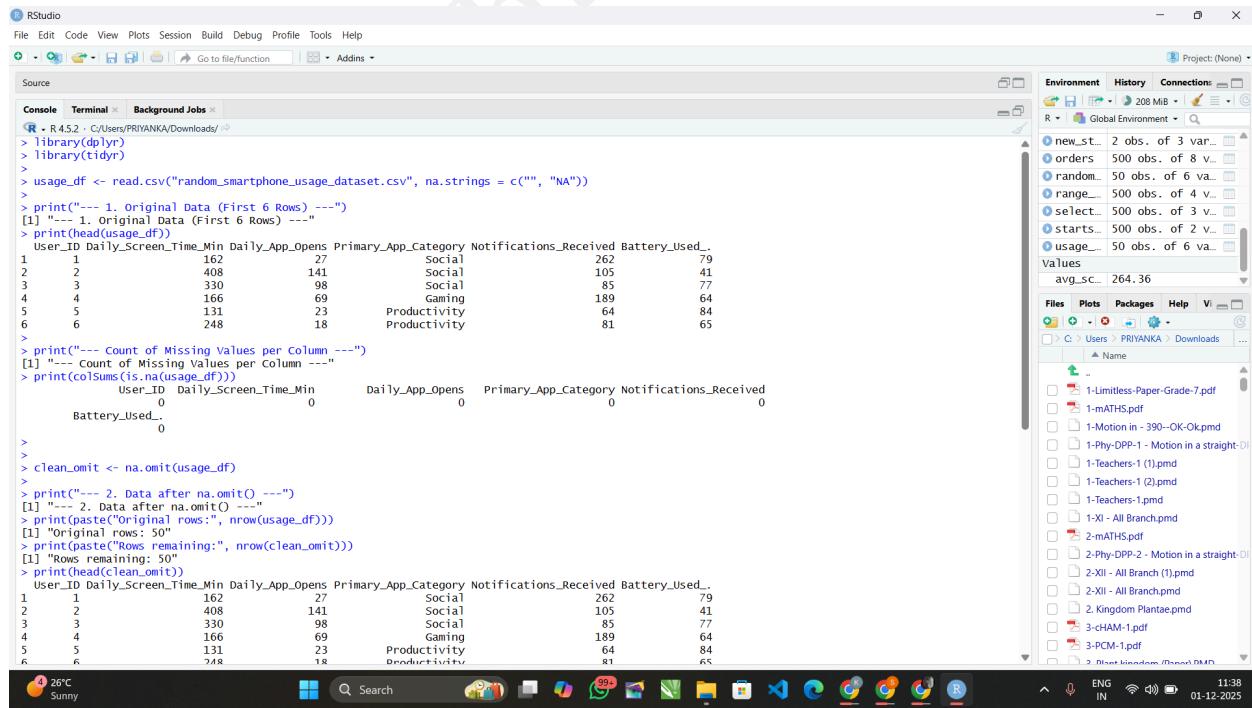


**SHETH L.U.J AND SIR M.V. COLLEGE**  
**SUBJECT NAME: DATA ANALYSIS WITH SAS/SPSS/R**

## Module 1 Practical 8

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

### OUTPUT:



The screenshot shows the RStudio interface with the following details:

- Console Tab:** Displays the R session code and its output. The code reads a CSV file, prints the first 6 rows, shows the count of missing values per column, and then uses `na.omit()` to remove rows with missing values. It also prints the original rows and the rows remaining after omitting missing values.
- Environment Tab:** Shows the global environment with objects like `new_st...`, `orders...`, `random...`, etc., and their respective dimensions.
- Files Tab:** Shows the file structure under `C:/Users/PRIYANKA/Downloads/`, including various PDF files and PMD files.
- Plots Tab:** No plots are present.
- Packages Tab:** No packages are present.
- Help Tab:** No help is present.

