data science

FINAL PROJECT

May 2024

Prepared by

ahmed shacker

Prepared for

Eng Huda Hemdan

Table of Contents

Introduction:

- Our project revolves around the development of a Data Analysis App using R Shiny. This interactive application allows users to upload CSV files containing transactional data and perform various analyses including data visualization, K-Means clustering, and Apriori association rule mining. The objective is to provide a user-friendly platform for exploring and analyzing data patterns.

Features:

1. Data Analysis Dashboard:
    - Users can upload CSV files.
    - Selection of various data visualization options like Payment Type Pie Chart, Age Group Box Chart, Distribution of Total Spending, and Total Spending by City.
2. K-Means Clustering:
    - Users can specify the number of clusters.
    - Runs K-Means clustering algorithm on uploaded data and displays clustered results.
3. Apriori Association Rule Mining:
    - Users can set support and confidence thresholds.
    - Performs association rule mining using the Apriori algorithm and displays the generated rules.

Technical Details:

1. **Libraries:**
    - The application utilizes several R packages including dplyr, arules, ggplot2, and shiny for data manipulation, visualization, and interactivity.
2. **User Interface (UI):**
    - The UI is designed using Shiny's UI functions, featuring a navigation bar with different tabs for each analysis module.
3. **Data Handling:**
    - Upon file upload, the application reads the CSV data, removes duplicates, and makes it available for analysis using reactive expressions.

Technical Details:

3-Data Visualization:

- Various plots are generated based on user selections using ggplot2. These include pie charts, bar charts, boxplots, and histogram for visualizing different aspects of the data.

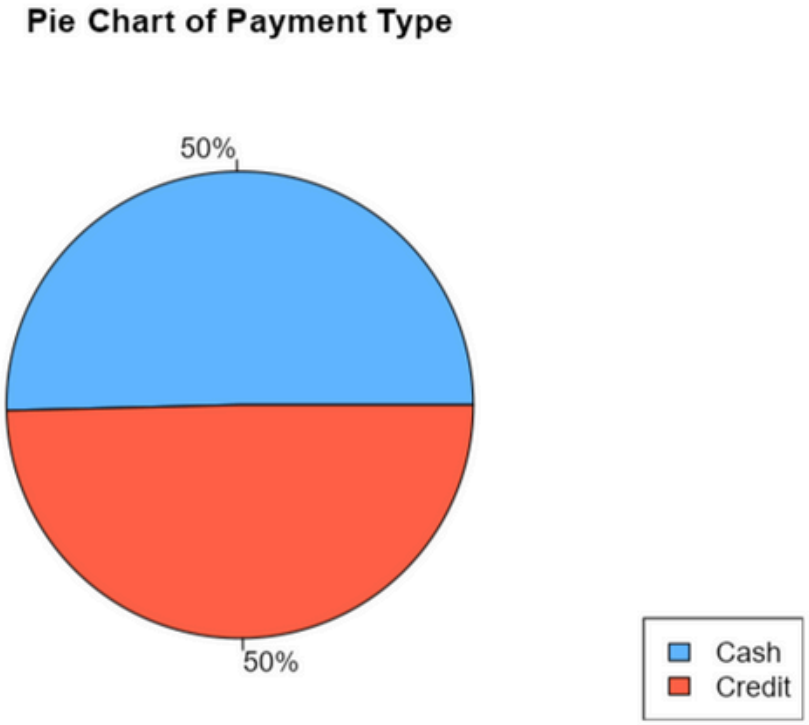
4-K-Means Clustering:

- K-Means clustering is performed on the uploaded data based on user-defined parameters.
- Results are displayed in a table showing clustered data points along with customer information, age, total spending, and assigned cluster.

