```
#                   ••• Walmart Store Sales Data Project •••


#   Objective: To analyse Walmart store sales data to practice R programming skills,
including data manipulation, statistical analysis, and data visualization.

#   Dataset: File - `walmart.csv`




#   1: Setting Up the Environment.

#   Installing the packages:

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

#   Loading the packages:

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



#   Reading the dataset by loading the CSV file from the local drive path:

walmart.csv <- read.csv("/Users/akheil/Downloads/Just_It - Data
Bootcamp/R/Walmart.csv")


#   Previewing the first few rows using 'head()':

head(walmart.csv)




#   2: Data Exploration.


#   Summarizing the dataset:

dfSummary(walmart.csv)


#   Checking for missing values:

summary(is.na(walmart.csv))




#   3: Statistical Analysis.


#   Descriptive Statistics:

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

#   Or,
descr(walmart.csv)
```

```
#   Correlation Analysis:

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




#   4: Data Visualization.


#   a). Histogram for Weekly Sales:

ggplot(walmart.csv, aes(x = Weekly_Sales)) +
  geom_histogram(binwidth = 95000, fill = "blue", color = "black")


#   b). Scatter Plot for 'Temperature' Vs. 'Weekly Sales':

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


#   c). (Optional) Time Series Plot:

ggplot(walmart.csv, aes(x = Date, y = Weekly_Sales)) +
   geom_line() + labs(title = "Weekly Sales Over Time")


#   d). Scatter Plot for 'Temperature' Vs. 'Fuel Price':

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



#   Creating a separate sub-table called 'Store_Weekly_Sales' that shows the Average of
Weekly Sales made my each Store (ie. Store no. from 1 to 45):

Store_Weekly_Sales <- walmart.csv %>%
  group_by(Store) %>%
  summarize(Avg_Weekly_Sales = mean(Weekly_Sales))


#   e). Line plot  displaying the Average of Weekly Sales made my each Store (ie. Store
no. from 1 to 45):

ggplot(data = Store_Weekly_Sales, aes(x = Store, y = Avg_Weekly_Sales, group = 1)) +
  geom_line() +
  geom_point() +
  labs(title = "Average Weekly_Sales by Store", x = "Store (Store no. from 1 to 45)", y
= "Average Weekly Sales") +
  theme_minimal()
```