



KIET
GROUP OF INSTITUTIONS
Connecting Life with Learning

Assessment Report
on
“Market Basket Aisle Data Analysis”
submitted as partial fulfillment for the award of
BACHELOR OF TECHNOLOGY
DEGREE

SESSION 2024-25

in
By

Vinay Rathore(202401100300280)

Under the supervision of
“Abhishek Shukla Sir”
KIET Group of Institutions, Ghaziabad

Affiliated to
Dr. A.P.J. Abdul Kalam Technical University, Lucknow

INTRODUCTION

Overview:

Market Basket Analysis (MBA) is a widely used data mining technique that uncovers patterns in consumer purchasing behavior by identifying products that are frequently bought together. In this analysis, we focus on aisle data from retail environments, where aisles represent different product categories. By examining aisle visit patterns, retailers can gain valuable insights to optimize store layouts, promotional strategies, and inventory management.

Objective:

The main objectives of this project are:

- To analyze aisle data and identify the most frequently visited aisles.
- To visualize aisle distribution using bar and pie charts.
- To provide actionable insights that can help retailers enhance store operations, improve customer experience, and optimize product placement.

Project Description:

This project focuses on analyzing aisle data from retail transactions. Each record in the dataset represents an aisle visited during a transaction, and our goal is to explore:

- **Aisle Distribution:** Identify the top aisles based on frequency of visits.

- **Data Visualization:** Create bar and pie charts to effectively represent aisle popularity.
- **Insights for Retail Optimization:** Use the findings to suggest improvements in store layout, stocking decisions, and promotional strategies.

By understanding aisle visit patterns, retailers can improve customer engagement and drive sales growth through better product placement, targeted promotions, and efficient inventory management.

Methodology

This project follows a structured approach to analyze aisle visit data, helping retailers optimize operations and enhance customer engagement.

1. Data Collection

- **Dataset:** Contains transaction records with aisle visit data in CSV format.
- **Upload:** Dataset is uploaded into the working environment (e.g., Google Colab, Jupyter Notebook). 

2. Data Preprocessing

- **Loading:** Data is imported into a Pandas DataFrame for easy analysis.
- **Check:** We inspect the data for missing values using `df.head()` and `df.info()`. 

3. Data Exploration 🔎

- **Aisles:** We calculate the number of unique aisles to understand product categories.
- **Frequency:** Identify the most visited aisles using `value_counts()`.


4. Data Visualization 🎨

- **Bar Chart:** Visualize the top 10 aisles with a bar chart created using seaborn.

- **Pie Chart:** Show aisle distribution with a pie chart from matplotlib.


5. Insight Extraction💡

- **Patterns:** Identify high-traffic aisles and understand customer preferences.

- **Recommendations:** Suggest improvements in product placement and promotions based on insights.


6. Reporting and Conclusion📝

- **Summary:** A final report highlights key findings and actionable strategies.

- **Future Work:** Suggest further analysis, such as customer demographics or advanced techniques.


Code

```
# Step 1: Import required libraries
```

```
import pandas as pd
```

```
import matplotlib.pyplot as plt
```

```
import seaborn as sns
```

```
from google.colab import files
```

```
# Step 2: Upload dataset file from user
```

```
print("👉 Please upload your CSV file (e.g., Market Basket Aisle  
data)...")
```

```
uploaded = files.upload() # User uploads file
```

```
filename = list(uploaded.keys())[0] # Get the uploaded file name
```

```
# Step 3: Load the dataset into a DataFrame
```

```
df = pd.read_csv(filename)
```

```
# Step 4: Show confirmation and preview
```

```
print("\n✅ Dataset Loaded Successfully!")
```

```
print("\n📄 First 5 Rows of the Dataset:")
```

```
print(df.head()) # Show top 5 rows
```

```
# Step 5: Display basic information about the dataset  
  
print("\n🔍 Dataset Info:")  
  
print(df.info()) # Data types, nulls, etc.
```

Step 6: Explore the data - count unique aisles and top frequent ones

```
unique_aisles = df['aisle'].nunique()  
  
print(f"\n📊 Total Unique Aisles: {unique_aisles}")
```

```
print("\n📦 Top 10 Most Common Aisles:")  
  
top_aisles = df['aisle'].value_counts().head(10)  
  
print(top_aisles)
```

