

Practical No 15

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

Output :

```

R - R 4.5.2 - C:/Users/mvuc/Downloads/
File Edit Code View Plots Session Build Debug Profile Tools Help
Source
Console Terminal Background Jobs
> 'BMW.sales.data.(2010.2024)'. <- read.csv("c:/users/mvuc/downloads/BMW sales data (2010-2024).csv")
> view('BMW.sales.data.(2010.2024)')
> library(dplyr)
> library(readr)
> # 1. Load the data
> df <- read_csv("BMW sales data (2010-2024).csv")
Rows: 50000 columns: 11
column specification
delimiter: ","
chr (6): Model, Region, Color, Fuel_Type, Transmission, Sales_Classification
dbl (5): Year, Engine_Size_L, Mileage_KM, Price_USD, Sales_Volume

i use 'spec()' to retrieve the full column specification for this data.
i specify the column types or set 'show_col_types = FALSE' to quiet this message.
> print("---- Data Loaded (First 6 Rows) ----")
[1] "---- Data Loaded (First 6 Rows) ----"
> print(head(df))
# A tibble: 6 x 11
  Model Year Region Color Fuel_Type Transmission Engine_Size_L Mileage_KM Price_USD Sales_Volume Sales_Classification
  <chr> <dbl> <chr> <chr> <chr> <chr> <dbl> <dbl> <dbl> <dbl> <chr>
1 5 Series 2016 Asia Red Petrol Manual 3.5 151748 98740 8300 High
2 i8 2013 North America Red Hybrid Automatic 1.6 121671 79219 3428 Low
3 5 Series 2022 North America Blue Petrol Automatic 4.5 10991 113265 6994 Low
4 x3 2024 Middle East Blue Petrol Automatic 1.7 22255 60971 4047 Low
5 7 Series 2020 South America Black Diesel Manual 2.1 122131 49898 3080 Low
6 5 Series 2017 Middle East Silver Diesel Manual 1.9 171362 42926 1232 Low
> print("---- OUTPUT OF str() ----")
[1] "---- OUTPUT OF str() ----"
> str(df)
'spec_tbl' [50,000 x 11] (s3: spec_tbl_df/tbl_df/tbl/data.frame)
 $ Model      : chr [1:50000] "5 Series" "i8" "5 Series" "x3" ...
 $ Year       : num [1:50000] 2016 2013 2022 2024 2020 ...
 $ Region     : chr [1:50000] "Asia" "North America" "North America" "Middle East" ...
 $ Color      : chr [1:50000] "Red" "Red" "Blue" "Blue" ...
 $ Fuel_Type  : chr [1:50000] "Petrol" "Hybrid" "Petrol" "Petrol" ...
 $ Transmission : chr [1:50000] "Manual" "Automatic" "Automatic" "Automatic" ...
 $ Engine_Size_L : num [1:50000] 3.5 1.6 4.5 1.7 2.1 1.9 1.8 1.6 1.7 3 ...
 $ Mileage_KM : num [1:50000] 151748 121671 10991 22255 122131 ...
 $ Price_USD  : num [1:50000] 98740 79219 113265 60971 49898 ...
 $ Sales_Volume : num [1:50000] 8300 3428 6994 4047 3080 ...
 $ Sales_Classification: chr [1:50000] "High" "Low" "Low" "Low" ...
 - attr(*, "spec")=
  
