9

library(readr)

library(dplyr)

library(ggplot2)

#Load the Dataset

purchase\_data <- read\_csv("D:\\R Lab 2024\\customer\_purchases.csv")

#Data Summary

cat("Total Number of Records:", nrow(purchase\_data), "\n")

cat("Total Number of Unique Customers:", n\_distinct(purchase\_data$CustomerID), "\n")

#Calculate Statistical Measures

Mean\_Purchase = mean(purchase\_data$PurchaseAmount)

Median\_Purchase = median(purchase\_data$PurchaseAmount)

SD\_Purchase = sd(purchase\_data$PurchaseAmount)

cat("MEan is: ", Mean\_Purchase)

cat("MEdian is: ", Median\_Purchase)

cat("sd is: ", SD\_Purchase)

# Define the threshold value

threshold <- 511

# Categorize spending based on the threshold

purchase\_data$newcol <- ifelse(

purchase\_data$PurchaseAmount < threshold,

"Low Spender", # Label for values less than 511

"High Spender" # Label for values 511 or more

)

print(purchase\_data)

#Visualize Data (Histogram)

ggplot(purchase\_data, aes(x = PurchaseAmount)) +

geom\_histogram(color = "black") +

labs(title = "Distribution of Purchase Amounts", x = "Purchase Amount", y = "Frequency")

#Visualize Relationship (Scatter Plot)

ggplot(purchase\_data, aes(x = CustomerID, y = PurchaseAmount)) +

geom\_point(color = "green") +

labs(title = "Customer Purchase Amounts", x = "Customer ID", y = "Purchase Amount")