Step 7: Visualization - Bar chart of top 10 most common aisles

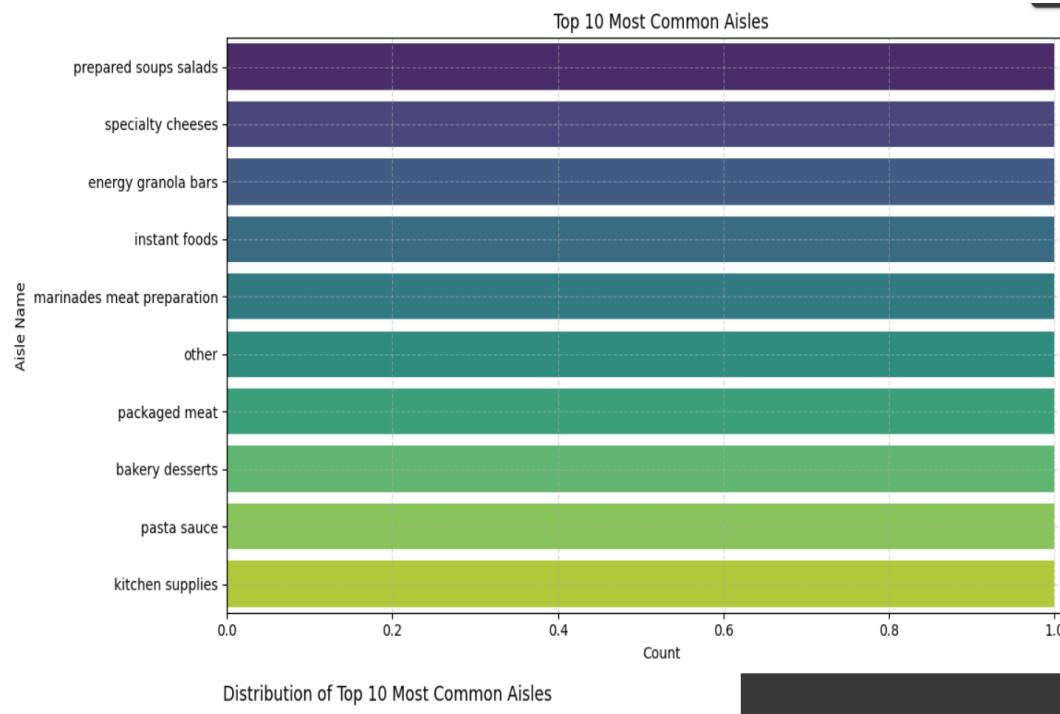
```
plt.figure(figsize=(12, 6)) # Set figure size  
  
sns.barplot(  
    y=top_aisles.index,  
    x=top_aisles.values,  
    palette='viridis'  
)
```

```
plt.title('Top 10 Most Common Aisles') # Title  
plt.xlabel('Count') # X-axis label  
plt.ylabel('Aisle Name') # Y-axis label  
plt.grid(True, linestyle='--', alpha=0.5) # Add grid for readability  
plt.tight_layout() # Adjust layout to avoid overlap  
plt.show()
```

