MARKET BASKET ANALYSIS

PROJECT REPORT

18IPE415J – FOUNDATION OF ANALYTICS
III Year/ V Semester
Academic Year: 2023 -2024
By

PALAASH SURANA(RA2111003010319)

KAMYA OJHA (RA2111003010343)

Under the guidance of

Dr. A. REVATHI

Assistant Professor

Department of Computational Intelligence

FACULTY OF ENGINEERING AND TECHNOLOGY

SRM INSTITUTE OF SCIENCE AND TECHNOLOGY

Kattankulathur, Chengalpattu District

NOVEMBER 2023

COLLEGE OF ENGINEERING & TECHNOLOGY
SRM INSTITUTE OF SCIENCE & TECHNOLOGY
SRM NAGAR, KATTANKULATHUR- 603203,
CHENGALPATTU DISTRICT

(Under Section 3 of UGC Act, 1956)

BONAFIDE CERTIFICATE

This is to certify that 18CSE415J – FOUNDATION OF ANALYTICS project report titled "MARKET BASKET ANALYSIS" is the bonafide work of KAMYA OJHA(RA2111003010343) and PALAASH SURANA (RA2111003010319) who undertook the task of completing the project within the allotted time.

SIGNATURE
Dr. A. Revathi
Course Faculty
Assistant Professor
Department of Computational Intelligence
SRM Institute of Science and Technology
Kattankulathur.

SIGNATURE
Dr. M. Pushpalatha
Head of the Department
Professor
Department of Computing Technologies
SRM Institute of Science and Technology
Kattankulathur.

ABSTRACT

Market Basket Analysis (MBA) is a data mining technique that plays a crucial role in uncovering hidden patterns and associations within transactional datasets. This analytical approach focuses on understanding the relationships between products frequently purchased together by customers, providing valuable insights for businesses in optimizing their marketing strategies, inventory management, and overall customer experience.

This abstract delves into the key aspects of Market Basket Analysis, highlighting its significance in the retail and e-commerce sectors. The analysis begins with an overview of the methodology, emphasizing the use of algorithms such as Apriori and FP-Growth to efficiently identify frequent itemsets and association rules. These rules offer businesses actionable intelligence to enhance cross-selling, promotional targeting, and product bundling strategies.

Furthermore, the abstract explores the various applications of Market Basket Analysis beyond the retail sector, extending its utility to diverse industries such as healthcare, telecommunications, and online streaming platforms. The adaptability of this technique underscores its versatility in extracting meaningful associations from transactional data, leading to informed decision-making and strategic planning.

Additionally, the abstract addresses the challenges and considerations associated with Market Basket Analysis, including data preprocessing, scalability issues, and privacy concerns. The evolution of advanced techniques, such as machine learning algorithms, in addressing these challenges is also discussed.

In conclusion, this abstract emphasizes the importance of Market Basket Analysis as a powerful tool for businesses to gain a deeper understanding of customer behavior, foster personalized marketing approaches, and ultimately enhance operational efficiency. As industries continue to evolve in the era of big data, Market Basket Analysis stands out as a valuable asset for organizations seeking to stay competitive and responsive to ever-changing consumer demands.

ACKNOWLEDGEMENT

We express our heartfelt thanks to our honorable Vice Chancellor Dr. C. MUTHAMIZHCHELVAN, for being the beacon in all our endeavors.

We would like to express my warmth of gratitude to our Registrar Dr. S. Ponnusamy, for his encouragement.

We express our profound gratitude to our Dean, College of Engineering and Technology, Dr. T. V.Gopal, for bringing out novelty in all executions.

We would like to express my heartfelt thanks to Chairperson, School of Computing Dr. Revathi Venkataraman, for imparting confidence to complete my course project

We are highly thankful to our my Course project Faculty Dr.A.Revathi, Assistant Professor, Department of Computational Intelligence, for his/her assistance, timely suggestion and guidance throughout the duration of this course project.

We extend my gratitude to our HoD Dr.M.Pushpalatha, Professor, Department of Computing Technologies and my Departmental colleagues for their Support.

Finally, we thank our parents and friends near and dear ones who directly and indirectly contributed to the successful completion of our project. Above all, I thank the almighty for showering his blessings on me to complete my Course project.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

Introduction:

In the ever-evolving landscape of business and commerce, understanding consumer behavior is paramount for organizations striving to stay ahead. Market Basket Analysis (MBA) stands out as a powerful analytical tool that allows businesses to uncover intricate patterns within their transactional data. Rooted in the principles of association rule mining, MBA provides a nuanced approach to unraveling the relationships between products that are frequently purchased together. This introduction aims to shed light on the significance of Market Basket Analysis, its underlying methodologies, and the far-reaching implications it holds for businesses across diverse industries.

Motivation:

In the dynamic landscape of commerce, businesses are constantly seeking innovative ways to gain a competitive edge. One such avenue that has garnered significant attention is Market Basket Analysis (MBA). The motivation behind embracing MBA lies in the profound desire to unravel the hidden patterns within consumer transactions, thereby empowering businesses with actionable insights that can drive strategic decision-making and enhance overall operational efficiency.

Objective:

Market Basket Analysis (MBA) serves a specific set of objectives aimed at extracting meaningful insights from consumer transactions. The primary goals revolve around understanding consumer behavior, optimizing operational processes, and informing strategic decision-making within various aspects of the business.

1. Understand Consumer Behavior: The fundamental objective of Market Basket Analysis is to gain a profound understanding of consumer behavior. By scrutinizing transactional data, businesses aim to uncover patterns and associations between products frequently purchased together. This knowledge goes beyond individual product preferences, providing insights into the interconnected choices consumers make during their shopping journey. Understanding these intricate relationships allows businesses to align their strategies with consumer expectations and preferences.

