Data Analysis and Visualization in R Programming

Student Details

Course Details

Institution

Instructor

Deadline

Data Analysis and Visualization in R Programming

Introduction

      This report analyzes and visualizes the sales data of a small online store in my organization. Using the (product.csv) in-house dataset, we find various data points like categories, sales, and revenue. The aim of this study is to understand what happened in the store over a certain period and create visual illustrations of how well the store performed.

Data analysis

Data import and summary statistics

      First, we should import the data and do some basic exploratory analysis. We use the following R code to load the dataset and display summary statistics (Wiley & Wiley, 2019).

```
# Load the data
data <- read.csv("product.csv")


# Summary statistics
summary(data)
```

Code

```
   Product            Category              Quantity           Revenue
 Length:10          Length:10          Min.   :20.00      Min.    : 400
 Class :character   Class :character   1st Qu.:31.25      1st Qu.: 725
 Mode  :character   Mode  :character   Median :42.50      Median :1200
                                       Mean   :43.50      Mean    :1340
                                       3rd Qu.:57.50      3rd Qu.:1725
                                       Max.   :70.00      Max.    :2800
```

Output

      Summary statistics summarize important features of the dataset like minimum, maximum, and mean, as well as the first and third quartiles for the 'quantity' and 'revenue' columns (Kaya, 2019).

Data exploration

We look at the data by studying product categories and how they were dispersed. We used the following code to generate a table of category frequencies.

```
# Count of products in each category
table(data$Category)
```

<div align="center">Code</div>

```
    Books     Clothing    Electronics  Home & Kitchen        Toys
      2            2            2            2               2
```
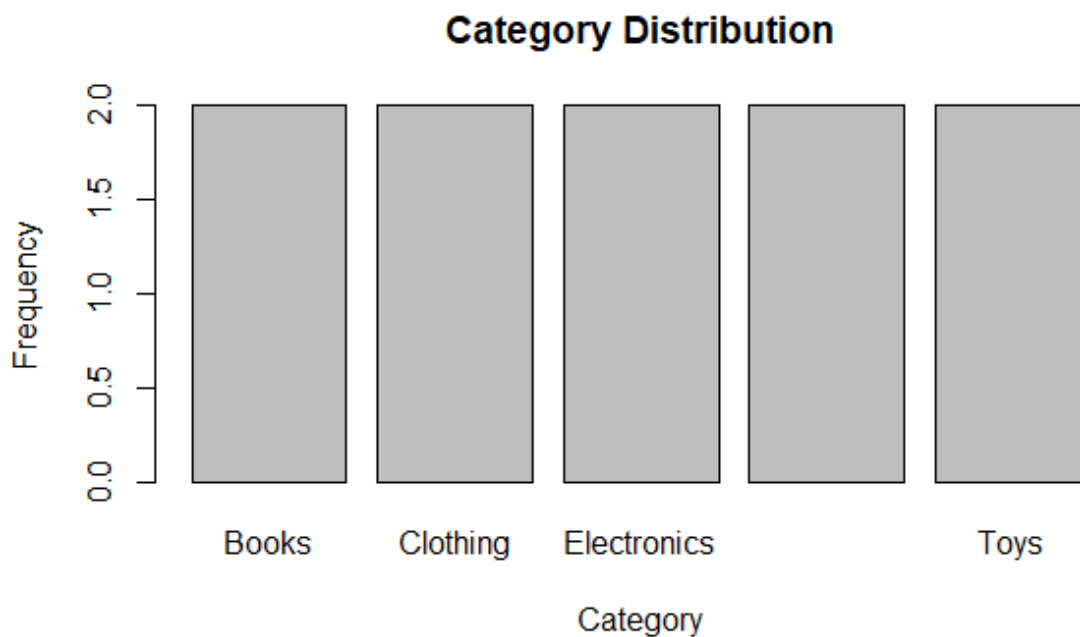
<div align="center">Output</div>

Data visualization

Category distribution

Category distribution bar plot

barplot(table(data$Category), main = "Category Distribution", xlab = "Category", ylab = "Frequency")
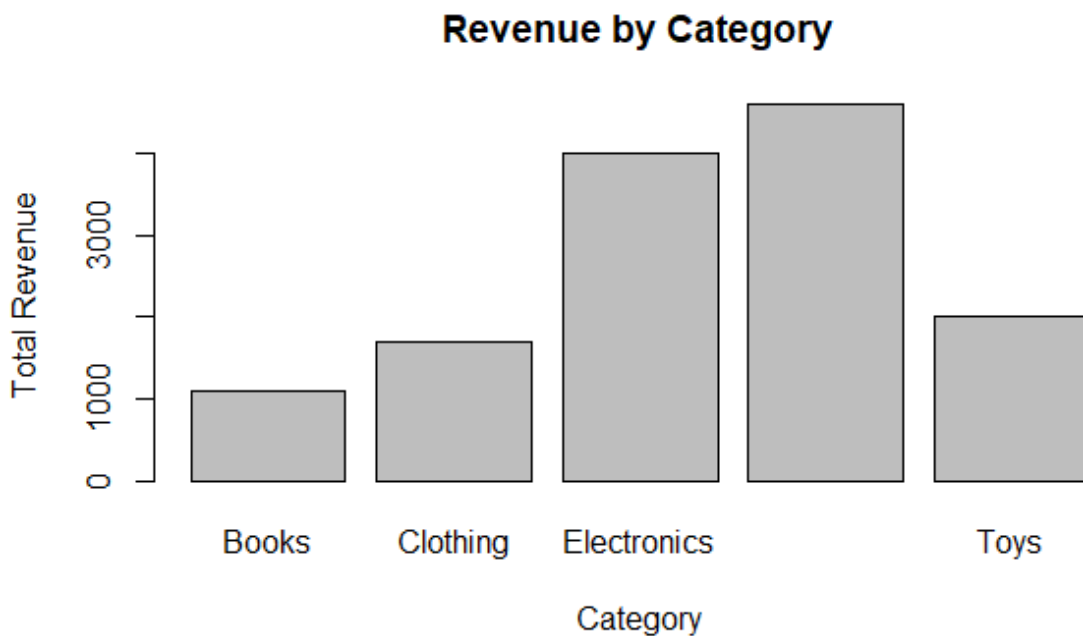
<div align="center">code</div>

Category Distribution

Revenue by category

We created a bar chart showing the total revenue for each category to visualize revenue by category.

Revenue by category bar plot

barplot(tapply(data$Revenue, data$Category, sum), main = "Revenue by Category", xlab = "Category", ylab = "Total Revenue")

code

Revenue by Category

Conclusion

These demonstrate how descriptive statistics were carried out based on a sample sales dataset. The use of R enabled us to import, explore, and visualize the data, and this gave us insights into the product category distribution and its related revenue. They help us to examine the business and make informed business decisions.

It is important to analyze and represent data as information for making sense of the data collected, with R having been an effective tool in undertaking this process. This report

shows that R skills for data analysis and visualization allow one to make the right decision based on the information.

References

Kaya, E., Agca, M., Adiguzel, F., & Cetin, M. (2019). Spatial data analysis with R

    programming for environment. *Human and ecological risk assessment: An*

    International Journal, *25*(6), 1521-1530.

Wiley, M., & Wiley, J. F. (2019). *Advanced R statistical programming and data models:*

    Analysis, machine learning, and visualization. Apress.