

## MVLU COLLEGE

### PRACTICAL NO :- 06

**AIM :- Combining and appending datasets using merge() or bind\_rows() in R.**

### CODE :-

```
library(dplyr)
```

```
data_main <-
```

```
read.csv("C:/Users/itlab/Downloads/S100/Mental_Health_and_Social_Media_Balance_Data  
set.csv")
```

```
names(data_main)
```

```
print("--- Original Dataset ---")
```

```
print(head(data_main))
```

### # 2. CREATE TWO SUBSETS FOR MERGING

```
# Dataset A → Basic Info
```

```
data_A <- data_main[, c("User_ID", "Age", "Gender")]
```

```
# Dataset B: Screen, Sleep, Stress, Happiness
```

```
data_B <- data_main[, c("User_ID",  
                        "Daily_Screen_Time.hrs.",  
                        "Sleep_Quality.1.10.",  
                        "Stress_Level.1.10.",  
                        "Happiness_Index.1.10.")]
```

```
print("--- Dataset A ---")
```

```
print(head(data_A))
```

```
print("--- Dataset B (Mental Health Info) ---")
```

```
print(head(data_B))
```

### # 3. MERGE (Joining Columns)

```
merged_data <- merge(data_A, data_B, by = "User_ID")
```

```
print("--- Merged Dataset ---")
```

```
print(head(merged_data))
```

```
data_new_users <- data.frame(  
  User_ID = c("U151", "U152"),  
  Age = c(29, 33),  
  Gender = c("Female", "Male"),
```

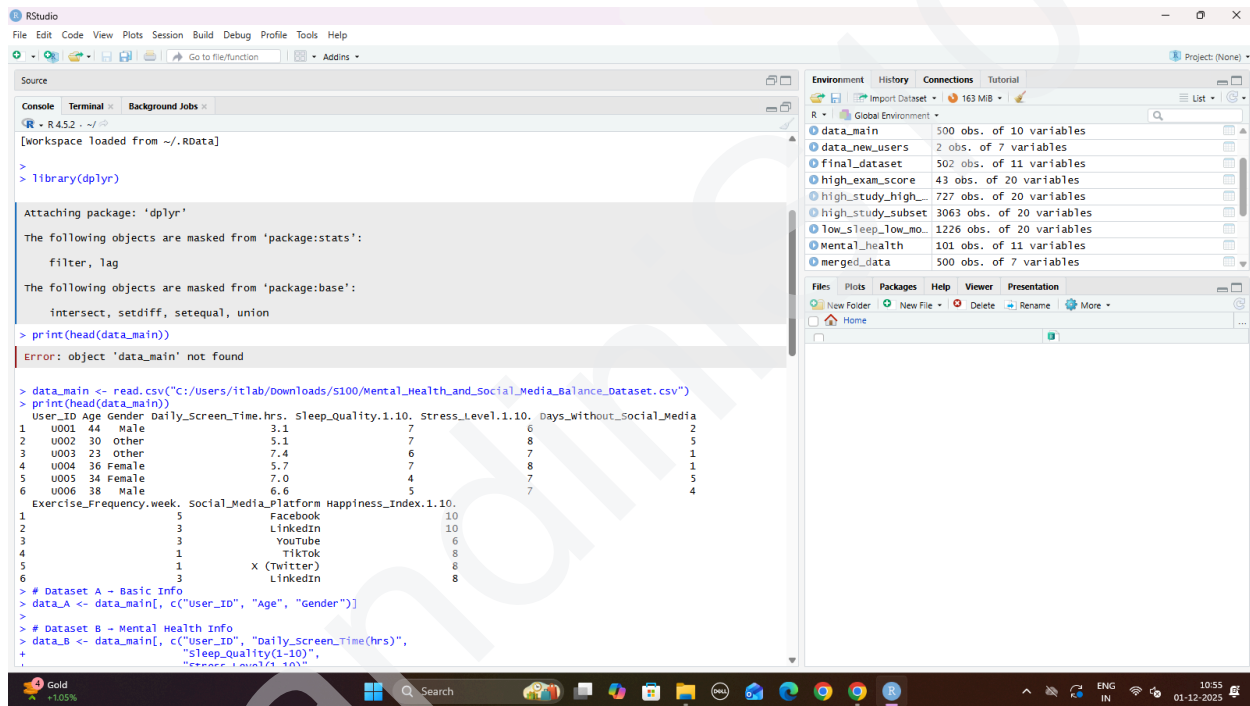
## MVLU COLLEGE

```
Daily_Screen_Time.hrs = c(4.5, 6.2),  
Sleep_Quality.1.10 = c(7, 6),  
Stress_Level.1.10 = c(5, 7),  
Happiness_Index.1.10 = c(8, 7)  
)
```

```
final_dataset <- bind_rows(merged_data, data_new_users)
```

```
print("--- Final Appended Dataset ---")  
print(head(final_dataset))
```