2. Identify Product Associations: Market Basket Analysis aims to identify and establish associations between different products. The objective is to unveil correlations, dependencies, and co-occurrences within consumer transactions. This involves recognizing which products are often bought together and understanding the potential reasons behind these associations. By identifying these patterns, businesses can create synergies between products, facilitating strategic decisions related to inventory management, product placement, and marketing strategies.

3. Optimize Marketing Strategies: An essential objective of MBA is to optimize marketing strategies. Once associations between products are identified, businesses can tailor their marketing efforts to capitalize on these connections. Whether through targeted advertising, bundled promotions, or cross-selling initiatives, the goal is to enhance the effectiveness of marketing campaigns. MBA allows businesses to tailor their messaging to align with consumer preferences, thereby increasing the impact of marketing efforts.

4. Improve Inventory Management: Efficient inventory management is a key

objective of Market Basket Analysis. By understanding which products are commonly purchased together, businesses can streamline their inventory to meet consumer demand more effectively. This involves ensuring the availability of complementary products, managing stock levels based on consumer preferences, and reducing the likelihood of stockouts or overstock situations.

5. Enhance Cross-Selling and Upselling: Market Basket Analysis aims to facilitate cross-selling and upselling opportunities. Once associations between products are identified, businesses can strategically bundle or suggest related items to customers. This objective is geared towards increasing the average transaction value, maximizing revenue, and providing customers with a more comprehensive and satisfying shopping experience.

5. Drive Informed Decision-Making: Ultimately, the overarching objective of Market Basket Analysis is to drive informed decision-making across various facets of the business. From marketing and inventory management to product development and customer engagement, the insights derived from MBA empower decision-makers with the knowledge needed to make strategic choices that align with consumer preferences and market trends.

Problem Statement:

In the dynamic landscape of modern commerce, businesses face a myriad of challenges in implementing and maximizing the effectiveness of Market Basket Analysis (MBA). Despite its potential for revealing valuable insights into consumer behavior, optimizing marketing strategies, and improving operational efficiency, several issues hinder the seamless integration and utilization of MBA in various industries. This problem statement aims to identify and address the challenges that organizations encounter when implementing Market Basket

Analysis, thus paving the way for more informed decision-making and strategic planning.

1. Data Quality and Preprocessing Challenges:

One of the primary challenges in Market Basket Analysis is the quality of input data. Incomplete, inaccurate, or inconsistent data can lead to misleading associations and hinder the extraction of meaningful patterns. Preprocessing these vast datasets is a cumbersome task, requiring careful consideration of data cleaning, normalization, and addressing missing values to ensure the accuracy and reliability of the results derived from MBA.

2. Scalability Issues:

As transactional datasets grow in size and complexity, scalability becomes a significant hurdle. Traditional algorithms, such as Apriori, may struggle to handle large datasets efficiently, impacting the speed and performance of MBA. Organizations grapple with the need for advanced algorithms and distributed computing frameworks to ensure timely and effective analysis, particularly in industries dealing with massive volumes of transactional data.

3. Dynamic Consumer Behavior:

Consumer preferences are inherently dynamic, influenced by trends, seasons, and external factors. The challenge lies in the ability of Market Basket Analysis to adapt to these changes effectively. Static analyses may fail to capture evolving consumer behaviors, requiring continuous monitoring and real-time adjustments to ensure the relevance and accuracy of MBA insights.

4. Interpretability of Results:

While MBA generates association rules and patterns, the interpretability of these results can pose a challenge. Businesses may struggle to translate complex data-driven insights into actionable strategies. The gap between analytical findings and practical decision-making requires bridging, often necessitating the involvement of cross-disciplinary teams to ensure effective communication and

implementation of MBA insights.

Scope:

Market Basket Analysis (MBA) offers a broad and versatile scope, providing organizations with opportunities to glean valuable insights and optimize various facets of their operations. The scope of MBA extends across industries, contributing to informed decision-making and strategic planning. Key aspects of the scope include:

1.Retail Optimization:

Enhance cross-selling and upselling.

Optimize inventory management.

E-commerce Enhancement:

2.Personalize recommendations.

Implement dynamic pricing strategies.

3. Healthcare Tailoring:

Inform personalized treatment plans

Optimize supply chain efficiency

4.Telecommunications Optimization:

Identify service bundling opportunities

Refine customer retention strategies

5. Online Streaming Personalization:

Improve content recommendations

Enhance viewer engagement

CHAPTER 2

REQUIREMENTS

Dataset: A diverse and representative dataset of resumes categorized by professional fields (e.g., IT, Marketing, Engineering, Finance, etc.). The dataset should be prepared in a structured format, allowing for easy integration into the project.

Google Colab: Utilize Google Colab, a cloud-based Jupyter notebook environment, for developing and implementing the MBA-based resume parser. Google Colab provides access to free GPU resources, which can significantly accelerate model training.

Machine Learning Framework: Use a machine learning framework such as TensorFlow or PyTorch for building and training MBA models.

Tableau Desktop: Employ Tableau Desktop for data visualization and analysis. Connect Tableau to the parser's output data to create informative dashboards and reports for recruiters and HR professionals.

Python Programming: Proficiency in Python is essential for scripting, data manipulation, and model development.

Natural Language Processing Expertise: A strong understanding of NLP techniques and methodologies is required to effectively preprocess and analyze text data.

Machine Learning Skills: Proficiency in machine learning techniques and deep learning architectures, particularly for fine-tuning models and optimizing parsing accuracy.

Data Integration: Develop a mechanism for integrating the parser's output with Tableau Desktop, ensuring seamless data flow and visualization.

User Interface (UI): Design a user-friendly interface for HR professionals to interact with the resume parser, view results, and generate reports using Tableau dashboards.

CHAPTER 3

DATASET DESCRIPTION

The data set captures the dynamics of customer transactions, providing a snapshot of purchases made on specific dates and times. Each transaction is uniquely identified by a Transaction ID, and individual items are distinguished by their Item ID and Item Name. This structured format enables a detailed exploration of customer preferences and associations between different products.

