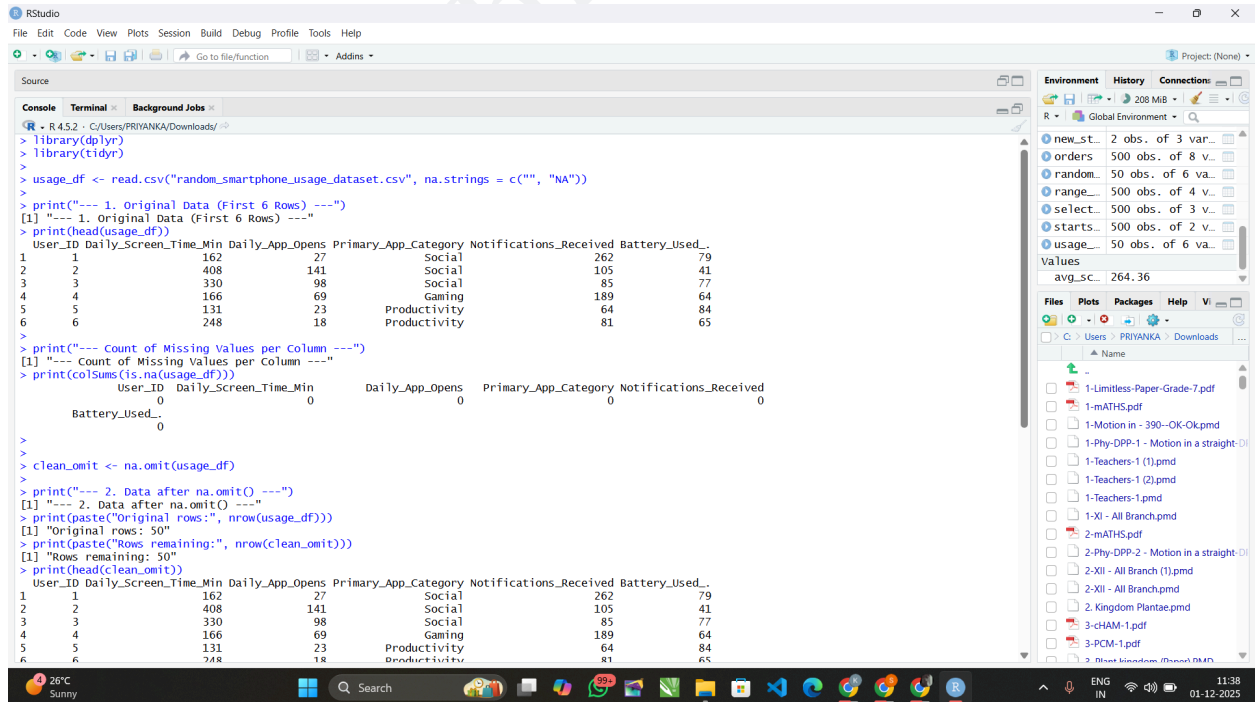


SHETH L.U.J AND SIR M.V. COLLEGE
SUBJECT NAME: DATA ANALYSIS WITH SAS/SPSS/R

Module 1 Practical 8

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

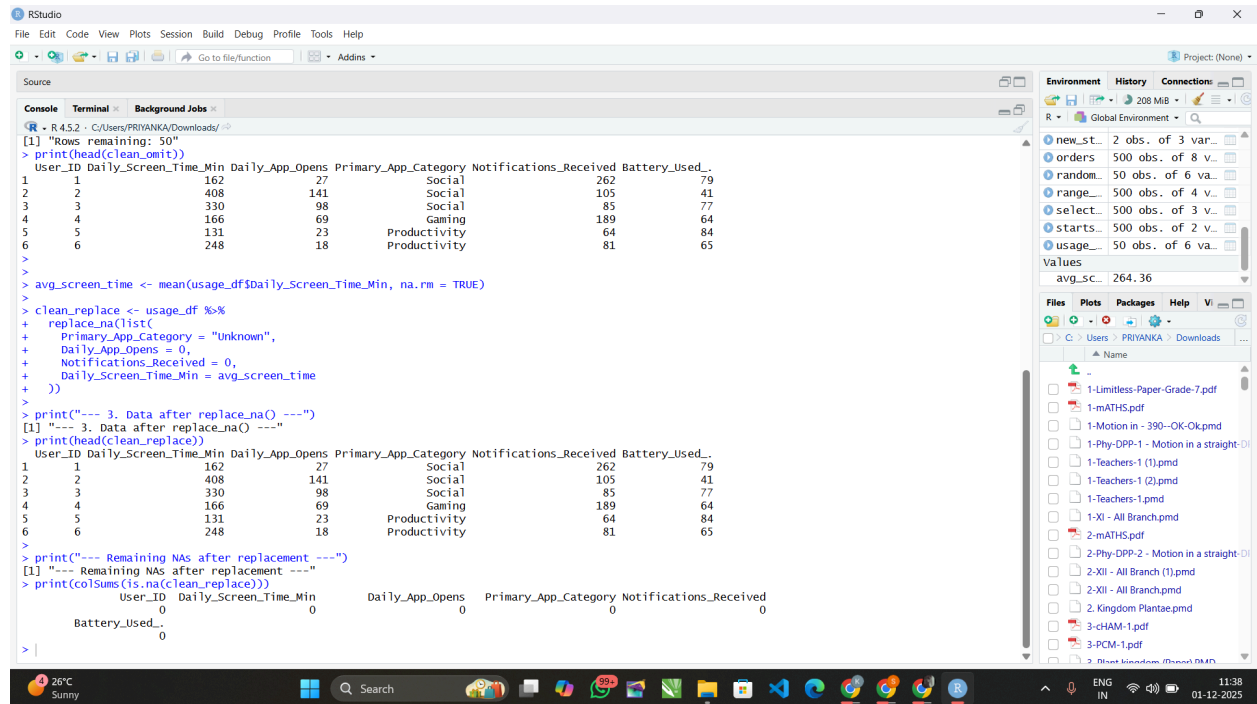
OUTPUT:



```
R
> library(dplyr)
> library(tidy)
> 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_Opens 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(colSums(is.na(usage_df)))
  User_ID Daily_Screen_Time_Min Daily_App_Opens Primary_App_Category Notifications_Received Battery_Used_
0         0                     0                0                    0                  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_Opens 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



The screenshot displays the RStudio environment with the following components:

- Source Editor:** Contains R code for data manipulation and summary statistics.
- Console:** Shows the output of the R code, including data frames and summary statistics.
- Environment Pane:** Lists the objects in the global environment, including 'new_st', 'orders', 'random', 'range', 'select', 'starts', 'usage', and 'avg_sc'.
- Files Pane:** Shows a list of files in the 'Downloads' folder, including PDFs and PPTs.

R Code:

```
[1] "Rows remaining: 50"
> print(head(clean_omit))
  User_ID Daily_Screen_Time_Min Daily_App_Opens 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_Opens = 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_Opens 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(colSums(is.na(clean_replace)))
  User_ID Daily_Screen_Time_Min Daily_App_Opens Primary_App_Category Notifications_Received
0      0          0          0          0          0
  Battery_Used_
0
```