SHETH L.U.J.AND SIR M.V. COLLEGE
PRACTICAL NO.15
Data Analysis with SAS / SPSS /R

AIM;- Generating basic summaries using str() or summary() (R).

INPUT:-

```
# R Script: Generating Basic Summaries
# Functions: str() and summary()
# Dataset: Car Data (cardata.csv)

# 1. LOAD DATASET

# Make sure your working directory contains cardata.csv
setwd("D:/S079_VIBHUTI/ADV PYTHON FOR DATA SCIENCE")

car_df <- read.csv("cardata.csv")
print("--- Car Data Loaded Successfully ---")

# View first few rows
head(car_df)

# 2. USING str() (Structure)

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

# 3. USING summary() (Statistical Summary)
```

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

# 4. IMPROVING summary() WITH FACTORS

# Convert character columns to factors where appropriate
# You may adjust depending on your columns

# Example common car dataset columns:
# "FuelType", "Transmission", "Owner", "Seller_Type"

factor_cols <- c("FuelType", "Transmission", "Owner", "Seller_Type")

# Only convert columns that actually exist
factor_cols <- factor_cols[factor_cols %in% names(car_df)]

car_df[factor_cols] <- lapply(car_df[factor_cols], as.factor)

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

# 5. ACCESSING SPECIFIC SUMMARIES

# Example numeric fields in typical car datasets:
# "Selling_Price", "Present_Price", "Kms_Driven"
```

SHETH L.U.J.AND SIR M.V. COLLEGE
PRACTICAL NO.15
Data Analysis with SAS / SPSS /R

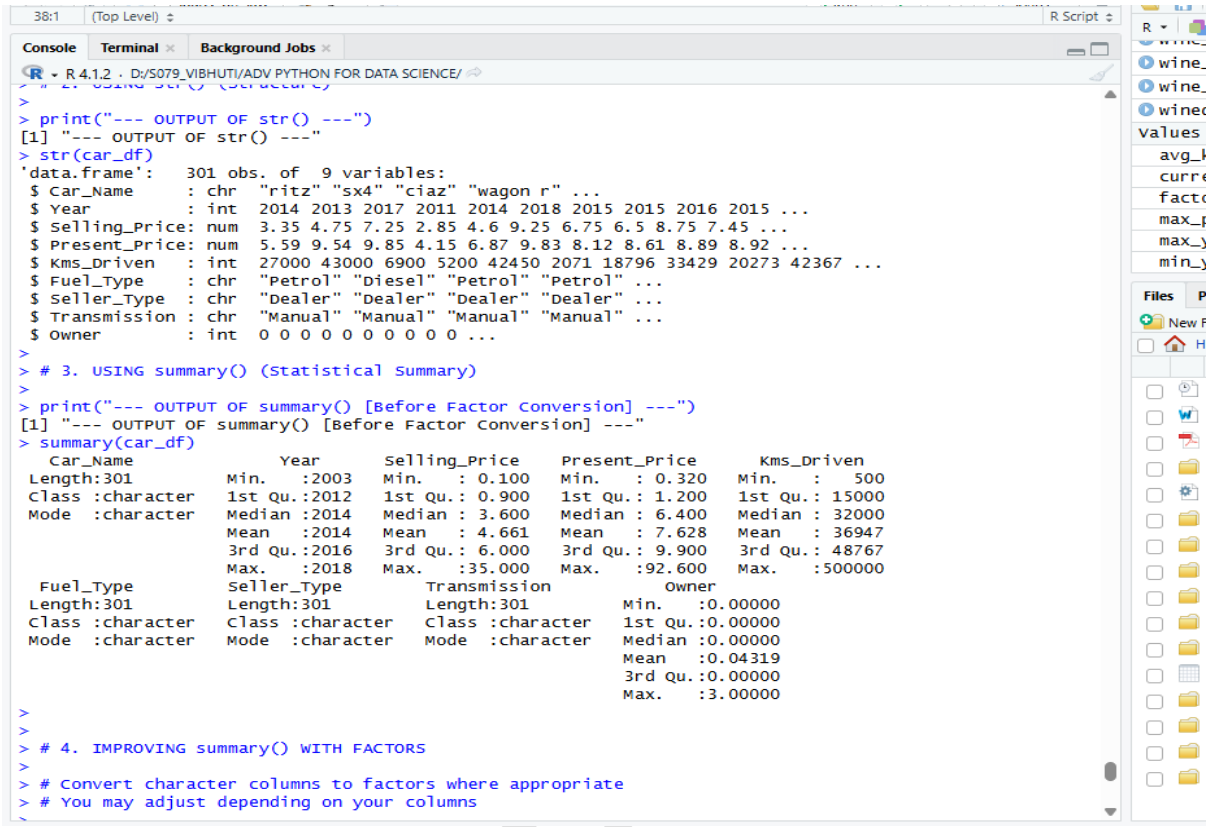
```
avg_kms <- mean(car_df$Kms_Driven, na.rm = TRUE)
max_price <- max(car_df$Selling_Price, na.rm = TRUE)
min_year <- min(car_df$Year, na.rm = TRUE)
max_year <- max(car_df$Year, na.rm = TRUE)