Sample Records:

Date	Time	Transaction	Item
30-10-2016	30-12-1899 09:58:11	1	Bread
30-10-2016	30-12-1899 10:05:34	2	Scandinavian
30-10-2016	30-12-1899 10:05:34	2	Scandinavian
30-10-2016	30-12-1899 10:07:57	3	Hot chocolate
30-10-2016	30-12-1899 10:07:57	3	Jam
30-10-2016	30-12-1899 10:07:57	3	Cookies
30-10-2016	30-12-1899 10:08:41	4	Muffin
30-10-2016	30-12-1899 10:13:03	5	Coffee
30-10-2016	30-12-1899 10:13:03	5	Pastry
30-10-2016	30-12-1899 10:13:03	5	Bread
30-10-2016	30-12-1899 10:16:55	6	Medialuna
30-10-2016	30-12-1899 10:16:55	6	Pastry

FIG 3.1: Sample Records

Dataset Description:

Date:

Description: The date when the transaction occurred.

Example: 2023-01-15

Time:

Description: The time when the transaction occurred.

Example: 14:30:00

Transaction ID:

Description: A unique identifier for each transaction.

Example: T001, T002, T003, ...

Item ID:

Description: A unique identifier for each item/product.

Example: I001, I002, I003, ...

Item Name:

Description: The name or description of the item/product.

Example: Milk, Bread, Eggs, ...

Dataset Characteristics:

Transactional Structure:

Description: The dataset follows a transactional structure, with each record representing a unique purchase transaction.

Significance: This structure is crucial for the analysis of item associations within the context of customer transactions.

Date and Time Information:

Description: The dataset includes columns for date and time, providing temporal context for each transaction.

Significance: Allows for the examination of temporal patterns, including daily, weekly, or seasonal variations in customer purchasing behavior.

Unique Identifiers:

Description: Unique identifiers are assigned to each transaction and item (e.g., Transaction ID, Item ID).

Significance: Enables the tracking of individual transactions and items, facilitating the identification of purchase patterns.

Item Details:

Description: Information about each item is included, often with an Item Name or description.

Significance: Facilitates the interpretation of specific products, contributing to the comprehensibility of the analysis results.

Quantitative Measures:

Description: The dataset may include quantitative measures such as quantity, price, and total purchase amount.

Significance: Allows for the consideration of quantitative aspects, such as the volume of items purchased and the monetary value of transactions.

Sparse Matrix Representation:

Description: The dataset may be represented as a sparse matrix, where most entries are zero, indicating items not purchased in a given transaction.

Significance: Supports the use of algorithms like Apriori, which are efficient for analyzing sparse data to discover frequent itemsets and association rules.

Granularity:

Description: Granularity can vary, with the dataset containing either individual items or

aggregated categories.

Significance: The level of granularity affects the specificity of insights; more granular data allows for detailed item-level analysis.

Relevance to Business Goals:

Description: The dataset aligns with specific business goals, such as improving sales, optimizing product placement, or enhancing customer satisfaction.

Significance: Ensures that the analysis results are directly applicable and beneficial to the business.

Volume and Diversity:

Description: The dataset should be sufficiently large and diverse to capture a representative sample of customer behavior.

Significance: A larger and more diverse dataset increases the likelihood of discovering meaningful patterns and associations.

Data Quality:

Description: The dataset is expected to be clean, accurate, and free of errors or inconsistencies.

Significance: High-quality data is essential for obtaining reliable insights and avoiding misinterpretations in the analysis.

These characteristics collectively contribute to the suitability of the dataset for Market Basket Analysis, providing a foundation for uncovering valuable patterns and associations in customer purchasing behavior based on date, time, transaction, and item information.

Purpose:

The primary purpose of this dataset is to train and test our resume parser and analysis

system. By processing and analyzing resumes, we aim to develop a tool that can extract meaningful information from unstructured text, including skills, qualifications, and experience. This tool will facilitate efficient recruitment processes, enabling the identification of qualified candidates based on predefined criteria.

Data Quality:

We have taken measures to ensure data quality by verifying the accuracy of categorization and the correctness of resume content. However, as with any real-world dataset, there may be variations in data quality and completeness.

CHAPTER 4

EXPLORATORY DATA ANALAYSIS

DATASET PREPARATION:

Punctuation Removal:

Punctuation marks are an integral part of written language, serving to convey meaning and structure in text. However, in data analysis and natural language processing, it's often necessary to remove punctuation for several reasons.

Cleaner Data: Punctuation removal results in cleaner and more structured data. It eliminates characters such as periods, commas, question marks, and exclamation marks, which are essential for sentence structure but may not provide substantial information for certain text analysis tasks.

Standardized Content: Removing punctuation ensures that the text data is standardized and consistent. This standardization is crucial when comparing, processing, or analyzing text, as it minimizes variations due to punctuation choices.

Improved Readability: Punctuation removal can enhance the readability of the text data. It reduces the visual clutter caused by punctuation marks, making it easier for both humans and machines to comprehend the text.

Preparation for Tokenization: Tokenization, the next step in text analysis, often requires punctuation removal. By eliminating punctuation beforehand, the tokenization process becomes more straightforward, as words are separated without interference from punctuation marks.

Feature Extraction: In some cases, punctuation removal is a part of feature extraction. For example, when analyzing sentiment in text, punctuation is often irrelevant and can be safely disregarded.

Tokenization:

Tokenization is a fundamental text analysis process that breaks down a continuous string of text into smaller units, usually words or phrases, known as tokens. This step is vital for various reasons:

Unit of Analysis: Tokenization defines the basic units of analysis in natural language processing. Each token represents a discrete element of meaning within the text.

Structure Recognition: Tokenization identifies the structural elements of the text, such as words, sentences, or paragraphs. This structural information is crucial for linguistic analysis.

Information Retrieval: Tokenization allows for efficient information retrieval. By dividing the text into tokens, it becomes easier to search for specific words or phrases within a document.