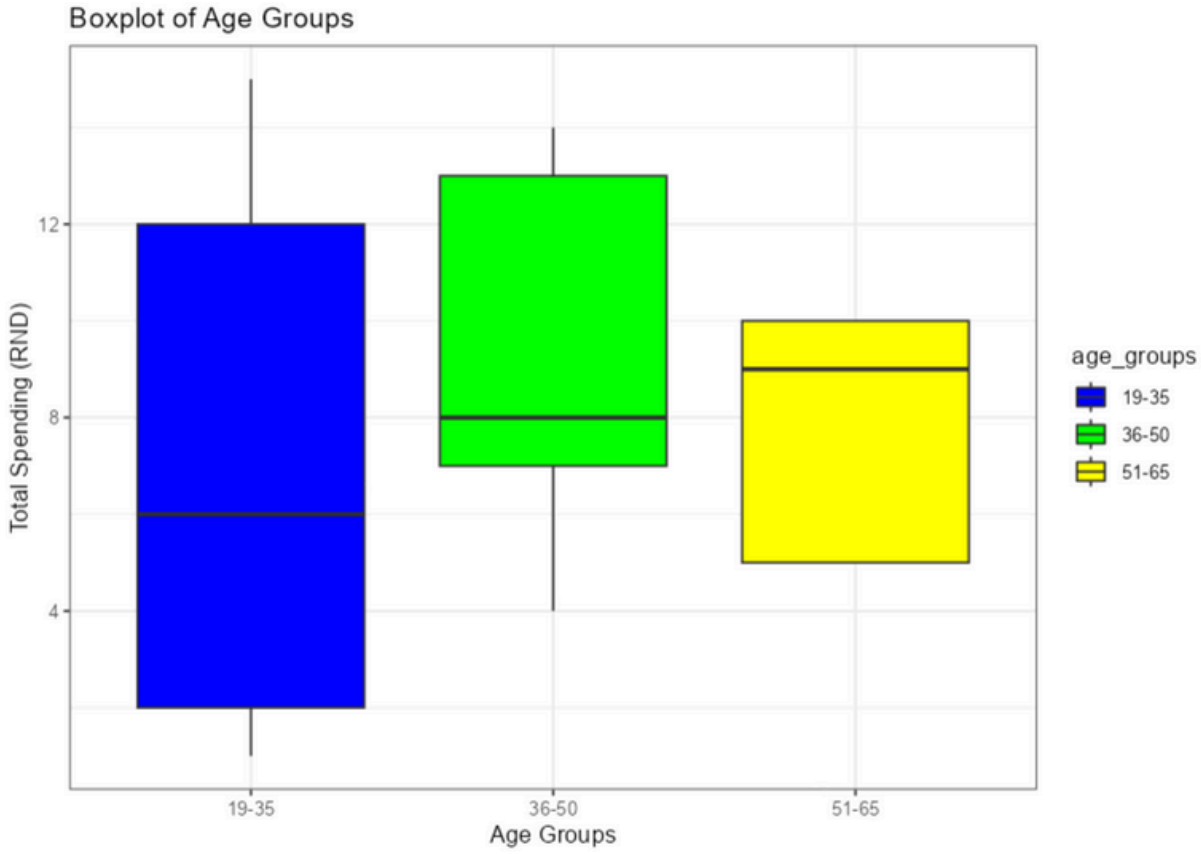
5-Apriori Association Rule Mining:

- The application conducts association rule mining using the Apriori algorithm based on user-defined support and confidence levels.
- The generated association rules are displayed in a table format.

