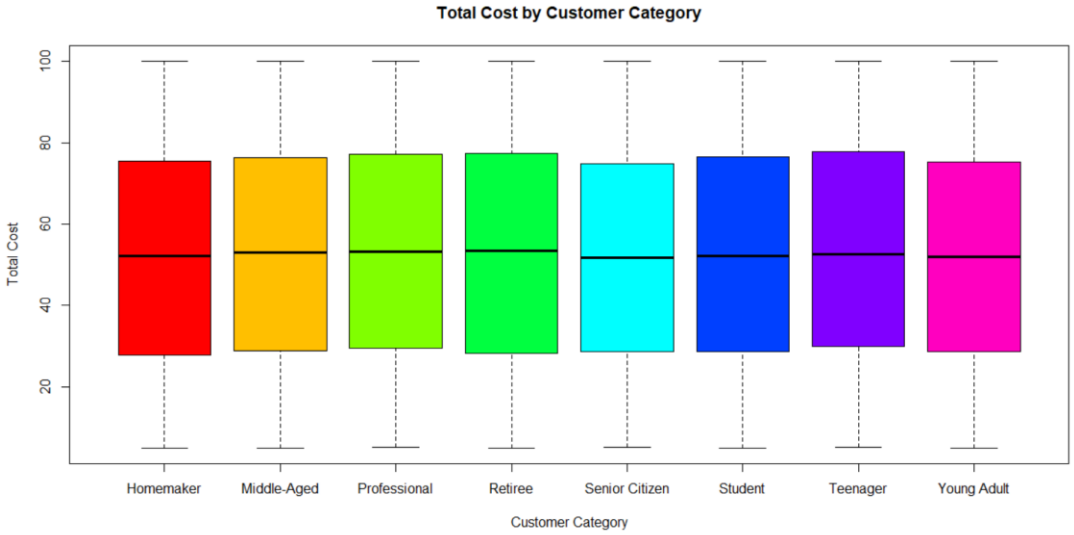
Responsi – Data Mining – CG24-2

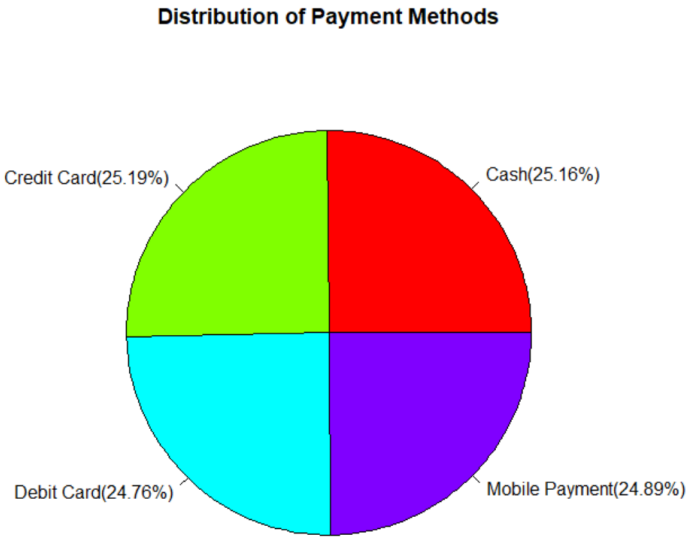
1. **Data Visualization**

   To describe the data better, you are asked to visualize the data in graph form. The data given should be named as **"retail.csv"**. Some data that needed to be visualized are:

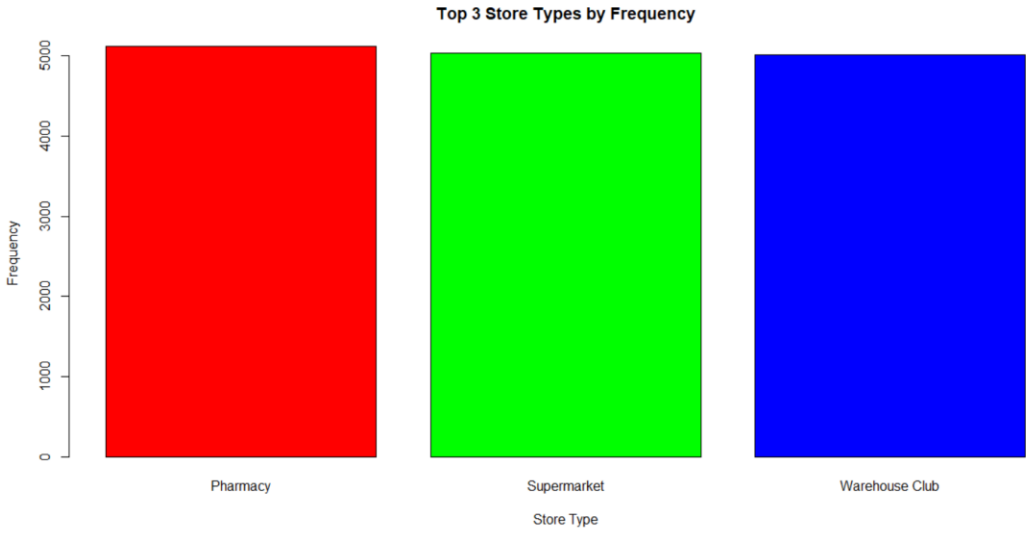
   a. **How does the total cost of purchases vary across different customer categories?**



   b. **What is the distribution of payment methods used by customers?**



c. **What are the top 3 popular store types where purchases are made?**



2. **Frequent Pattern Analysis**

You are asked to do frequent pattern analysis to know the **frequent product** that the people bought. To get the frequent product, "**market_data.csv**" and follow all steps below:

a. **Data Preprocessing**

In this phase, some data can't be used for further analysis. Do the following task to **cleanse** the data:

- Process the data that only hails from the department of **"beverages"**
- Remove all **product** which aisle is **"soft drinks"**
- Remove all **duplicated** data for the analysis

b. **Data Transformations**

In this phase, you need to change the data, so it is suitable to be used in the Apriori analysis. Prepare the product data in terms of the product's name.

c. **Data Mining**

- Show frequent product using Apriori algorithm with **minimum support: 0.02** based on the data that have already pre-processed

```
    items                                 support    count
[1] {Italian Sparkling Mineral Water} 0.02098951  42
[2] {Original Orange Juice}           0.02048976  41
[3] {100% Raw Coconut Water}          0.02148926  43
[4] {Sparkling Natural Mineral Water} 0.03048476  61
[5] {Spring Water}                    0.04147926  83
[6] {Sparkling Lemon Water}           0.02898551  58
[7] {Lime Sparkling Water}            0.03748126  75
[8] {Sparkling Water Grapefruit}      0.06396802 128
```

- Show the association rules using minimum **confidence: 0.05** based on the frequent product that resulted from step above.

```
    lhs    rhs                          support    confidence coverage lift count
[1] {}  => {Sparkling Water Grapefruit} 0.06396802 0.06396802 1        1    128
```