Statistical Analysis: Tokenized data is amenable to statistical analysis. It enables the calculation of word frequencies, co-occurrence patterns, and other linguistic statistics, which are valuable in various MBA tasks.

Text Classification: Tokenization is a prerequisite for text classification tasks. Texts are converted into numerical feature vectors, where each token may

become a feature with its associated weight.

This Punctuation removal enhances data cleanliness and standardization, while tokenization establishes the fundamental units for further linguistic and statistical analysis. These processes set the stage for a wide range of natural language processing and text analysis tasks.

DATA ANALYSIS:

Performing data analysis for Market Basket Analysis involves exploring the relationships between items purchased in transactions. The dataset structure you provided includes columns for date, time, transaction, and item. Below are the steps for conducting Market Basket Analysis using this dataset:

1. Data Loading:

Load the dataset into a suitable data analysis tool (e.g., with pandas, R, or any other tool of your choice).

```
import pandas as pd

# Load the dataset
df = pd.read_csv("market_basket_data.csv")
```

2. Data Exploration:

Explore basic statistics and information about the dataset.

```
# Display basic information about the dataset
print(df.info())

# Display summary statistics
print(df.describe())
```

3. Preprocessing:

Ensure data consistency, handle missing values, and convert date and time columns to datetime objects.

```
# Convert date and time columns to datetime objects
df['Date'] = pd.to_datetime(df['Date'])
df['Time'] = pd.to_datetime(df['Time'])
# Handle any other preprocessing steps as needed (e.g., handling missing values)
```

4. Transaction Level Data:

Group the data by transactions to create a transaction-level dataset.

```
# Group by transaction ID and aggregate items
transactions = df.groupby(['Date', 'Time', 'Transaction'])['Item'].agg(lambda x: list(x)).reset_index()
```

5. Association Rule Mining:

Apply association rule mining algorithms to identify frequent itemsets and association rules.

```
from mlxtend.frequent_patterns import apriori, association_rules
```

```
# Convert the transaction data to a one-hot encoded format
one_hot_encoded = transactions['Item'].str.join('|').str.get_dummies('|')

# Apply Apriori algorithm to find frequent itemsets
frequent_itemsets = apriori(one_hot_encoded, min_support=0.02,
use_colnames=True)

# Generate association rules
rules = association_rules(frequent_itemsets, metric="lift", min_threshold=1)
```

6. Analyzing Results:

Explore and interpret the generated association rules.

```
# Display the top association rules based on lift
top_rules = rules.sort_values(by='lift', ascending=False).head(10)
print(top_rules)
```

7. Visualization:

Visualize key patterns or associations to aid interpretation.

```
import networkx as nx
import matplotlib.pyplot as plt

# Create a network graph of association rules
G = nx.from_pandas_edgelist(top_rules, 'antecedents', 'consequents')
nx.draw(G, with_labels=True, node_size=2000, node_color="skyblue",
font_size=8, font_color="black", font_weight="bold")
```

plt.show()

8. Interpretation:

Interpret the results to make actionable insights for business decisions.

Note:

Adjust parameters such as min_support and other threshold values based on the characteristics of your dataset and the desired level of significance for association rules.

This basic example provides a framework for performing Market Basket Analysis. The effectiveness of the analysis depends on the dataset's characteristics and the business context. Adjustments and additional steps may be necessary based on the specifics of your dataset and analysis goals.

DATA VISUALIZATION

1) **BestSellers**: The resulting bar plot visualizes the top 10 best-selling items, showing the number of transactions for each item. The y-axis represents the count of transactions, and the x-axis represents the individual items. The plot provides a quick overview of which items are the most popular or frequently purchased based on the provided dataset. (FIG 4.1)

2) **Business During Past Months**: The resulting bar plot visualizes the business activity over different months, providing insights into the variations in transaction volume over time. Each bar represents a specific month, and the height of the bar corresponds to the number of transactions that occurred during that month. The plot is useful for identifying trends and patterns in business activity across different

periods. Peaks in transaction counts might indicate busier months, while valleys could represent slower periods. This information is valuable for businesses to understand the seasonality of customer behavior, plan for fluctuations in demand, and make informed decisions regarding inventory management, marketing strategies, and resource allocation. (FIG 4.2)

3) **Daily Business During Past Months**: The resulting plot provides a time series view of the business activity over the past months. By resampling the data on a monthly basis, you can observe trends, seasonality, and fluctuations in transaction counts over time. Trends: The overall direction of the line can reveal whether the business is experiencing growth, decline, or stability in transaction volume. Seasonality: Repeated patterns or cycles in the data may indicate seasonality. For example, certain months may consistently have higher or lower transaction counts due to factors like holidays, promotions, or specific events. Fluctuations: Sharp increases or decreases in transaction counts may highlight specific events or changes in business strategy that impact customer behavior. (FIG 4.3)

4) **Transactions**: Analyzing weekly business activity is valuable for businesses that experience variations in customer demand throughout the week. It provides insights that can be used to optimize staffing levels, adjust marketing strategies for specific days, and enhance overall operational efficiency. The plot allows stakeholders to understand the short-term dynamics of the business and make data-driven decisions at a more granular level. (FIG 4.4)

