Practical 1

**Code:**

**#Step 1: Loading and Exploring the Data** # Load necessary libraries library(tidyverse)

# Load the dataset

superstore\_data <- read.csv("D:/MSc DS/Semester 1/Retail Market Analysis/Practical/superstore\_data.csv")

**#Step 2: Cleaning and Preprocessing**

# Check for missing values

missing\_values <- colSums(is.na(superstore\_data))

# Replace missing values with the mean of each column for (col in names(superstore\_data)) {

if (sum(is.na(superstore\_data[[col]])) > 0 & is.numeric(superstore\_data[[col]])) { superstore\_data[[col]][is.na(superstore\_data[[col]])] <- mean(superstore\_data[[col]], na.rm = TRUE)

}}

# Remove duplicates or irrelevant columns if necessary.

selected\_data <- superstore\_data[, c("Year\_Birth", "Marital\_Status", "Education", "Dt\_Customer", "Recency", "NumStorePurchases", "NumWebPurchases", "NumWebVisitsMonth")]

selected\_data <- unique(selected\_data)

#Convert data types if needed (e.g., converting a column to the correct date format). selected\_data$Dt\_Customer <- as.Date(selected\_data$Dt\_Customer, format = "%Y-%m-%d") selected\_data$Year\_Birth <- as.numeric(as.character(selected\_data$Year\_Birth))

**#Step 3: Visualization with Histogram Plot**

# Simple histogram plot for a numerical variable (e.g., 'age')

selected\_data$Age <- as.integer(format(Sys.Date(), "%Y")) - selected\_data$Year\_Birth hist(selected\_data$Age,

main = "Histogram of Age", xlab = "Age",

col = "salmon")

**#Step 4: Tabulation and Summary** # Summary statistics summary(superstore\_data)

# Tabulation for categorical variables table(selected\_data$NumWebPurchases) table(selected\_data$NumStorePurchases)

### Interpretation and Implications:

Histogram Visualization: Analysing the distribution of age provides insight into the target audience's age range, aiding in targeted marketing campaigns. The majority of the target audience falls between 40-60.

Tabulation and Summary: Understanding the distribution of categorical variables like Web purchases/Store Purchases and summary statistics of numerical features such as income, age, etc helps identify customer demographics. Insights into instore and web purchase levels provide indications of purchasing power and potential market segments is higher for instore purchase than web purchases.