

Sheth L.U.J & Sir M.V College Of Science
Subject :- Data Analysis with SAS / SPSS / R
Practical No 8

Aim :- Applying basic data cleaning functions: handling missing values using `na.omit()/replace_na()` in R. import dataset.

```
Console Terminal Background Jobs
R - R452 - ~/
> library(readr)
> library(dplyr)
> library(tidyverse)
> car_df <- read_csv("BMW sales data (2010-2024) (1).csv",
+                   na = c("", "NA"))
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("--- Original Data ---")
[1] "--- Original Data ---"
> print(head(car_df))
# A tibble: 6 x 11
  Model Year Region Color Fuel_Type Transmission Engine_Size_L Mileage_KM Price_USD Sales_Volume
<chr> <dbl> <chr> <chr> <chr> <chr> <dbl> <dbl> <dbl> <dbl>
1 5 Series 2016 Asia Red Petrol Manual 3.5 151748 98740 8300
2 i8 2013 North America Red Hybrid Automatic 1.6 121671 79219 3428
3 5 Series 2022 North America Blue Petrol Automatic 4.5 10991 113265 6994
4 X3 2024 Middle East Blue Petrol Automatic 1.7 27255 60971 4047
5 7 Series 2020 South America Black Diesel Manual 2.1 122131 49898 3080
6 5 Series 2017 Middle East Silver Diesel Manual 1.9 171362 42926 1232
# i 1 more variable: Sales_Classification <chr>
> print("--- Missing Values per Column ---")
[1] "--- Missing Values per Column ---"
> print(colSums(is.na(car_df)))
      Model      Year      Region      Color      Fuel_Type
      0         0         0         0         0
      Transmission      Engine_Size_L      Mileage_KM      Price_USD      Sales_Volume
      0         0         0         0         0
Sales_Classification
      0
> # Create clean_omit
> clean_omit <- na.omit(car_df)
> print("--- After Removing Rows with NA ---")
[1] "--- After Removing Rows with NA ---"
> print(paste("Original rows:", nrow(car_df)))
[1] "Original rows: 50000"
```

Name :- Vinod Mali
Roll No :- S092

Sheth L.U.J & Sir M.V College Of Science
Subject :- Data Analysis with SAS / SPSS /R
Practical No 8

```
[1] "Original rows: 50000"
> print(paste("Remaining rows:", nrow(clean_omit)))
[1] "Remaining rows: 50000"
> print(head(clean_omit))
# A tibble: 6 × 11
  Model      Year Region      Color Fuel_Type Transmission Engine_Size_L Mileage_KM Price_USD Sales_Volume
  <chr>    <dbl> <chr>    <chr> <chr>    <chr>    <dbl>    <dbl>    <dbl>    <dbl>
1 5 Series  2016 Asia      Red    Petrol    Manual      3.5    151748    98740    8300
2 i8       2013 North America Red    Hybrid    Automatic    1.6    121671    79219    3428
3 5 Series  2022 North America Blue   Petrol    Automatic    4.5    10991    113265    6994
4 X3       2024 Middle East Blue   Petrol    Automatic    1.7    27255    60971    4047
5 7 Series  2020 South America Black  Diesel    Manual      2.1    122131    49898    3080
6 5 Series  2017 Middle East Silver Diesel    Manual      1.9    171362    42926    1232
# 1 more variable: Sales_Classification <chr>
> |
```

29°C
Sunny



Search



Name :- Vinod Mali
Roll No :- S092