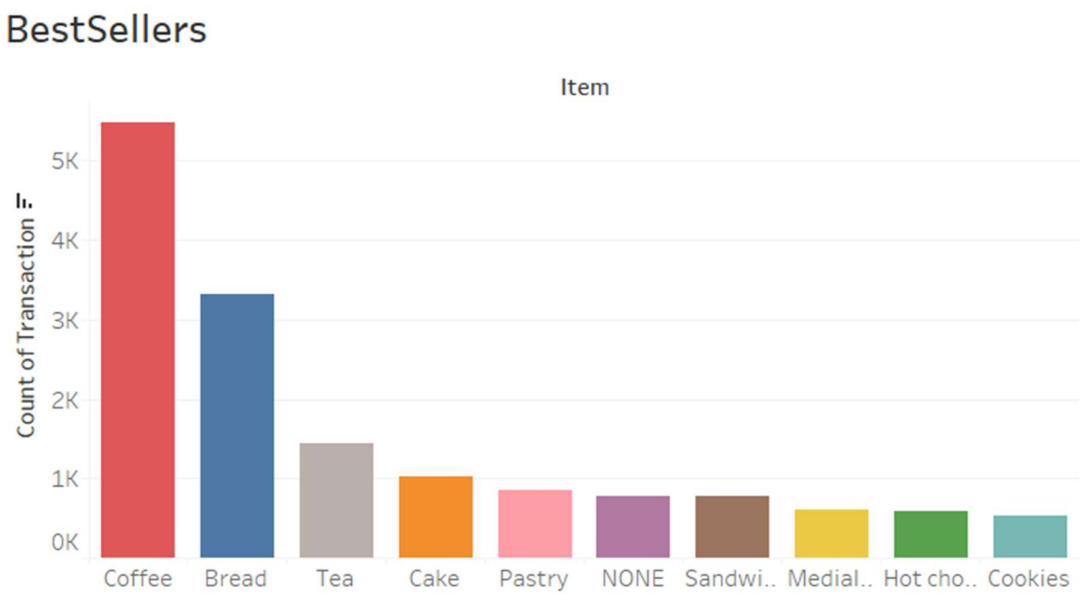

FIG 4.1: **BestSellers**

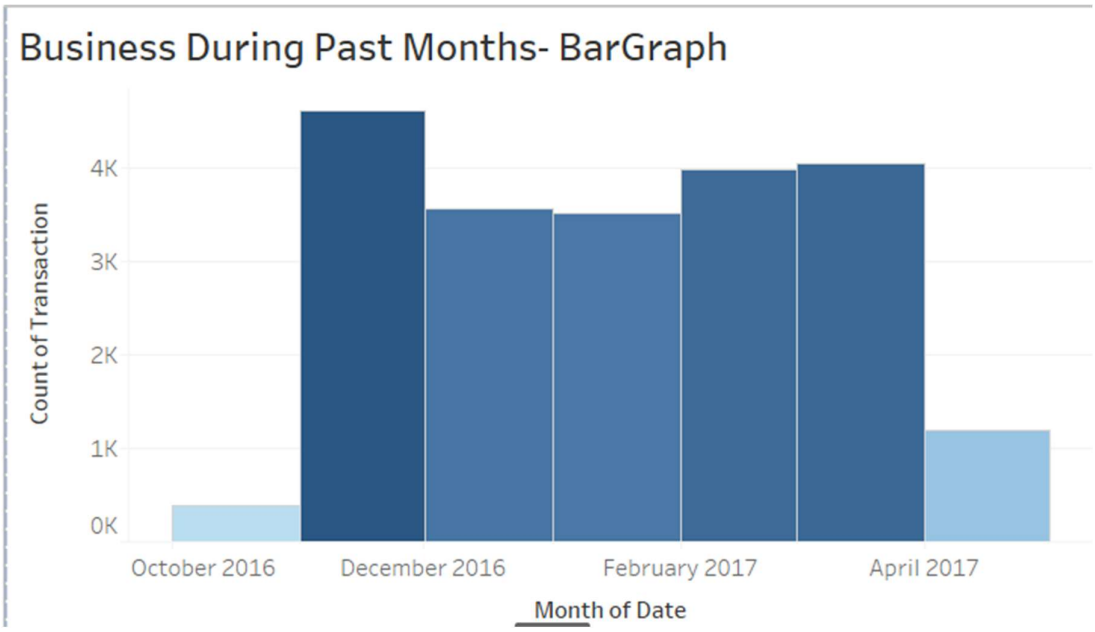

FIG 4.2: **Business During Past Months**

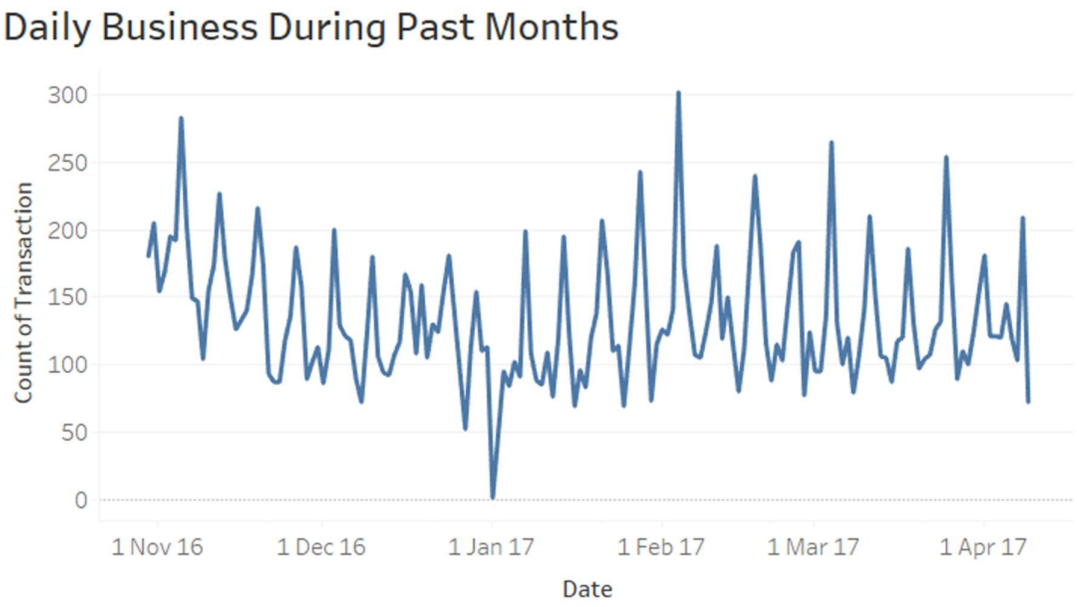

FIG 4.3: **Daily Business During Past Months**

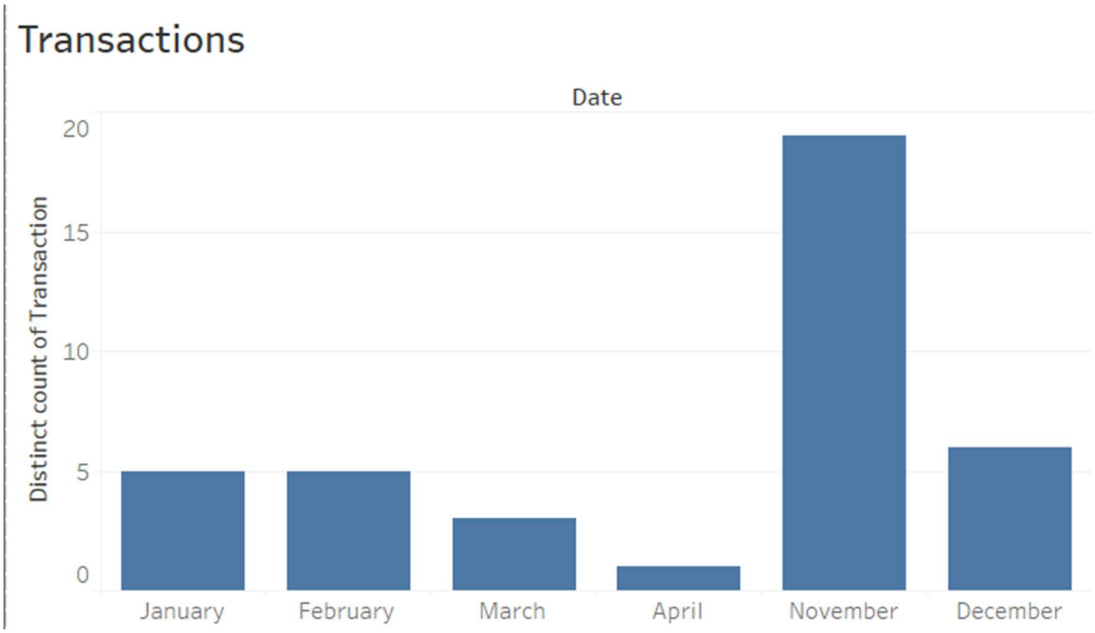

FIG 4.4: **Transactions**

CHAPTER 5

INTERACTIVE DASHBOARD USING TABLEAU

Creating an interactive dashboard for Market Basket Analysis in Tableau involves a systematic approach to leverage the platform's capabilities. Beginning with data preparation and connection, users ensure that the dataset is organized and relevant columns are present. Exploring and cleaning the data in Tableau follows, addressing any missing values or outliers. Calculated fields and time-based dimensions are then created to facilitate in-depth analysis. Utilizing Tableau's visualization features, Market Basket Analysis is visualized, often in the form of a heatmap displaying item associations in transactions. Filters are incorporated to focus on specific elements, and dashboards are designed for a coherent layout. Interactivity is enhanced through Tableau actions, allowing users to explore data dynamically. Additional insights and visualizations are created to provide a comprehensive overview. Customization of appearance, publication to Tableau Server or Tableau Public, and sharing the interactive dashboard with stakeholders conclude the process. Regular updates ensure the dashboard remains relevant and continues to deliver actionable insights from transactional data.

PROCEDURE:

Step 1: Prepare Your Data

Ensure that your dataset is in a suitable format for Tableau. It should include relevant columns such as Date, Time, Transaction ID, Item ID, and Item Name.

Step 2: Connect to Data in Tableau

Open Tableau Desktop.

Connect to your dataset by selecting the appropriate data source (e.g., CSV, Excel, database).

Step 3: Explore and Clean Data

Examine the data in Tableau to ensure it's correctly loaded.

