جامعة أم القرى
UMM AL-QURA UNIVERSITY

REPORT
DATA ANALYSIS (2)
DS3114

Dr. Omaima Fallatah

Team Members:

- Faten Matouq Almowallad 444000266
- Ayah Bakur Alhawsawi 444006678

https://colab.research.google.com/drive/193dc74CvY2OJ3leKVXrPT-69u9AsCa3t

Tasks: 2

Market Basket Analysis

INTRODUCTION

Market Basket Analysis is a data mining technique used to discover patterns and relationships between items in transactional data. The goal is to identify associations between frequently purchased items to improve decision-making for promotions, product placement, and cross-selling strategies.

ABOUT DATASET

The dataset contains transaction records with attributes such as BillNo, Itemname, Quantity, and Country. Each row represents a unique item bought in a particular transaction. The task focuses on analyzing these transactions to discover meaningful item associations.

Data Preprocessing

- **Loading Data**: The dataset was loaded using Pandas, handling missing values and formatting issues by using different delimiters and skipping bad lines if necessary.
- **Handling Missing Values**: Rows with missing values were dropped to ensure data quality.
- **Transforming Data**: The transactional data was transformed into a format suitable for association rule mining, where each transaction was converted into a matrix of binary values indicating whether an item was purchased.

```
#Check Missing Values
print("Missing Values:")
print(data.isnull().sum()) # Changed df to data

#Drop Rows with Missing Values
data.dropna(inplace=True) # Changed df to data
```

```
Missing Values:
BillNo             0
Itemname        1351
Quantity           0
Date               0
Price              1
CustomerID    103787
Country            1
dtype: int64
```

```
[ ] description = data.describe()
    print(description)
```

```
            Quantity     CustomerID
count  295406.000000  295406.000000
mean       13.271907   15308.041462
std       143.508647    1723.880693
min         1.000000   12346.000000
25%         2.000000   13882.000000
50%         6.000000   15249.000000
75%        12.000000   16818.000000
max     74215.000000   18287.000000
```

ANALYSIS & RESULTS

- **Apriori Algorithm**: The Apriori algorithm was used to find frequent itemsets in the transaction data. Itemsets with a minimum support of 1% were considered, meaning that these itemsets appear in at least 1% of all transactions.

- **Association Rules**: From these frequent itemsets, association rules were generated to uncover relationships between items, with metrics like **support**, **confidence**, and **lift** used to evaluate the strength of these rules.

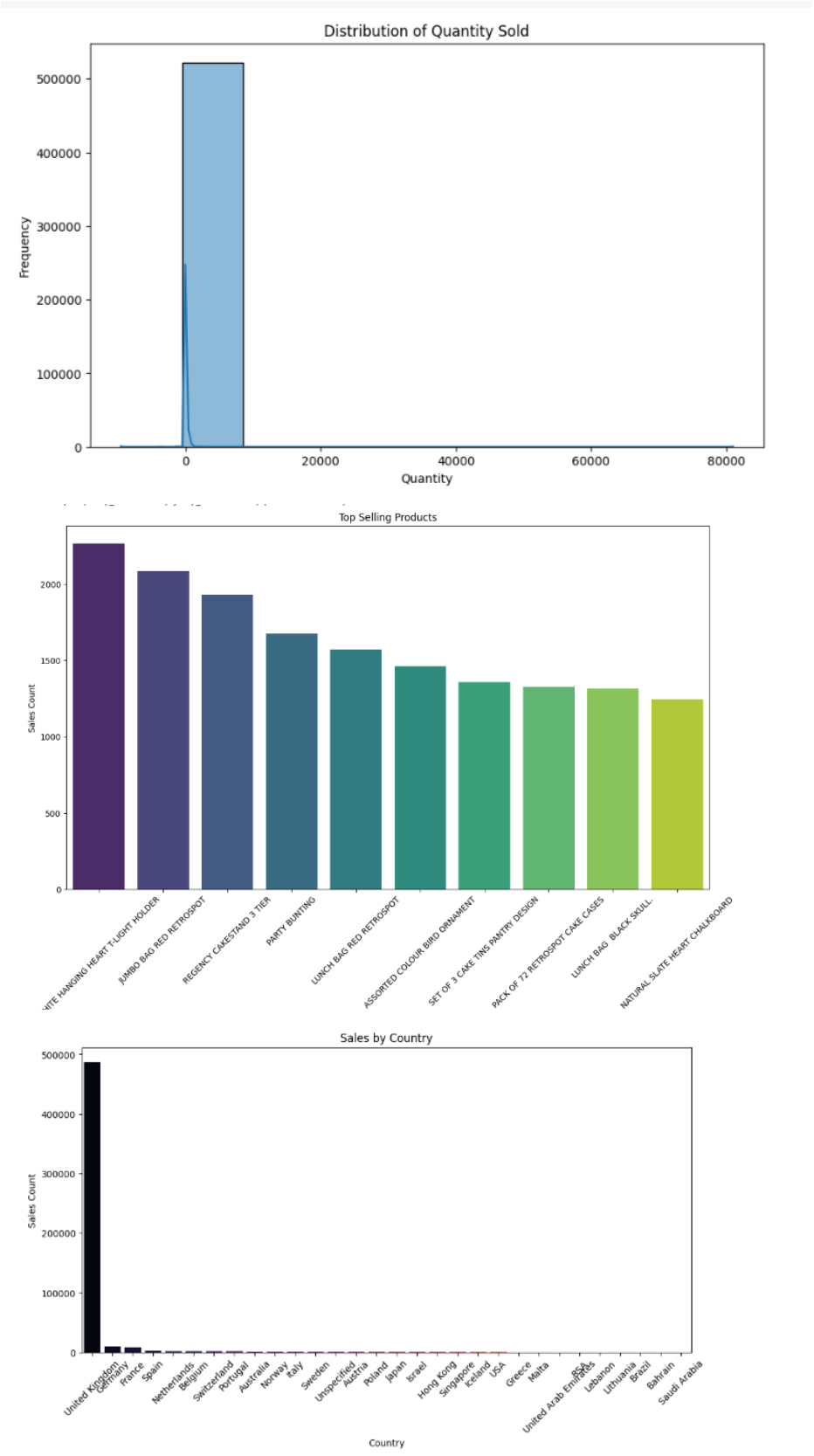
- **Visual Analysis**:

  - **Top Selling Products**: A bar plot of the top 10 most frequently purchased items highlighted key products driving sales.

  - **Quantity Sold**:

    - A distribution of the quantity sold per item revealed variations in purchase volumes.
    - A histogram of Quantity sold per item was created to visualize how often different quantities of items are sold.

- **Sales by Country**: Sales were also analyzed by country, showing how sales volume differs across regions.



CONCLUSION

Market Basket Analysis helped uncover valuable insights into customer purchasing behavior. The Apriori algorithm identified frequent itemsets and association rules, which can be used to optimize store layout, recommend products, or design promotional offers. Visual analyses of top-selling items and country-wise sales further helped in understanding the broader trends in the dataset.

REFERENCES

Ahmedov, A. (2021) *Market Basket Analysis*. Kaggle. Available at: https://www.kaggle.com/datasets/aslanahmedov/market-basket-analysis (Accessed: 25 October 2024).