**FINAL PROJECT SUBMISSION**

**Phase-5**

**Market Basket Insights**

|  |  |
| --- | --- |
|  |  |
| **Date** | **01.11.2023** |
| **Team ID** | **Proj-212168-Team-1** |
| **Maximum Mark** |  |

**Acknowledgement:**

IBM Corporation: I am deeply appreciative of IBM for providing the opportunity and resources to undertake this project. The access to data, tools, and support from IBM was instrumental in our analysis.Project Team: My sincere thanks to the dedicated team members who collaborated on this project. Your hard work, expertise, and commitment were essential in achieving our goals.Family and Friends: To my family and friends, thank you for your unwavering support and understanding during the demanding phases of this project.

**Abstract**

Market Basket Analysis is an important part of the analytical system in the retail organization to determine the placement of goods, designing sales promotion for different segments of customers to improve customer satisfaction and hence the profit of the supermarkets.MBA is well known activity of ARM ultimately used for business intelligent decisions. Mining frequent item sets and hence deduce rules to build classifiers with good accuracy is essential for efficient algorithm. The issues for a leading supermarket are addressed here using frequent item set mining. The project uses file as database. Here, the itemsets and transactions of items are kept in a matrix form representing rows as list of items and column as transactions. The frequent item sets are mined from database using the Apriori algorithm and then the association rules are generated. The project is beneficial for supermarket managers to determine the relationship between the items that are purchased by their customers.

