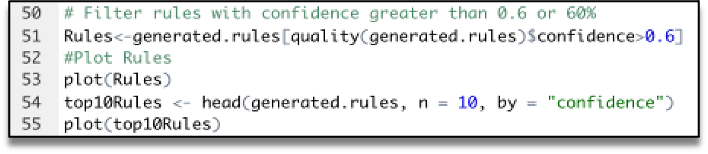
****

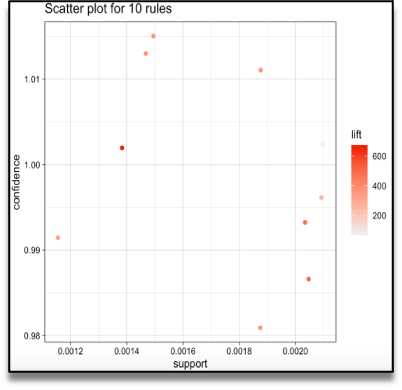
* **Pre-processing data:**

****

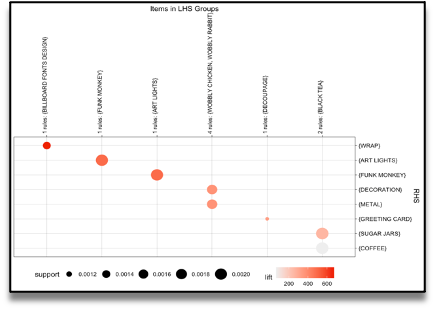
****

* **Scatter plot:**

****

****

#### Graph - Based Visualization and Group Method:



### Conclusion:

Based on the results of these calculations can be used as a recommendation for retail owners to arrange the arrangement of product catalogs and take strategic steps to improve product marketing.. By utilizing the association rules which are discovered as a result of the analyses, the retailer can apply effective marketing and sales promotion strategies, he will be able increase customer engagement and improve customer experience and identify customer behavior.

**SUBMITTED BY:**

**STUDENT REG NO:711221104011**

**NAAN MUDHALVAN: au711221104011**