Below the RStudio window, the Windows taskbar shows the date and time as 01-12-2025, 11:38, ENG IN.

```
RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
R - R 4.5.2 · C:/users/PRIYANKA/Downloads/ ...
Source
Console Terminal Background Jobs
[R - R 4.5.2 · C:/users/PRIYANKA/Downloads/ ...]
> library(dplyr)
> library(tidyverse)
>
> usage_df <- read.csv("random_smartphone_usage_dataset.csv", na.strings = c("", "NA"))
> print("--- 1. Original Data (First 6 Rows) ---")
[1] "--- 1. Original Data (First 6 Rows) ---"
> print(head(usage_df))
#> #> #> #> #> #>
User_ID Daily_Screen_Time_Min Daily_App_Openes Primary_App_Category Notifications_Received Battery_Used_
1       1           162             27          Social            262              79
2       2           408             141          Social            105              41
3       3           330             98          Social             85              77
4       4           166              69         Gaming            189              64
5       5           131              23    Productivity            64              84
6       6           248              18    Productivity            81              65
> print("---- Count of Missing Values per Column ----")
[1] "---- Count of Missing Values per Column ----"
> print(columns(is.na(usage_df)))
#> #> #> #> #> #>
      User_ID Daily_Screen_Time_Min Daily_App_Openes Primary_App_Category Notifications_Received
      0                  0                  0                      0                      0
      Battery_Used_
      0
>
> clean.omit <- na.omit(usage_df)
>
> print("---- 2. Data after na.omit() ----")
[1] "---- 2. Data after na.omit() ----"
> print(paste("Original rows:", nrow(usage_df)))
[1] "Original rows: 50"
> print(paste("Rows remaining:", nrow(clean.omit)))
[1] "Rows remaining: 50"
> print(head(clean.omit))
#> #> #> #> #> #>
User_ID Daily_Screen_Time_Min Daily_App_Openes Primary_App_Category Notifications_Received Battery_Used_
1       1           162             27          Social            262              79
2       2           408             141          Social            105              41
3       3           330             98          Social             85              77
4       4           166              69         Gaming            189              64
5       5           131              23    Productivity            64              84
6       6           248              18    Productivity            81              65
```

**SHETH L.U.J AND SIR M.V. COLLEGE**  
**SUBJECT NAME: DATA ANALYSIS WITH SAS/SPSS/R**

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Source

Console Terminal Background Jobs

R - R 4.5.2 - C:/Users/PRIYANKA/Downloads/

```
[1] "Rows remaining: 50"
> print(head(clean_omit))
#> #> #> #> #> #>
User_ID Daily_Screen_Time_Min Daily_App_Openes Primary_App_Category Notifications_Received Battery_Used_.
1       1             162                  27           Social            262                79
2       2             408                  141           Social            105                41
3       3             330                  98           Social             85                77
4       4             166                  69          Gaming            189                64
5       5             131                  23 Productivity            64                84
6       6             248                  18 Productivity            81                65
>
>
> avg_screen_time <- mean(usage_df$Daily_Screen_Time_Min, na.rm = TRUE)
>
> clean_replace <- usage_df %>%
+   replace_na(list(
+     Primary_App_Category = "Unknown",
+     Daily_App_Openes = 0,
+     Notifications_Received = 0,
+     Daily_Screen_Time_Min = avg_screen_time
+   ))
>
> print("## 3. Data after replace_na() ##")
[1] "## 3. Data after replace_na() ##"
> print(head(clean_replace))
#> #> #> #> #> #>
User_ID Daily_Screen_Time_Min Daily_App_Openes Primary_App_Category Notifications_Received Battery_Used_.
1       1             162                  27           Social            262                79
2       2             408                  141           Social            105                41
3       3             330                  98           Social             85                77
4       4             166                  69          Gaming            189                64
5       5             131                  23 Productivity            64                84
6       6             248                  18 Productivity            81                65
>
> print("## Remaining NAs after replacement ##")
[1] "## Remaining NAs after replacement ##"
> print(coltums(is.na(clean_replace)))
#> #> #> #> #> #>
User_ID Daily_Screen_Time_Min Daily_App_Openes Primary_App_Category Notifications_Received Battery_Used_.
0                   0                  0           0                 0                  0
Battery_Used_.
0
```

Environment History Connections

Project: (None)

26°C Sunny

Search

11:38 01-12-2023