**Keywords:**

Basket Analysis, Association Rule Mining, Apriori Algorithm

INDEX

**S.no CONTENTS page no**

1. List of Figures and Tables………………………………………………………………1

1.1) Histogram……………………………………………………………………1

1.2) Scatter Plot………………………………………………………… ……...1

1.3) Heatmap…………………………………………………………………. ....2

1.4) Box plot……………………………………………………………………......2

2. Introduction ……………………………………………………………………………...…….3

2.1) Market Basket Analysis………………………………………………..3

2.2) Goals of Project………………………………………………………………..3

2.3) Overview of Approach…………………………………………………….3

2.4)Document Structure………………………………………………………….4

3. Literature Survey…………………………………………………………………………..5

4. Problem definition, design thinking, Innovation problem solving…6

4.1. Problem Definition……………………………………………………………6

4.2. Design Thinking………………………………………………………………..…6

4.3. Innovation…………………………………………………………………………...6

4.4. Problem Solving………………………………………………………………….6

5. Importing dataset & performing data cleaning & Analysis………8

5.1. Import the Dataset……………………………………………………8

5.2. Data Cleaning………………………………………………………….....8

5.3. Data Analysis……………………………………………………………..8

5.4. Market Basket Analysis…………………………………………….8

5.5 Data Visualization………………………………………………………8

6. Data Visualization………………………………………………………………….......10

6.1 Bar Charts………………………………………….…………….10

6.2 Heatmaps………………………………………………………..10

6.3 Histogram……………………………………………..…………10

6.4 Pair Plot…………………………………………………………..10

6.5 Scatter Plot…………………………………………………….10

6.6 Box Plot………………………………………………………..10

7. Model development & Evaluation……………………………………11

7.1. Data Preparation……………………………………….11

7.2. Model Development……………………………………11

7.3. Rule Generation…………………………………………11

7.4. Evaluation Metrics……………………………………11

7.5. Evaluation Process…………………………………..11

7.6. Application…………………………………………………11

8. Code Sample…………………………………………………………………..10

9. Output Screenshot………………………………………………………… 13

9.1 Histogram…………………………………………….13

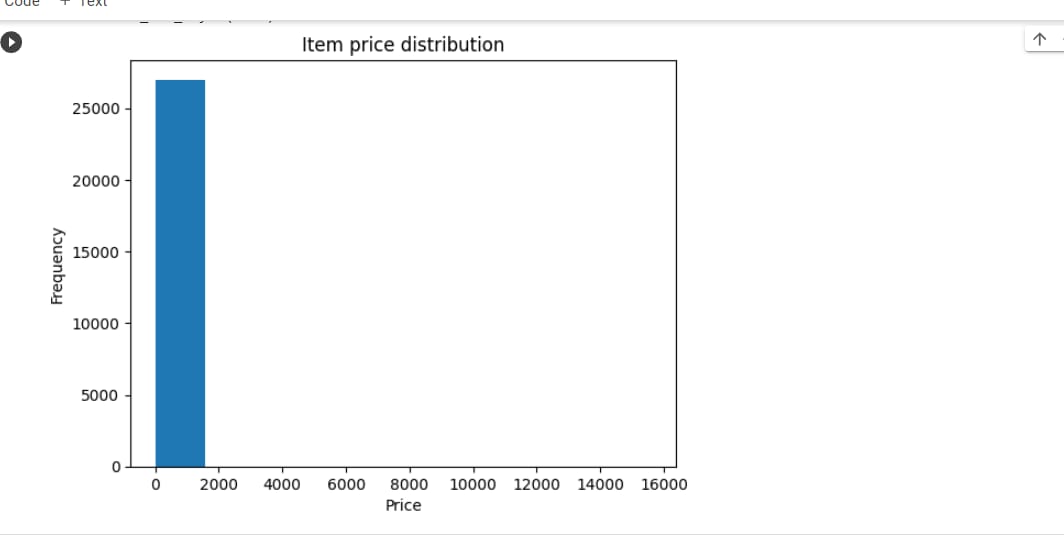
9.2 Heatmap……………………………………………….....13

9.3 Train & Test…………………………………………....15

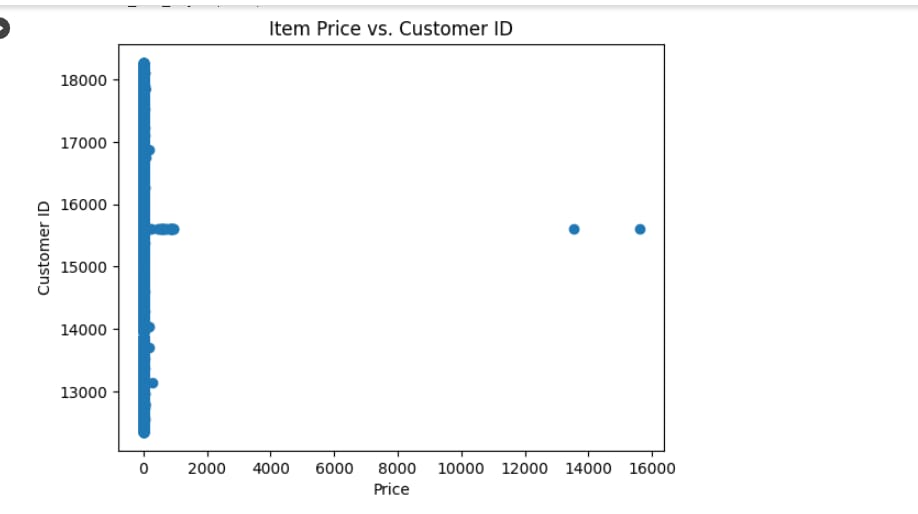
10. Conclusion & future enhancement, References………………1