Handle any missing values or outliers if necessary.

Step 4: Create Calculated Fields

Depending on your analysis goals, create calculated fields for metrics like Total Purchase Amount or Quantity per Transaction.

Step 5: Create Time-Based Dimensions

If applicable, create time-based dimensions like Month, Day of Week, or Hour from the Date and Time columns.

Step 6: Create Market Basket Analysis

Drag the Item ID or Item Name field to Rows.

Drag Transaction ID to Columns.

Tableau may automatically generate a table. Convert it to a heatmap by changing the chart type to "Square" or "Filled Map" under the "Show Me" menu.

Step 7: Create Filters

Use filters to focus on specific time periods, items, or transactions.

Create a filter for the top N items to analyze.

Step 8: Create Dashboards

Go to the "Dashboard" tab.

Drag sheets created in previous steps onto the dashboard.

Arrange and resize elements to create an organized layout.

Step 9: Add Interactivity

Use Tableau actions to enable interactivity between sheets.

Link filters and selections to update multiple sheets simultaneously.

Step 10: Create Insights and Visualizations

Create additional sheets for specific visualizations or insights (e.g., association rules, trends over time).

Add these sheets to the dashboard.

Step 11: Customize Appearance

Adjust formatting, colors, and fonts to make the dashboard visually appealing.

Add titles and captions to provide context.

Step 12: Publish to Tableau Server or Tableau Public

Save your Tableau workbook.

Publish it to Tableau Server if you have access, or save it to Tableau Public for online sharing.

Step 13: Share and Collaborate

Share the interactive dashboard link with relevant stakeholders.

Encourage exploration and collaboration using the interactive features.

Step 14: Update and Maintain

As your dataset evolves or new data becomes available, update your Tableau workbook to maintain relevance and accuracy.