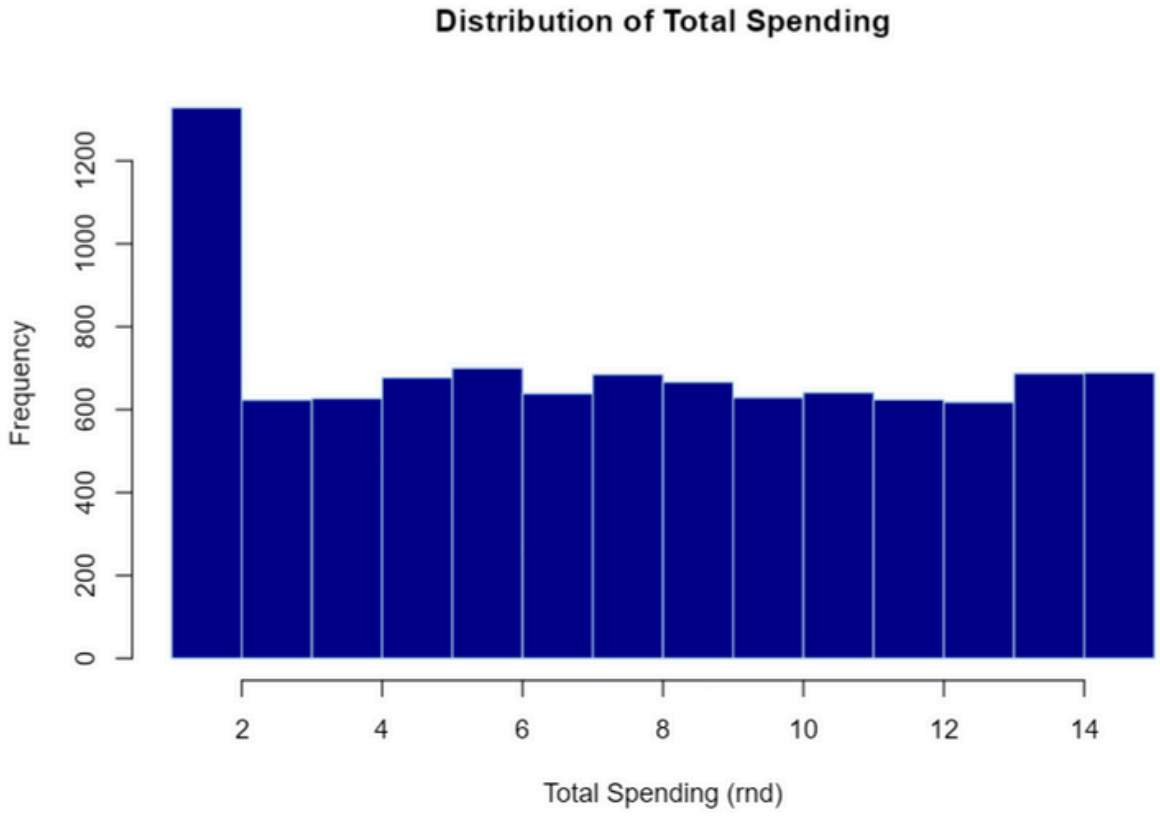
graphs:
1-pie chart



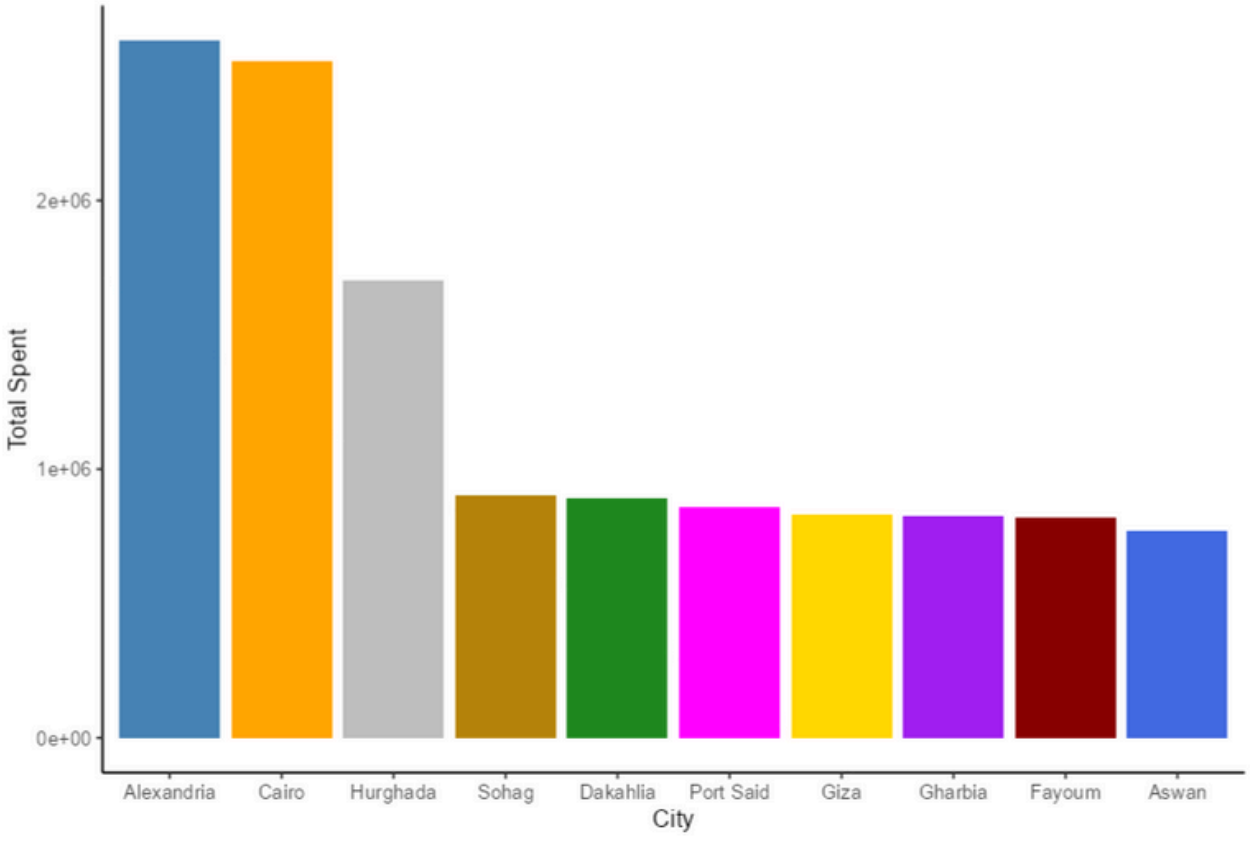
2-box plot



3-scatter plot



4-barplot



conclusion

In conclusion, our Data Analysis App offers a comprehensive platform for users to explore and analyze their data through intuitive visualizations, clustering analysis, and association rule mining. With its user-friendly interface and powerful analytical capabilities, the application serves as a valuable tool for gaining insights into transactional datasets.