**1. List of Figures**

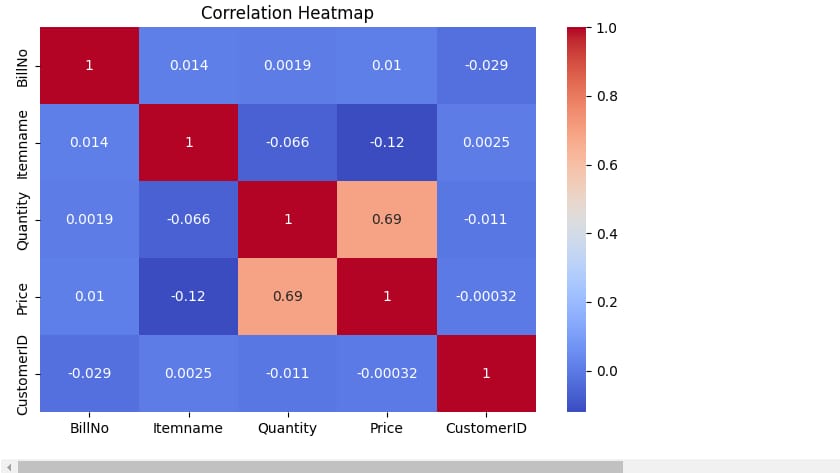
**1.1 Histogram:**

****

**1.2 Scatter Plot:**

****

**1.3 Heatmap**

****

**1.4 Box Plot**

****

**2.Introduction**

**2.1 Market Basket Analysis**

* Market Basket Analysis is a key method known and utilized by substantial retailers to reveal relationships between products, like bread, butter, etc. It works by searching for a mix of products that happen together every now and then in exchanges. To give it another perspective, it enables retailers to recognize connections between things that individuals purchase. With the continuous growth of information technology, massive amounts of data are collected and stored by enterprises. It is very important for enterprises to transform this data into useful information and knowledge for decision making in dynamic markets. This value added information discovered from Market Basket Analysis can be used to support decision making.
* If it is known that customers who purchase one product are likely to purchase another product, it is possible for retailers to market these products together, or to make the purchasers of a target prospects for the second product. If customers who purchase diapers are likely to purchase beer, they will be more likely to if beer is displayed just beside a diaper aisle. Though the – young fathers result does make sense, stocking up on supplies for themselves and for children before the weekend starts is something that someone would normally think of right away. The strength of those relationships is valuable information and can be used to cross-sell or up-sell.

**2.2 Goals of Project**

The main objective of the project is to make Instacart retailers to understand the current customer's behaviour and to predict future customers’ purchasing behaviour. Leveraging customer transaction data can help in understanding customers’ purchasing behaviour, offering right bundles and promotions, assortment planning and inventory management to retain customers, improve sales and extend their relationship with customers. The specific objectives of the project are as listed below

* To understand the purchasing pattern of products that comprise the customers’ basket.
* To study about many products usually purchased by the customers.
* To study the most likely products purchased by the customers along with a particular product category.
* To recommend and suggest products to individual customers.

**2.3 Overview of Approach**

The present technological time that we live in has made it feasible for business organisations to accumulate extensive data. Currently, Database technology innovation has sufficiently grown to keep these information stacks solid, however, it is significant not to simply keep that information, yet to assess the information to increase the value of the organisation. In today’s customer-centred markets, business needs to establish adequate and low advertising techniques that can react to changes in customer perceptions and demands for products. It might also assist business to recognise a whole new market strategy that can effectively target. All together for making key choices on the market strategy, stable, as much as could be and secured proof-based data is needed. With innovation, Data Mining has gotten perhaps the best response to this requirement. Data Mining is the process of refining important data from enormous Databases which includes a tremendous assortment of statistical and computational methods, neural network analysis, clustering, classification and summing up information. The computation of association rules, which is one of the Data Mining techniques implemented by Market Basket Research, is part of this project. The analysis is carried out on the Grocery stores' transaction data for the customers. The research goal is to consider the category of product that is likely to be marketed in conjunction by implementing Apriori and FP-Growth algorithms.

**2.4 Document Structure**

The document consists of seven chapters. Chapter One provides a brief introduction to the market basket analysis, project objectives, project overview and technology lists. In Chapter Two, addresses a literature review, results of similar research works and the related work has been discussed which provided insights to work on my project. The methodology and high-level design of the proposed system are defined in Chapter Three. The chapter Four, brief the architecture of the system and specifications, including the hardware and software used. The implementation details for the project are provided in Chapter Five. Details about working model, such as evaluation methods and results are provided in Chapter Six. Finally, Chapter 7 outlines findings and work for the future.

1. **Literature Survey**

In this section we have concentrated on presenting different areas where data mining algorithm and apriori algorithm used. The existing algorithms that were designed by the researchers in context of association rule mining in IBM project .Apply the rules and insights to improve marketing, recommendations,store layouts, or inventory management.

Monitor the impact of these strategies on sales and customer satisfaction.

**4.Problem definition, designthinking, Innovation problem solving:**

**4.1. Problem Definition:**

* Define the problem by identifying the goals, challenges, and opportunities in market basket analysis.
* Ask questions like: How can we increase cross-selling? What patterns can we discover in customer purchases? How can we improve the customer experience?