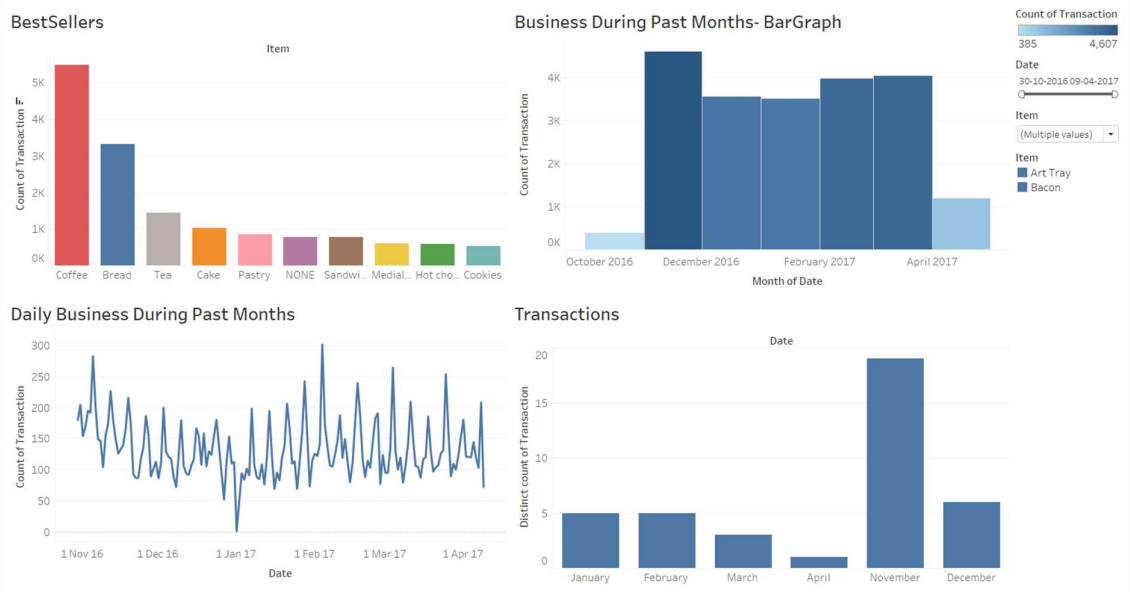

FIG 5.1: **Dashboard**

CHAPTER 6

CONCLUSION

In conclusion, the selection of an appropriate dataset for Market Basket Analysis is crucial for deriving meaningful insights into customer purchasing behavior. The dataset should ideally include key attributes such as transaction details (date, time, transaction ID), item details (item ID, item name), and potentially additional information like quantity or price. A well-prepared dataset allows for the application of advanced algorithms, such as the Apriori algorithm, to uncover patterns and associations between items purchased together. The insights gained from Market Basket Analysis can inform strategic decisions related to product placement, inventory management, and targeted marketing efforts. Additionally, the dataset's quality and completeness play a significant role in the accuracy of the analysis results. In conclusion, a well-curated dataset is the foundation for successful Market Basket Analysis, providing businesses with actionable intelligence to enhance customer satisfaction and drive overall business success.

FUTURE ENHANCEMENT:

The field of Market Basket Analysis (MBA) is dynamic, and continuous improvements and innovations can enhance its effectiveness. Here are some potential future enhancements for Market Basket Analysis:

1. Integration with Advanced Machine Learning Models:
Combine traditional MBA techniques with advanced machine learning models, such as deep learning, to uncover more complex patterns and relationships in large datasets.

2. Real-Time Analysis:
Develop real-time MBA capabilities to analyze transactions and update insights immediately. This is particularly valuable for industries where rapidly changing customer preferences drive business decisions.

3. Dynamic Pricing Strategies:
Integrate MBA insights into dynamic pricing strategies, allowing businesses to adjust pricing based on current item associations, demand patterns, and market conditions.

4. Customer Segmentation Refinement:
Enhance customer segmentation by incorporating additional variables such as customer demographics, preferences, and historical behavior. This can lead to more personalized marketing and product recommendations.

5. Visualization and Interpretability Tools:
Develop advanced visualization tools and interpretable models to help business stakeholders understand and trust the results of MBA, promoting wider adoption and implementation.