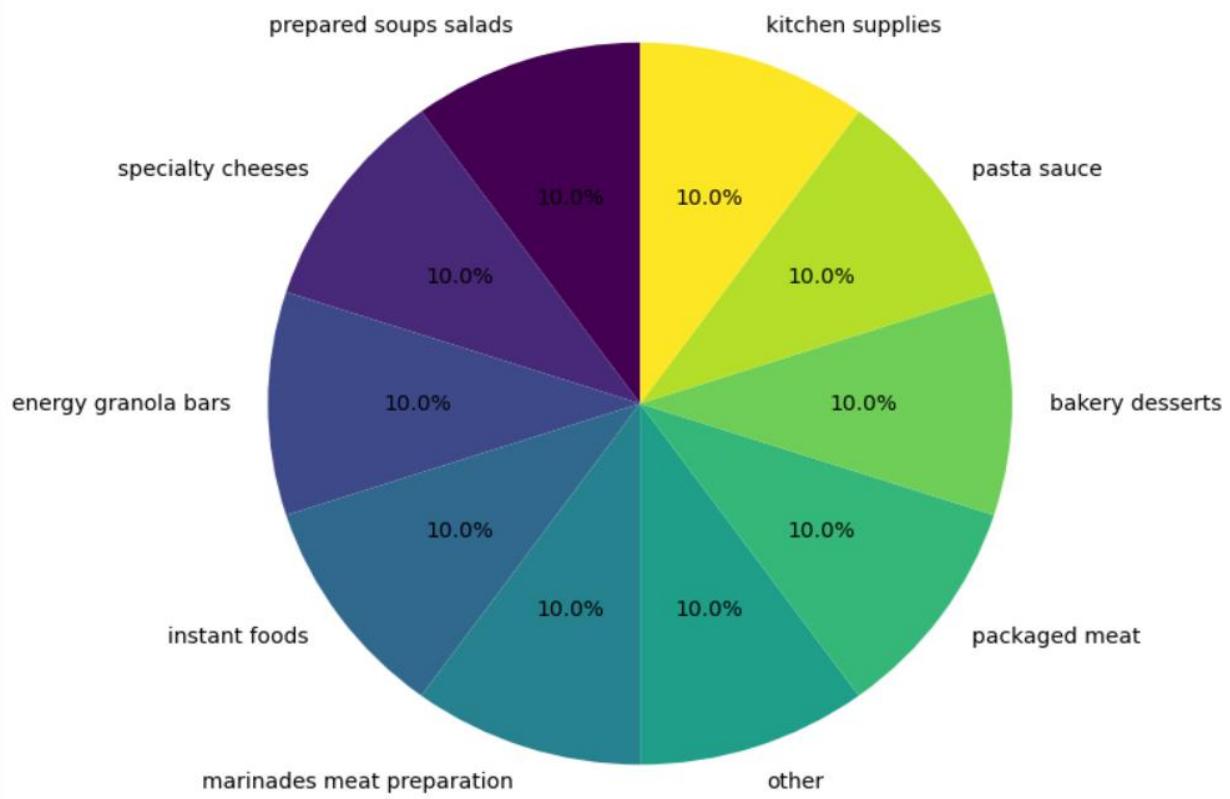
```
# Step 8: Visualization - Pie chart of top 10 most common aisles  
plt.figure(figsize=(8, 8))  
top_aisles.plot.pie(  
    autopct='%.1f%%',  
    startangle=90,  
    cmap='viridis',  
    legend=False  
)  
plt.title('Distribution of Top 10 Most Common Aisles')  
plt.ylabel("") # Remove the y-axis label  
plt.tight_layout() # Adjust layout to avoid overlap  
plt.show()
```

Output

```
Please upload your CSV file (e.g., Market Basket Aisle data)...  
Choose File: 10. Market ...Analysis.csv  
• 10. Market Basket Analysis.csv(text/csv) - 2603 bytes, last modified: 4/18/2025 - 100% done  
Saving 10. Market Basket Analysis.csv to 10. Market Basket Analysis (.).csv  
 Dataset Loaded Successfully!  
First 5 Rows of the Dataset:  
aisle_id      aisle  
0            1    prepared soups salads  
1            2    specialty cheeses  
2            3    energy granola bars  
3            4    instant foods  
4            5    marinades meat preparation  
Dataset Info:  
<class 'pandas.core.frame.DataFrame'>  
RangeIndex: 134 entries, 0 to 133  
Data columns (total 2 columns):  
 #   Column   Non-Null Count  Dtype    
 ---    
 0   aisle_id  134 non-null   int64    
 1   aisle     134 non-null   object   
dtypes: int64(1), object(1)  
memory usage: 2.2+ KB  
None  
Total Unique Aisles: 134  
Top 10 Most Common Aisles:  
aisle  
prepared soups salads      1  
specialty cheeses          1  
energy granola bars        1  
instant foods              1  
marinades meat preparation  1  
other                      1  
packaged meat              1
```



Distribution of Top 10 Most Common Aisles



References

1. Market Basket Analysis:

- Tan, P. N., Steinbach, M., & Kumar, V. (2005). *Introduction to Data Mining*. Pearson.
- Aggarwal, C. C. (2015). *Data Mining: The Textbook*. Springer.

2. Datasets:

- **Kaggle**: Retail Market Basket Dataset 
- **UCI Repository**: [Online Retail Dataset](#) 
- **Data.gov**: [Retail and Transaction Datasets](#) 

3. Visualization & Tools:

- **Matplotlib**: <https://matplotlib.org/> 
- **Seaborn**: <https://seaborn.pydata.org/> 
- **Pandas**: <https://pandas.pydata.org/> 

4. Additional Reading:

- Nguyen, M., & Tran, Q. (2018). "Data mining for retail market basket analysis." *Journal of Retailing and Consumer Services*, 42, 22-30.