**4.2. Design Thinking:**

* Empathize: Understand the needs, motivations, and pain points of customers and stakeholders.
* Define: Clearly define the problem based on your research and insights.
* Ideate: Generate creative ideas for solutions, such as improving product placement, suggesting complementary items, or enhancing the shopping journey.
* Prototype: Develop and test potential solutions, which might include algorithms
* for recommendation systems or changes to store layouts.
* Test: Gather feedback through user testing and refine your solutions accordingly.

**4.3. Innovation:**

* Encourage a culture of innovation within your organization, where employees are encouraged to think creatively and contribute new ideas.
* Explore emerging technologies like machine learning and AI for advanced market basket analysis and personalized recommendations.
* Collaborate with other businesses or industry experts to share insights and ideas.

**4.4. Problem Solving:**

* Use data analysis techniques to discover patterns and associations in market baskets, such as frequent itemset mining and association rule mining.
* Implement solutions based on your findings, like real-time recommendations on e-commerce platforms or in-store promotions.

**5.Importing Dataset & Performing Datacleaning & Analysis**

**5.1. Import the Dataset:**

* Identify the source of your market basket data. It could be point-of-sale transactions, e-commerce logs, or any other relevant source.
* Use data import tools or libraries, such as Pandas for Python, to load your dataset.

**5.2. Data Cleaning:**

* This step involves preparing the data for analysis.
* Address missing data: Identify and decide how to handle missing values (e.g., imputation or removal).
* Remove duplicates: Check for and eliminate duplicate transactions or items.

**5.3. Data Analysis:**

* Perform exploratory data analysis (EDA) to understand your market basket dataset.
* Calculate basic statistics: Use summary statistics like count, mean, and standard deviation for numerical variables.
* Identify patterns: Explore the frequency of items purchased together and item co-occurrence patterns.
* Visualize the data: Create visualizations like bar charts, heatmaps, or network diagrams to better understand item relationships.

**5.4. Market Basket Analysis:**

* Conduct association rule mining to discover patterns in the dataset.
* Frequent Itemset Mining: Identify itemsets that occur frequently together.
* Association Rule Mining: Discover rules like "If a customer buys item A, they are likely to buy item B."

**5.5 Data Visualization:**

* Visualize the market basket insights using appropriate charts or graphs to make the findings more accessible to stakeholder .

**6.Data Visualization**

**6.1 Bar Charts:**

* Create bar charts to display the frequency of items in the market basket. This can help identify popular and less popular items.

**6.2 Heatmaps:**

* Use heatmaps to visualize item co-occurrence. The rows and columns represent items, and cells show the frequency or association strength between item pairs.

**6.3 Histogram:**

* The shape of this histogram can provide insights into which items are frequently bought together or which products are best-sellers. The items with the highest bars are the most commonly purchased, while those with low bars are less frequently purchased.

**6.4 Pair Plot:**

* Pair plots are often used in data analysis to visualize relationships between pairs of variables in a dataset. In the context of market basket analysis, you might consider creating pair plots to visualize associations between pairs of products or item categories.

**6.5 Scatter Plot:**

* Scatter plots can be a useful way to visually explore relationships between pairs of items, but they are best suited for understanding the relationships between two specific items at a time.

**6.6 Box Plot:**

* creating box plots for different items or categories in your market basket data, you can gain insights into how frequently certain items are included in transactions, how they vary across different transactions, and whether there are any unusual or outlier occurrences.

**7.Model Development Evaluation**

**7.1. Data Preparation:**

* Gather transaction data, where each row represents a customer transaction and each column represents items or product categories.
* Encode the data, typically as binary values (1 for item presence, 0 for absence).

**7.2. Model Development:**

* Choose a market basket analysis technique, such as Apriori, association rule mining.
* Set parameters like minimum support and confidence levels to define what constitutes a meaningful association.

**7.3. Rule Generation:**

* Generate association rules that capture item relationships. These rules often have the form "If item A is purchased, then item B is likely to be purchased."

**7.4. Evaluation Metrics:**

* Support: Measures how frequently a rule occurs in the dataset.
* Confidence: Indicates the likelihood that item B is bought when item A is purchased.
* Lift: Measures how much more likely item B is bought when item A is purchased compared to random chance.
* Leverage: Measures the difference between the observed co-occurrence and the expected co-occurrence of items.