CHAPTER 7

REFERENCES

1) https://www.kaggle.com/datasets/aslanahmedov/market-basket-analysis/

2) https://www.google.com/

3) https://www.tableau.com/products/desktop

APPENDIX- A (CODE)

```
#Importing necessary modules
import numpy as np
import pandas as pd
%matplotlib inline
import matplotlib.pyplot as plt
import seaborn as sns
from mlxtend.preprocessing import TransactionEncoder
from mlxtend.frequent_patterns import apriori, association_rules
import networkx as nx
import warnings
warnings.filterwarnings('ignore')

data.shape

data.describe()

data.info()

data.loc[data['Item']=='NONE',:].count()

data=data.drop(data.loc[data['Item']=='NONE'].index)

data['Item'].nunique()

data['Item'].value_counts().sort_values(ascending=False).head(10)

fig, ax=plt.subplots(figsize=(6,4))

data['Item'].value_counts().sort_values(ascending=False).head(10).plot(kind='bar')

plt.ylabel('Number of transactions')

plt.xlabel('Items')

ax.get_yaxis().get_major_formatter().set_scientific(False)

plt.title('Best sellers')

data.loc[(data['Time']<'12:00:00'),'Daytime']='Morning'

data.loc[(data['Time']>='12:00:00')&(data['Time']<'17:00:00'),'Daytime']='Afternoon'

data.loc[(data['Time']>='17:00:00')&(data['Time']<'21:00:00'),'Daytime']='Evening'

data.loc[(data['Time']>='21:00:00')&(data['Time']<'23:50:00'),'Daytime']='Night'


fig, ax=plt.subplots(figsize=(6,4))

sns.set_style('darkgrid')

data.groupby('Daytime')['Item'].count().sort_values().plot(kind='bar')

plt.ylabel('Number of transactions')

ax.get_yaxis().get_major_formatter().set_scientific(False)

plt.title('Business during the day')


data.groupby('Daytime')['Item'].count().sort_values(ascending=False)
```

```
data['Date_Time']=pd.to_datetime(data['Date']+' '+data['Time'])
data['Day']=data['Date_Time'].dt.day_name()
data['Month']=data['Date_Time'].dt.month
data['Month_name']=data['Date_Time'].dt.month_name()
data['Year']=data['Date_Time'].dt.year
data['Year_Month']=data['Year'].apply(str)+' '+data['Month_name'].apply(str)
data.drop(['Date','Time'], axis=1, inplace=True)

data.index=data['Date_Time']
data.index.name='Date'
data.drop(['Date_Time'],axis=1,inplace=True)
data.head()

data4=data.pivot_table(index='Day',columns='Item', aggfunc={'Item':'count'}).fillna(0)
data4['Max']=data4.idxmax(axis=1)
data4

lst=[]
for item in data['Transaction'].unique():
    lst2=list(set(data[data['Transaction']==item]['Item']))
    if len(lst2)>0:
        lst.append(lst2)
print(lst[0:3])
print(len(lst))

te=TransactionEncoder()
te_data=te.fit(lst).transform(lst)
data_x=pd.DataFrame(te_data,columns=te.columns_)
print(data_x.head())

frequent_items= apriori(data_x, use_colnames=True, min_support=0.03)
print(frequent_items.head())

rules = association_rules(frequent_items, metric="lift", min_threshold=1)
rules.antecedents = rules.antecedents.apply(lambda x: next(iter(x)))
rules.consequents = rules.consequents.apply(lambda x: next(iter(x)))
rules
```

```
fig, ax=plt.subplots(figsize=(10,4))
GA=nx.from_pandas_edgelist(rules,source='antecedents',target='consequents')
nx.draw(GA,with_labels=True)
plt.show()
```

APPENDIX-B SCREENSHOTS

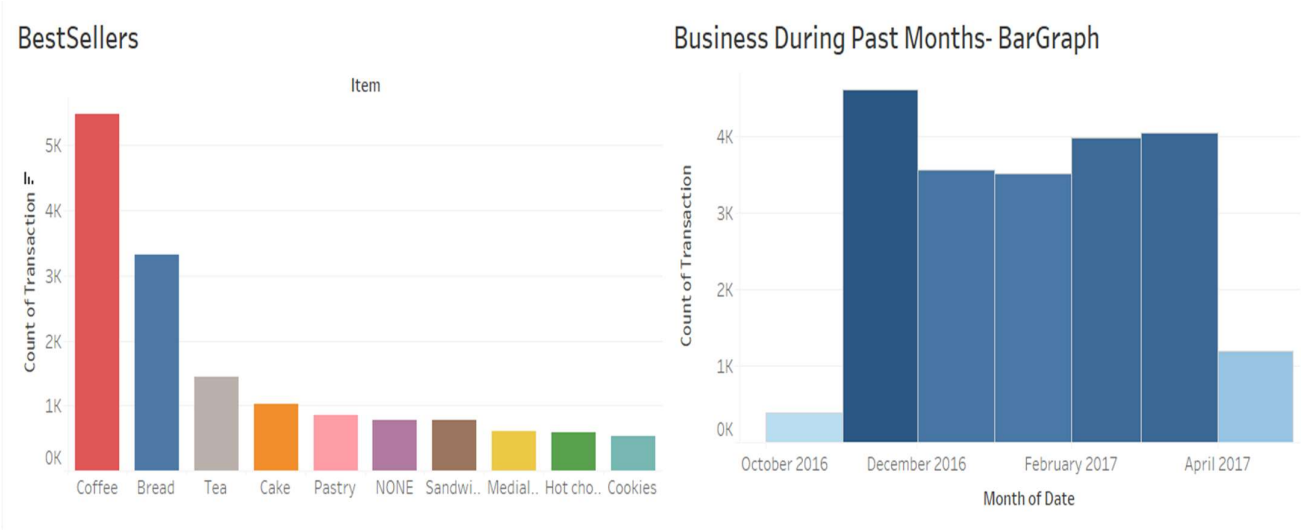

FIG: BESTSELEERS AND TRANSACTIONS PAST MONTH

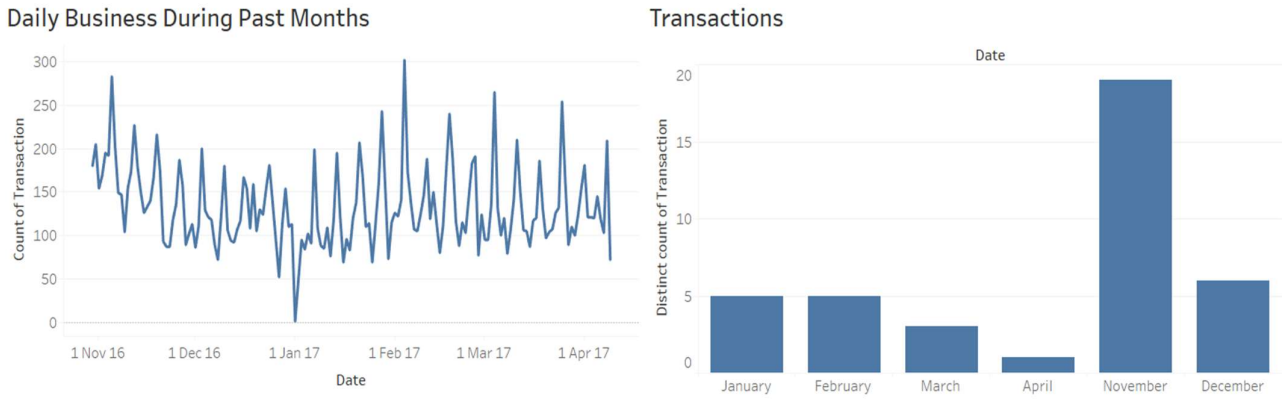

FIG: DAILY TRANSACTIONS