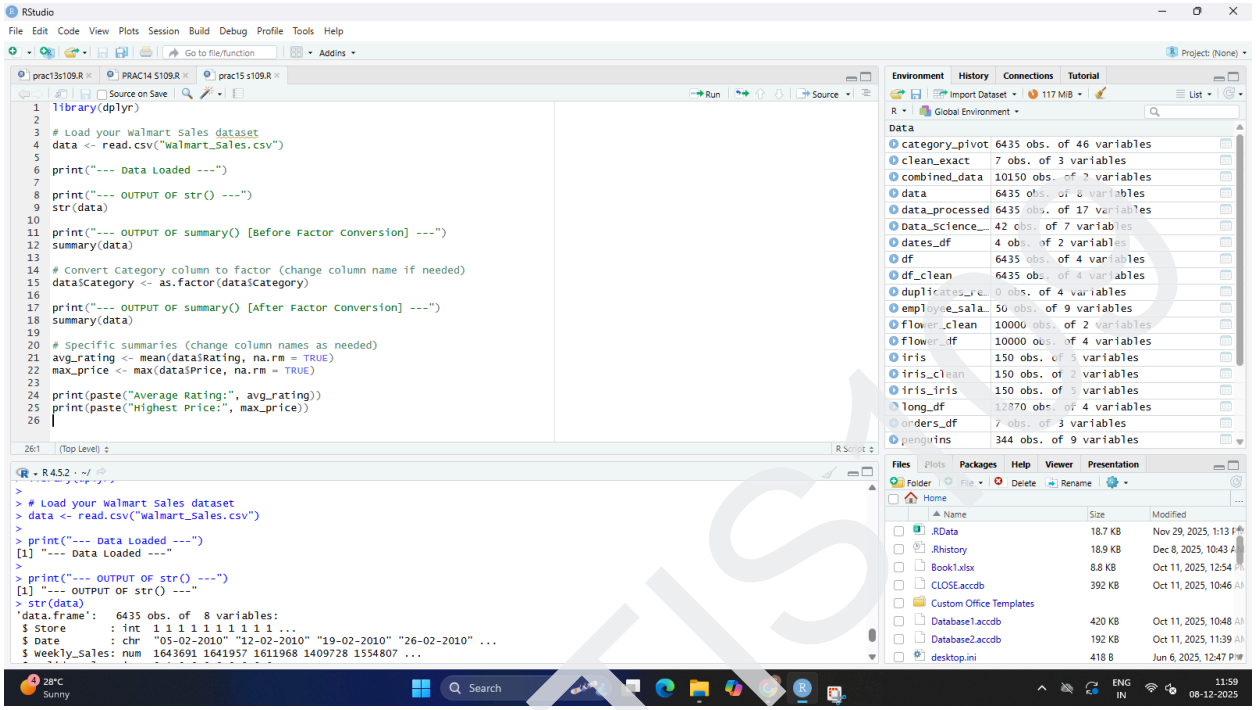
SHETH L.U.J AND M.V COLLEGE
PRACTICAL NO .15
SUBJECT - DATA ANALYSIS

Aim - 15. Generating basic summaries using str() or summary() (R).