**7.5. Evaluation Process:**

* Evaluate the generated rules using the chosen metrics.
* Focus on high-confidence and high-support rules to identify meaningful associations.

**7.6. Application:**

* Apply the rules and insights to improve marketing, recommendations, store layouts, or inventory management.
* Monitor the impact of these strategies on sales and customer satisfaction.

**8.Code Sample:**

import numpy as np

import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sns

from mlxtend.frequent\_patterns import apriori

from mlxtend.frequent\_patterns import association\_rules

dataset=pd.read\_csv('insights (1).csv',encoding="UTF-8")

df=dataset.fillna({'Itemname':'abc'})

df

df1=dataset.fillna(value=dataset['CustomerID'].mean())

df1

df1.isnull().sum()

df2=pd.DataFrame(df1)

item\_prices = df['Price']

item\_prices.plot(kind='hist', bins=10)

plt.title('Item Price Distribution')

plt.xlabel('Price')

plt.ylabel('Frequency')

plt.show()

df2['Itemname'] = df2['Itemname'].astype(str).str.replace('[^0-9.]', '', regex=True)

df2['Itemname'] = pd.to\_numeric(df2['Itemname'], errors='coerce')

plt.figure(figsize=(8, 5))

correlation\_matrix = df2.corr()

sns.heatmap(correlation\_matrix, annot=True, cmap='coolwarm')

plt.title('Correlation Heatmap')

plt.show()

basket = (dataset[dataset['Country'] == 'Germany' ].groupby(['BillNo','Itemname'])['Quantity'].sum().unstack().fillna(0))

basket

defencode(x):

    if x <= 0:

        return0

    if x >= 1:

        return1

basket = basket.applymap(encode)

basket

from mlxtend.frequent\_patterns import apriori

from mlxtend.frequent\_patterns import association\_rules

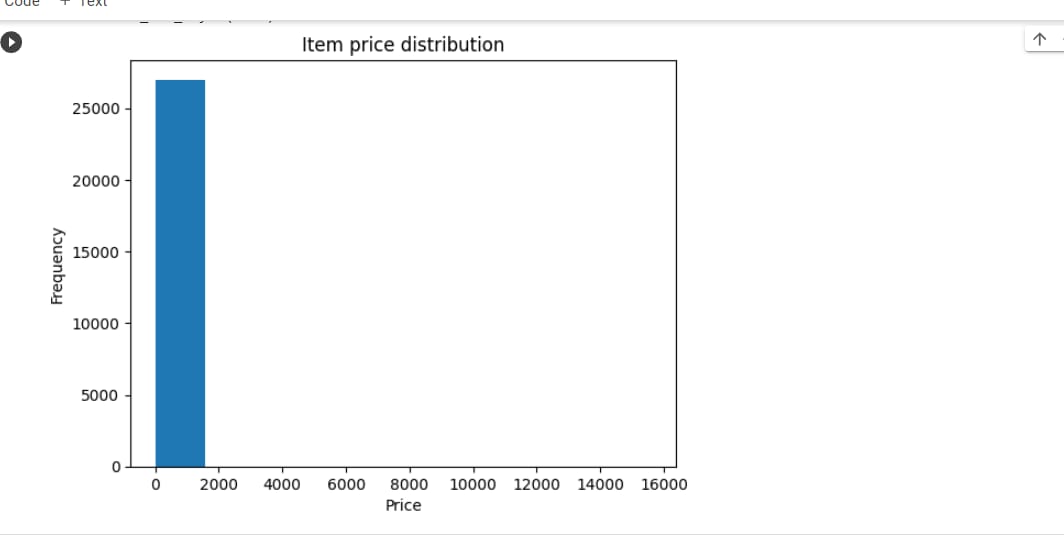
frequent\_itemsets = apriori(basket, min\_support = 0.06, use\_colnames= True)

rules =association\_rules(frequent\_itemsets,metric='lift',min\_threshold=1)

