Walmart Store Sales Data Project

Objective:

Analyse Walmart store sales data to practice R programming skills, including data manipulation, statistical analysis, and data visualization.

Dataset:

- File: `walmart.csv`

Instructions:

Step 1: Setting Up the Environment

1. Install and load the necessary packages:

```
install.packages("tidyverse")
install.packages("summarytools")
install.packages("ggplot2")
```

```
library(tidyverse)
library(summarytools)
library(ggplot2)
```

2. Read the dataset:

- Load the CSV file using `read.csv()`:

```
walmart.csv <- read.csv("path\\Walmart.csv")
```

- Preview the first few rows using `head()`:

```
head(Walmart.csv)
```

Step 2: Data Exploration

1. Summarize the dataset:

- Generate a summary using `dfSummary()`:

```
dfSummary(walmart.csv)
```

2. Check for missing values:

- Identify missing values using `summary(is.na())`:

```
summary(is.na(walmart.csv))
```

Step 3: Statistical Analysis

1. Descriptive Statistics:

- Calculate mean, median, and standard deviation for key financial columns (e.g., Sales, Weekly_Sales, etc.):

```
mean(walmart.csv$Weekly_Sales, na.rm = TRUE)
median(walmart.csv$Weekly_Sales, na.rm = TRUE)
sd(walmart.csv$Weekly_Sales, na.rm = TRUE)
```

2. Correlation Analysis:

- Create a correlation matrix for key metrics (e.g., Weekly_Sales, Temperature, Fuel_Price):

```
cor(walmart.csv[, c("Weekly_Sales", "Temperature", "Fuel_Price")], use = "complete.obs")
```

Step 4: Data Visualization

1. Histogram:

- Create a histogram to visualize `Weekly_Sales`:

```
ggplot(walmart.csv, aes(x = Weekly_Sales)) +
  geom_histogram(binwidth = 5000, fill = "blue", color = "black")
```

2. Boxplot:

- Generate a boxplot for `Weekly_Sales`:

```
ggplot(walmart.csv, aes(y = Weekly_Sales)) +
  geom_boxplot()
```

3. Scatter Plot:

- Create a scatter plot for `Temperature` vs `Weekly_Sales`:

```
ggplot(walmart.csv, aes(x = Temperature, y = Weekly_Sales)) +
  geom_point() +
  geom_smooth(method = "lm", se = FALSE)
```

4. (Optional) Time Series Plot:

- If the dataset includes date information, create a time series plot for `Weekly_Sales` over time:

```
ggplot(walmart.csv, aes(x = Date, y = Weekly_Sales)) +
  geom_line() +
  scale_x_date(date_labels = "%b %Y") +
  labs(title = "Weekly Sales Over Time")
```