print(paste("Average KM Driven:", avg_kms))
print(paste("Highest Selling Price:", max_price))
print(paste("Oldest Car Year:", min_year))
print(paste("Newest Car Year:", max_year))
```

OUTPUT:-

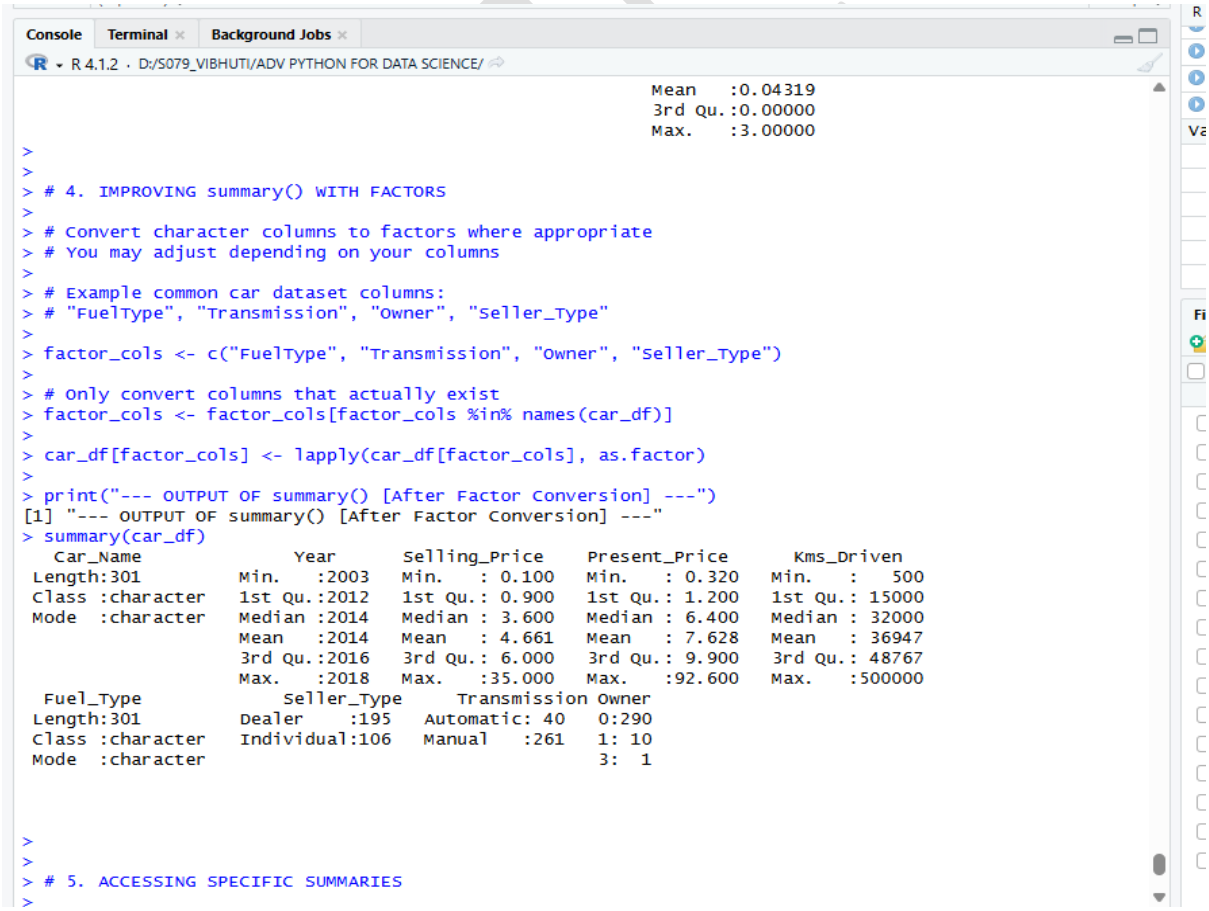
```
>
>
> # 1. LOAD DATASET
>
> # Make sure your working directory contains cardata.csv
> setwd("D:/S079_VIBHUTI/ADV PYTHON FOR DATA SCIENCE")
>
> car_df <- read.csv("cardata.csv")
> print("--- Car Data Loaded Successfully ---")
[1] "--- Car Data Loaded Successfully ---"
>
> # View first few rows
> head(car_df)
       Car_Name Year Selling_Price Present_Price Kms_Driven Fuel_Type Seller_Type Transmission
1          ritz 2014          3.35          5.59       27000    Petrol      Dealer       Manual
2           sx4 2013          4.75          9.54       43000    Diesel      Dealer       Manual
3          ciaz 2017          7.25          9.85        6900    Petrol      Dealer       Manual
4       wagon r 2011          2.85          4.15        5200    Petrol      Dealer       Manual
5         swift 2014          4.60          6.87       42450    Diesel      Dealer       Manual
6 vitara brezza 2018          9.25          9.83        2071    Diesel      Dealer       Manual
  Owner
1     0
2     0
3     0
4     0
5     0
6     0
>
>
> # 2. USING str() (Structure)
>
> print("--- OUTPUT OF str() ---")
[1] "--- OUTPUT OF str() ---"
> str(car_df)
'data.frame':   301 obs. of  9 variables:
 $ Car_Name     : chr  "ritz" "sx4" "ciaz" "wagon r" ...
 $ Year         : int  2014 2013 2017 2011 2014 2018 2015 2015 2016 2015 ...
 $ Selling_Price: num  3.35 4.75 7.25 2.85 4.6 9.25 6.75 6.5 8.75 7.45 ...
 $ Present_Price: num  5.59 9.54 9.85 4.15 6.87 9.83 8.12 8.61 8.89 8.92 ...
```


VIBHUTI GAWADE
S079

SHETH L.U.J.AND SIR M.V. COLLEGE
PRACTICAL NO.15
Data Analysis with SAS / SPSS /R