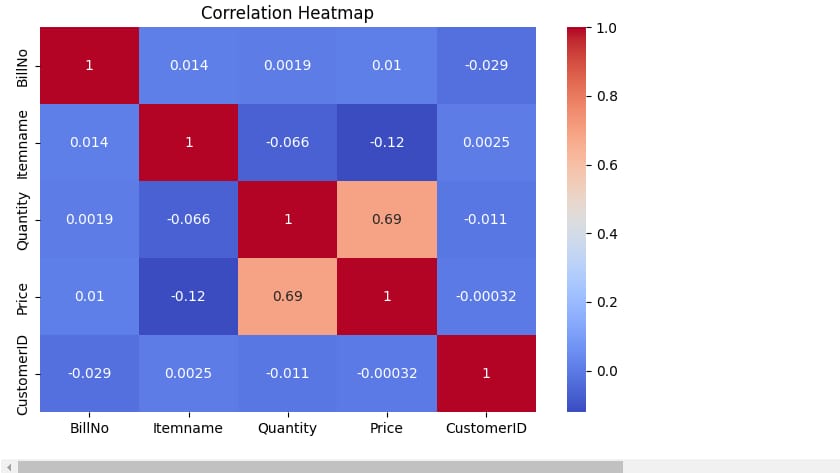
rules[(rules['confidence']>0.4) & (rules['lift']>1)]