```

SHETH L.U.J. AND SIR M.V. COLLEGE OF ARTS SCIENCE AND COMMERCE

SUBJECT : R Programming

```
RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function Addins
Source
Console Terminal Background Jobs
R - R 4.5.2 - C:/Users/mvuc/Downloads/
> Model <- chr [1:50000] "5 Series" "18" "3 Series" "X3" ...
> Year <- num [1:50000] 2016 2013 2022 2024 2020 ...
> Region <- chr [1:50000] "Asia" "North America" "North America" "Middle East" ...
> Color <- chr [1:50000] "Red" "Red" "Blue" "Blue" ...
> Fuel_Type <- chr [1:50000] "Petrol" "Hybrid" "Petrol" "Petrol" ...
> Transmission <- chr [1:50000] "Manual" "Automatic" "Automatic" "Automatic" ...
> Engine_Size_L <- num [1:50000] 3.5 1.6 4.5 1.7 2.1 1.9 1.8 1.6 1.7 3 ...
> Mileage_KM <- num [1:50000] 151748 121671 10991 27255 122131 ...
> Price_USD <- num [1:50000] 98740 79219 113265 60971 49898 ...
> Sales_Volume <- num [1:50000] 8300 3428 6994 4047 3080 ...
> Sales_Classification <- chr [1:50000] "High" "Low" "Low" "Low" ...
> attr(,"spec")=
.. cols(
.. Model = col_character(),
.. Year = col_double(),
.. Region = col_character(),
.. Color = col_character(),
.. Fuel_Type = col_character(),
.. Transmission = col_character(),
.. Engine_Size_L = col_double(),
.. Mileage_KM = col_double(),
.. Price_USD = col_double(),
.. Sales_Volume = col_double(),
.. Sales_Classification = col_character()
.. )
> attr(,"problems")=externalptr
> print("---- OUTPUT of summary() [Before Factor Conversion] ----")
> summary(df)
  Model      Year      Region      Color      Fuel_Type      Transmission      Engine_Size_L      Mileage_KM      Price_USD
Length:50000 Min.   :2010 Length:50000 Length:50000 Length:50000 Length:50000 Min.   :1.500 Min.   : 3 Min.   :30000
Class :character 1st Qu.:2013 Class :character Class :character Class :character Class :character 1st Qu.:2.400 1st Qu.: 50178 1st Qu.: 52435
Mode :character Median :2017 Mode :character Mode :character Mode :character Median :3.200 Median :100389 Median : 75012
Mean :2017 3rd Qu.:2021 Max.   :2024 Mean :3.247 Mean :100307 Mean : 75035
3rd Qu.:4.100 3rd Qu.:150630 3rd Qu.: 97628
Max.   :5.000 Max.   :199996 Max.   :119998

  Sales_Volume Sales_Classification
Min.   :100 Length:50000
1st Qu.:2588 Class :character
Median :5087 Mode :character
Mean :5068
3rd Qu.:7537
Max. :9999

> # 4. Convert character columns to factors to improve summary() output
> df$Region <- as.factor(df$Region)
> df$Fuel_Type <- as.factor(df$Fuel_Type)
> print("---- OUTPUT of summary() [After Factor Conversion] ----")
> summary(df)
  Model      Year      Region      Color      Fuel_Type      Transmission      Engine_Size_L      Mileage_KM      Price_USD
Length:50000 Min.   :2010 Africa :8253 Length:50000 Diesel :12263 Length:50000 Min.   :1.500 Min.   : 3 Min.   :30000
Class :character 1st Qu.:2013 Asia   :8454 Class :character Electric:12471 Class :character 1st Qu.:2.400 1st Qu.: 50178 1st Qu.: 52435
Mode :character Median :2017 Europe :8334 Mode :character Hybrid :12716 Mode :character Median :3.200 Median :100389 Median : 75012
Mean :2017 Middle East :8373 Mean :3.247 Mean :100307 Mean : 75035
3rd Qu.:2021 North America:8335 3rd Qu.:4.100 3rd Qu.:150630 3rd Qu.: 97628
Max.   :2024 South America:8251 Max.   :5.000 Max.   :199996 Max.   :119998

  Sales_Volume Sales_Classification
Min.   :100 Length:50000
1st Qu.:2588 Class :character
Median :5087 Mode :character
Mean :5068
3rd Qu.:7537
Max. :9999

> # 5. Accessing Specific Summaries
> # na.rm = TRUE handles any missing values, though the data is complete
> avg_price <- mean(df$Price_USD, na.rm = TRUE)
> max_sales_volume <- max(df$Sales_Volume, na.rm = TRUE)
> print("---- Accessing Specific Summaries ----")
[1] "---- Accessing Specific Summaries ----"
> print(paste("Average Price (USD):", format(round(avg_price, 2), big.mark = ",")))
[1] "Average Price (USD): 75,034.6"
> print(paste("Highest Sales Volume:", format(max_sales_volume, big.mark = ",")))
[1] "Highest Sales Volume: 9,999"
> |
```

NAME : SHUBHAM SANJAY KARAPE
ROLL NO : S085