```
> print("--- OUTPUT OF str() ---")
[1] "--- OUTPUT OF str() ---"
> str(car_df)
'data.frame':   301 obs. of  9 variables:
 $ Car_Name     : chr  "ritz" "sx4" "ciaz" "wagon r" ...
 $ Year         : int  2014 2013 2017 2011 2014 2018 2015 2015 2016 2015 ...
 $ Selling_Price: num  3.35 4.75 7.25 2.85 4.6 9.25 6.75 6.5 8.75 7.45 ...
 $ Present_Price: num  5.59 9.54 9.85 4.15 6.87 9.83 8.12 8.61 8.89 8.92 ...
 $ Kms_Driven   : int  27000 43000 6900 5200 42450 2071 18796 33429 20273 42367 ...
 $ Fuel_Type    : chr  "Petrol" "Diesel" "Petrol" "Petrol" ...
 $ Seller_Type  : chr  "Dealer" "Dealer" "Dealer" "Dealer" ...
 $ Transmission : chr  "Manual" "Manual" "Manual" "Manual" ...
 $ Owner        : int  0 0 0 0 0 0 0 0 0 0 ...
>
> # 3. USING summary() (Statistical Summary)
>
> print("--- OUTPUT OF summary() [Before Factor Conversion] ---")
[1] "--- OUTPUT OF summary() [Before Factor Conversion] ---"
> summary(car_df)
   Car_Name              Year      Selling_Price    Present_Price      Kms_Driven
 Length:301         Min.   :2003   Min.   : 0.100   Min.   : 0.320   Min.   :   500
 Class :character   1st Qu.:2012   1st Qu.: 0.900   1st Qu.: 1.200   1st Qu.: 15000
 Mode  :character   Median :2014   Median : 3.600   Median : 6.400   Median : 32000
                    Mean   :2014   Mean   : 4.661   Mean   : 7.628   Mean   : 36947
                    3rd Qu.:2016   3rd Qu.: 6.000   3rd Qu.: 9.900   3rd Qu.: 48767
                    Max.   :2018   Max.   :35.000   Max.   :92.600   Max.   :500000
  Fuel_Type          Seller_Type        Transmission          Owner
 Length:301         Length:301         Length:301         Min.   :0.00000
 Class :character   Class :character   Class :character   1st Qu.:0.00000
 Mode  :character   Mode  :character   Mode  :character   Median :0.00000
                                                          Mean   :0.04319
                                                          3rd Qu.:0.00000
                                                          Max.   :3.00000
>
>
> # 4. IMPROVING summary() WITH FACTORS
>
> # Convert character columns to factors where appropriate
> # You may adjust depending on your columns
```