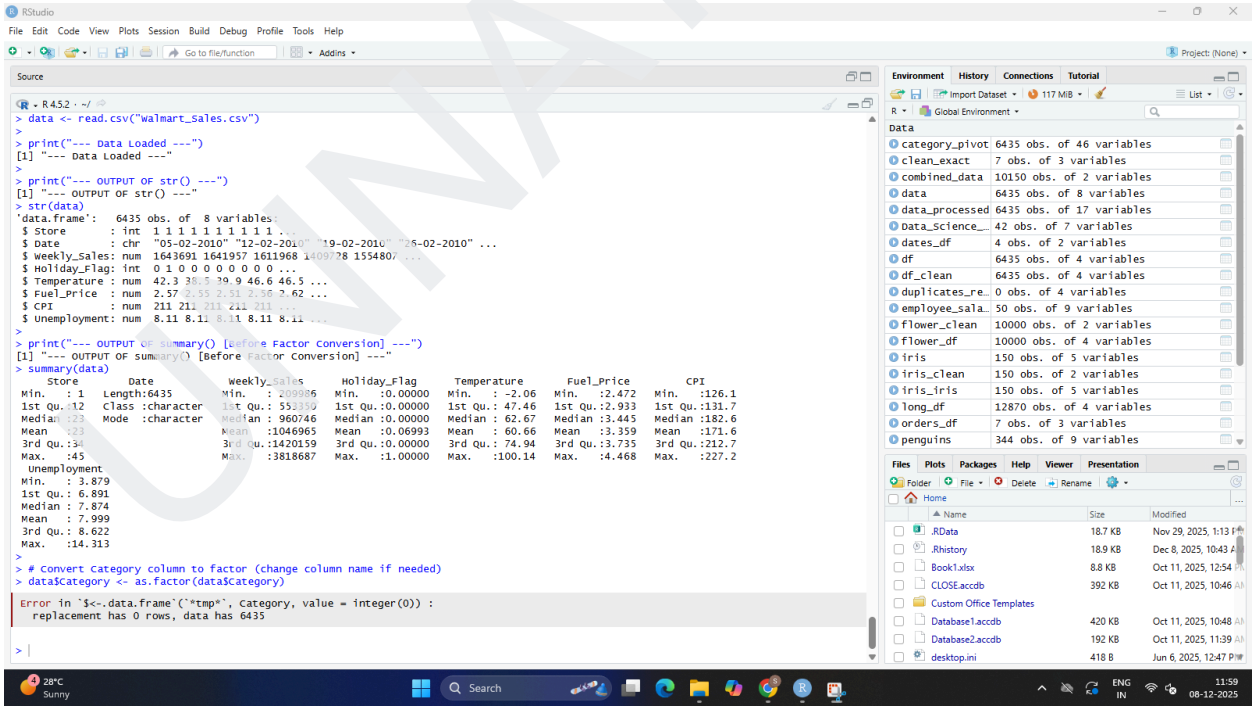
Input -



```
library(dplyr)

# Load your Walmart Sales dataset
data <- read.csv("Walmart_Sales.csv")

print("--- Data Loaded ---")

print("--- OUTPUT OF str() ---")
str(data)

print("--- OUTPUT OF summary() [Before Factor Conversion] ---")
summary(data)

# Convert Category column to factor (change column name if needed)
data$Category <- as.factor(data$Category)

print("--- OUTPUT OF summary() [After Factor Conversion] ---")
summary(data)

# Specific summaries (change column names as needed)
avg_rating <- mean(data$Rating, na.rm = TRUE)
max_price <- max(data$Price, na.rm = TRUE)

print(paste("Average Rating:", avg_rating))
print(paste("Highest Price:", max_price))
```

```
> # Load your Walmart Sales dataset
> data <- read.csv("Walmart_Sales.csv")
>
> print("--- Data Loaded ---")
[1] "--- Data Loaded ---"
>
> print("--- OUTPUT OF str() ---")
[1] "--- OUTPUT OF str() ---"
> str(data)
'data.frame':   6435 obs. of  8 variables:
 $ Store       : int  1 1 1 1 1 1 1 1 1 1 ...
 $ Date        : chr  "05-02-2010" "12-02-2010" "19-02-2010" "26-02-2010" ...
 $ Weekly_Sales: num  1643691 1641957 1611968 1409728 1554807 ...
```

OUTPUT -



```
> data <- read.csv("Walmart_Sales.csv")
>
> print("--- Data Loaded ---")
[1] "--- Data Loaded ---"
>
> print("--- OUTPUT OF str() ---")
[1] "--- OUTPUT OF str() ---"
> str(data)
'data.frame':   6435 obs. of  8 variables:
 $ Store       : int  1 1 1 1 1 1 1 1 1 1 ...
 $ Date        : chr  "05-02-2010" "12-02-2010" "19-02-2010" "26-02-2010" ...
 $ Weekly_Sales: num  1643691 1641957 1611968 1409728 1554807 ...
 $ Holiday_Flag: int  0 1 0 0 0 0 0 0 0 0 ...
 $ Temperature : num  42.3 38.5 39.9 46.6 46.5 ...
 $ Fuel_Price  : num  2.57 2.55 2.51 2.56 2.62 ...
 $ CPI         : num  211 211 211 211 211 ...
 $ Unemployment: num  8.11 8.11 8.11 8.11 8.11 ...
>
> print("--- OUTPUT OF summary() [Before Factor Conversion] ---")
[1] "--- OUTPUT OF summary() [Before Factor Conversion] ---"
> summary(data)
     Store          Date           Weekly_Sales      Holiday_Flag     Temperature      Fuel_Price         CPI
 Min.   : 1   Length:6435        Min.   : 209986   Min.   :0.00000   Min.   : -2.06   Min.   :2.472   Min.   :126.1
 1st Qu.:12   Class :character   1st Qu.: 553350   1st Qu.:0.00000   1st Qu.: 47.46   1st Qu.:2.933   1st Qu.:131.7
 Median :23   Mode  :character   Median : 960746   Median :0.00000   Median : 62.67   Median :3.445   Median :182.6
 Mean   :23                      Mean   :1046965   Mean   :0.06993   Mean   : 60.66   Mean   :3.359   Mean   :171.6
 3rd Qu.:34                      3rd Qu.:1420159   3rd Qu.:0.00000   3rd Qu.: 74.94   3rd Qu.:3.735   3rd Qu.:212.7
 Max.   :45                      Max.   :3818687   Max.   :1.00000   Max.   :100.14   Max.   :4.468   Max.   :227.2
  Unemployment
 Min.   : 3.879
 1st Qu.: 6.891
 Median : 7.874
 Mean   : 7.999
 3rd Qu.: 8.622
 Max.   :14.313
>
> # Convert Category column to factor (change column name if needed)
> data$Category <- as.factor(data$Category)
Error in `$<-.data.frame`(`*tmp*`, Category, value = integer(0)) :
  replacement has 0 rows, data has 6435
```

NAME - UNNATI RATHOD
ROLL NO S109