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")
```