```
> # 4. IMPROVING summary() WITH FACTORS
>
> # Convert character columns to factors where appropriate
> # You may adjust depending on your columns
>
> # Example common car dataset columns:
> # "FuelType", "Transmission", "Owner", "Seller_Type"
>
> factor_cols <- c("FuelType", "Transmission", "Owner", "Seller_Type")
>
> # Only convert columns that actually exist
> factor_cols <- factor_cols[factor_cols %in% names(car_df)]
>
> car_df[factor_cols] <- lapply(car_df[factor_cols], as.factor)
>
> print("--- OUTPUT OF summary() [After Factor Conversion] ---")
[1] "--- OUTPUT OF summary() [After Factor Conversion] ---"
> summary(car_df)
   Car_Name              Year      Selling_Price    Present_Price      Kms_Driven
 Length:301         Min.   :2003   Min.   : 0.100   Min.   : 0.320   Min.   :   500
 Class :character   1st Qu.:2012   1st Qu.: 0.900   1st Qu.: 1.200   1st Qu.: 15000
 Mode  :character   Median :2014   Median : 3.600   Median : 6.400   Median : 32000
                    Mean   :2014   Mean   : 4.661   Mean   : 7.628   Mean   : 36947
                    3rd Qu.:2016   3rd Qu.: 6.000   3rd Qu.: 9.900   3rd Qu.: 48767
                    Max.   :2018   Max.   :35.000   Max.   :92.600   Max.   :500000
  Fuel_Type            Seller_Type     Transmission Owner
 Length:301         Dealer     :195   Automatic: 40   0:290
 Class :character   Individual:106   Manual   :261   1: 10
 Mode  :character                                    3:  1
>
>
> # 5. ACCESSING SPECIFIC SUMMARIES
>
```

VIBHUTI GAWADE
S079

SHETH L.U.J.AND SIR M.V. COLLEGE
PRACTICAL NO.15
Data Analysis with SAS / SPSS /R

```
>
>
> # 5. ACCESSING SPECIFIC SUMMARIES
>
>
> # Example numeric fields in typical car datasets:
> # "Selling_Price", "Present_Price", "Kms_Driven"
>
> avg_kms <- mean(car_df$Kms_Driven, na.rm = TRUE)
> max_price <- max(car_df$Selling_Price, na.rm = TRUE)
> min_year <- min(car_df$Year, na.rm = TRUE)
> max_year <- max(car_df$Year, na.rm = TRUE)
>
> print(paste("Average KM Driven:", avg_kms))
[1] "Average KM Driven: 36947.2059800664"
> print(paste("Highest Selling Price:", max_price))
[1] "Highest Selling Price: 35"
> print(paste("Oldest Car Year:", min_year))
[1] "Oldest Car Year: 2003"
> print(paste("Newest Car Year:", max_year))
[1] "Newest Car Year: 2018"
>
```


VIBHUTI GAWADE
S079