**9.Output:**

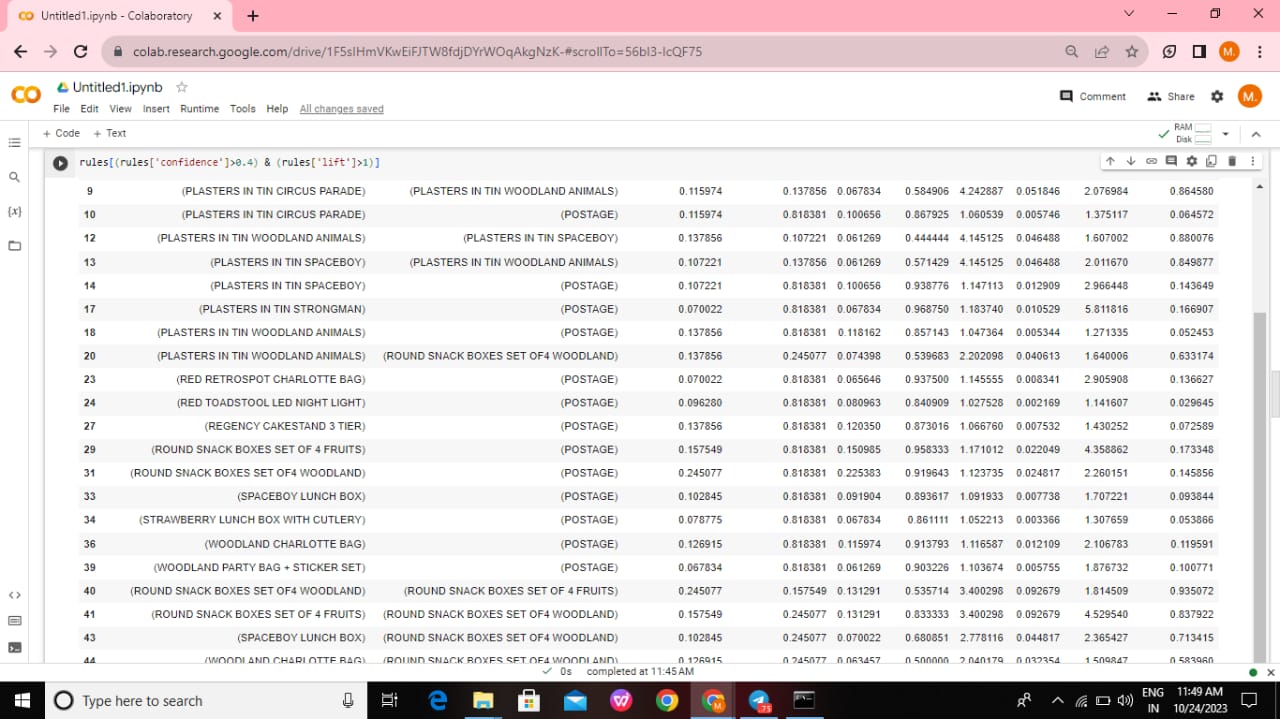
**9.1 Histogram:**

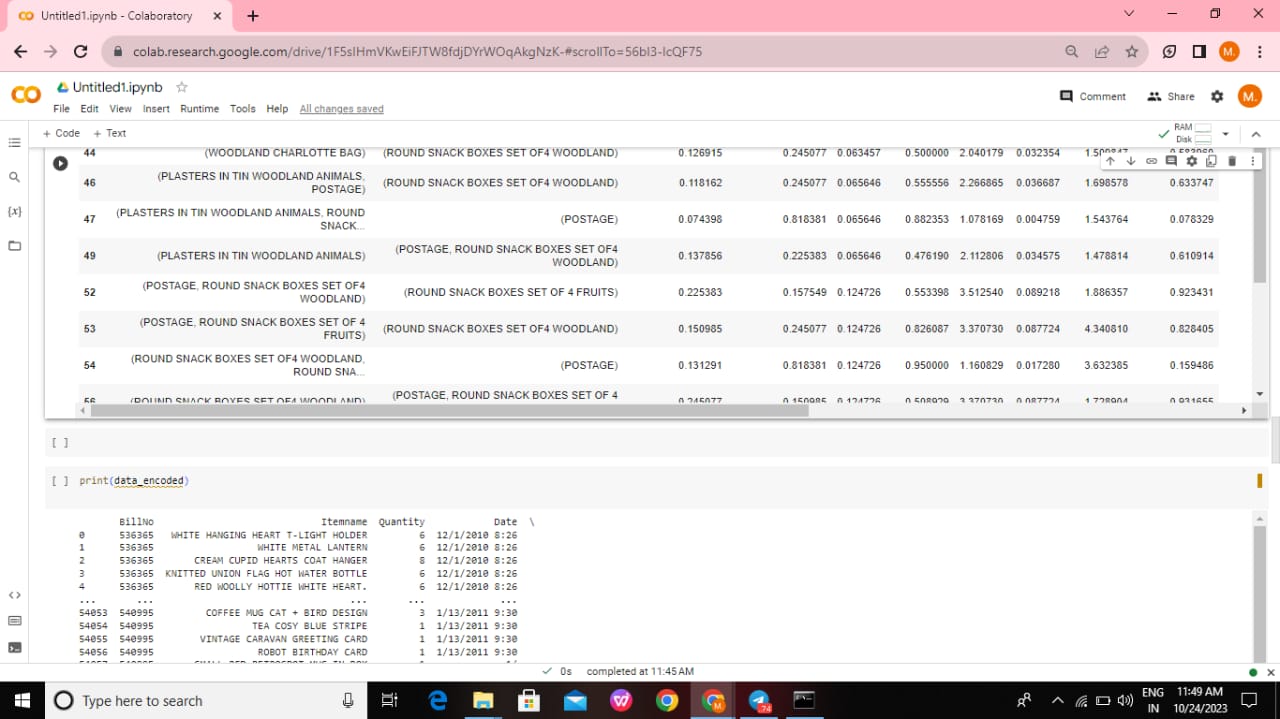


**9.2 Heatmap:**



**9.3 Train & Test:**

****



**10.Conclusion:**

Understanding Customer Behavior: We have gained a deeper understanding of customer purchasing behavior, uncovering associations and patterns in their product choices. This knowledge is instrumental in tailoring marketing strategies and enhancing customer experiences. Improved Product Recommendations: The project has enabled us to create more accurate and personalized product recommendations, leading to increased sales and customer satisfaction. By leveraging association rules, we can suggest complementary products effectively.Enhanced Decision-Making: With the use of market basket analysis, businesses can make data-driven decisions, such as product placement, pricing strategies, and marketing campaigns. This leads to more efficient and effective operations.Data Privacy Considerations: We acknowledge the importance of data privacy and ethical concerns when handling customer transaction data.

**References:**

Manideepavadootha16. (Year). Market Basket Dataset. Dataset on kaggle

<https://www.kaggle.com/code/manideepavadootha16/market-basket>