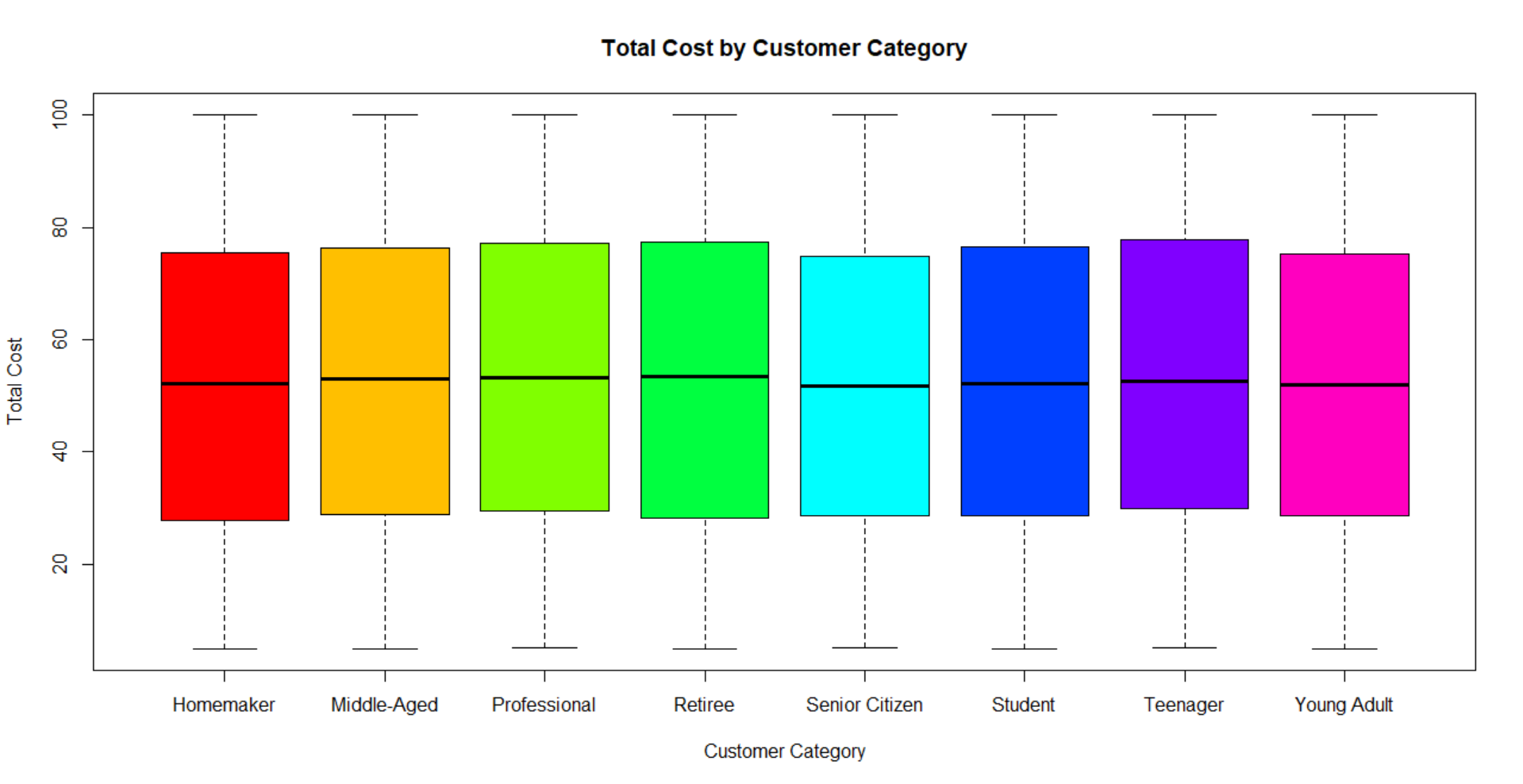
**Responsi – Data Mining – CG24-2**

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:

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

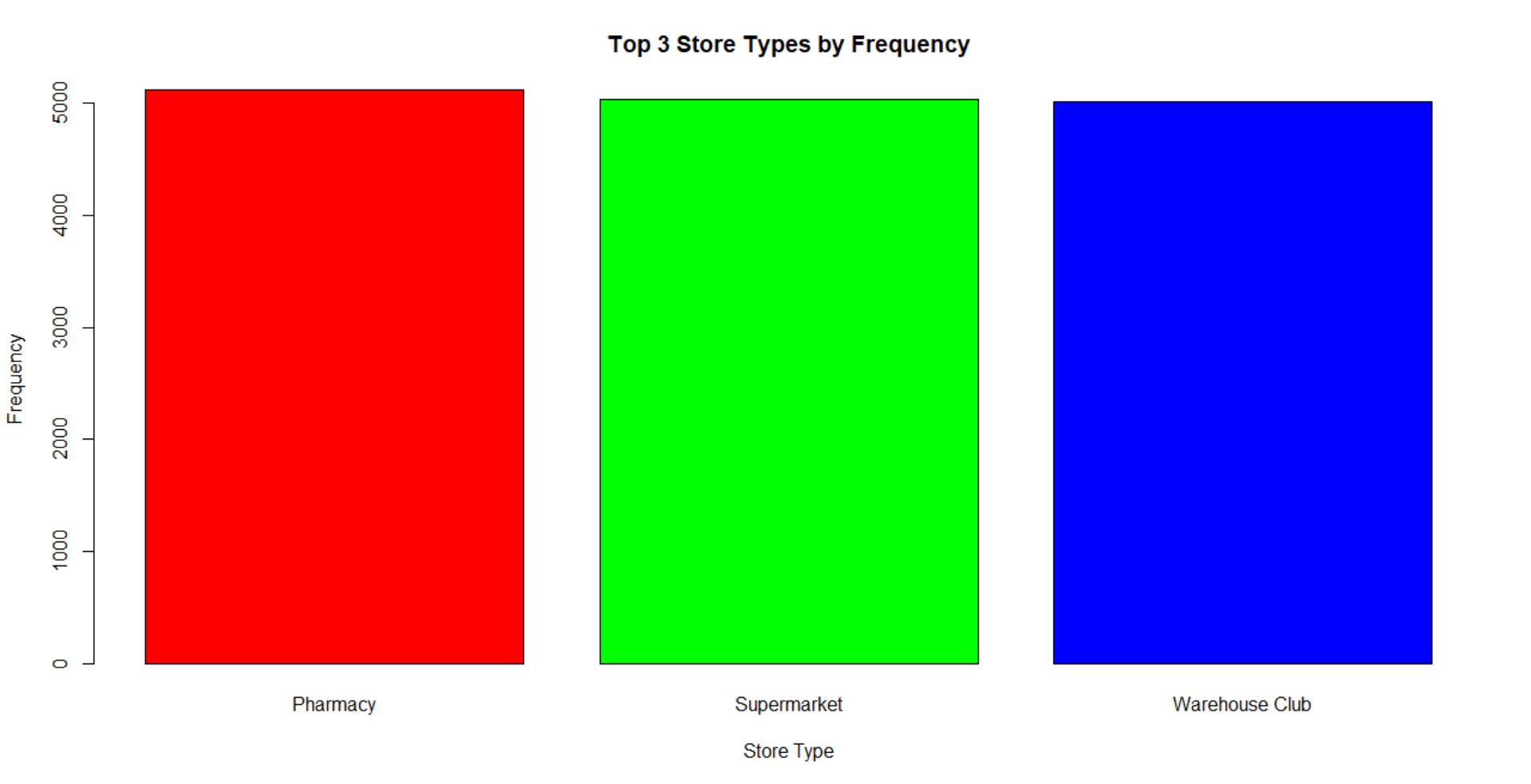


1. **What is the distribution of payment methods used by customers?**

A diagram of a distribution of payment methods

Description automatically generated

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



1. **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:

1. **Data Preprocessing**

In thisphase, 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”**
* Removeall **product** which aisle is **“soft drinks”**
* Removeall **duplicated** data for the analysis

1. **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.

1. **Data Mining**

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

A screenshot of a computer program

Description automatically generated

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