### OUTPUT :-



The screenshot shows the RStudio interface. The console on the left displays the following output:

```
[workspace loaded from ~/.RData]  
> library(dplyr)  
Attaching package: 'dplyr'  
The following objects are masked from 'package:stats':  
  filter, lag  
The following objects are masked from 'package:base':  
  intersect, setdiff, setequal, union  
> print(head(data_main))  
Error: object 'data_main' not found  
  
> data_main <- read.csv("C:/Users/itlab/downloads/S100/Mental_Health_and_Social_Media_Balance_Dataset.csv")  
> print(head(data_main))  
  User_ID Age Gender Daily_Screen_Time.hrs Sleep_Quality.1.10 Stress_Level.1.10 Days_Without_Social_Media  
1  U001  44  Male         3.1             7                6                2  
2  U002  30  other         5.1             7                8                5  
3  U003  23  other         7.4             6                7                1  
4  U004  36  Female        5.7             7                8                1  
5  U005  34  Female        7.0             4                7                5  
6  U006  38  Male         6.6             5                7                4  
  
  Exercise_Frequency.week Social_Media_Platform Happiness_Index.1.10  
1             5             Facebook             10  
2             3             LinkedIn             10  
3             3             YouTube              6  
4             1             Tiktok              8  
5             1             X (Twitter)           8  
6             3             LinkedIn             8  
  
> # Dataset A - Basic Info  
> data_A <- data_main[, c("User_ID", "Age", "Gender")]  
>  
> # Dataset B - Mental Health Info  
> data_B <- data_main[, c("User_ID", "Daily_Screen_Time(hrs)",  
+ "Sleep_Quality(1-10)",  
+ "Stress_Level(1-10)",  
+ "Days_Without_Social_Media")]
```

The Environment pane on the right shows the following objects:

Object	Size
data_main	500 obs. of 10 variables
data_new_users	2 obs. of 7 variables
final_dataset	502 obs. of 11 variables
high_exam_score	43 obs. of 20 variables
high_study_high	727 obs. of 20 variables
high_study_subset	3063 obs. of 20 variables
low_sleep_low_mo	1226 obs. of 20 variables
Mental_health	101 obs. of 11 variables
merged_data	500 obs. of 7 variables

# MVLU COLLEGE

The image displays two screenshots of the RStudio interface, showing the progression of an R script execution. The top screenshot shows an error in the console: "Error in '[.data.frame](data\_main, , c('User\_ID', 'daily\_Screen\_Time(hrs)', : undefined columns selected". The code in the source editor attempts to select columns from 'data\_main' that do not exist. The bottom screenshot shows the successful execution of the script. The console displays the output of several print statements, showing the structure and content of the data frames. The environment pane on the right shows the creation of new data frames: 'data\_new\_users' (2 obs. of 7 variables), 'final\_dataset' (502 obs. of 11 variables), 'high\_exam\_score' (43 obs. of 20 variables), 'high\_study\_high' (727 obs. of 20 variables), 'high\_study\_subset' (3063 obs. of 20 variables), 'low\_sleep\_low\_mo' (1226 obs. of 20 variables), 'Mental\_health' (101 obs. of 11 variables), and 'merged\_data' (500 obs. of 7 variables).

```
R - R4.5.2 - ~/RStudio
Source
Console Terminal Background Jobs
R - R4.5.2 - ~/RStudio
Error in '[.data.frame](data_main, , c('User_ID', 'daily_Screen_Time(hrs)', :
undefined columns selected

>
> view(data_main)
> view(data_A)
> names(data_main)
[1] "User_ID" "Age" "Gender" "daily_Screen_Time(hrs)."
[5] "sleep_quality.1.10." "Stress_Level.1.10." "Days_Without_Social_Media" "Exercise_Frequency.week."
[9] "Social_Media_Platform" "Happiness_Index.1.10."
# Dataset B: Screen, sleep, stress, happiness
data_B <- data_main[, c("User_ID",
+ "daily_Screen_Time(hrs).",
+ "sleep_quality.1.10.",
+ "Stress_Level.1.10.",
+ "Happiness_Index.1.10.")]
> print(head(data_A))
  User_ID Age Gender
1  U001  44  Male
2  U002  30  Other
3  U003  23  Other
4  U004  36  Female
5  U005  34  Female
6  U006  38  Male
> print(head(data_B))
+
+
+ print(head(data_B))
Error: unexpected symbol in:
"
print"

> print(head(data_B))
  User_ID daily_Screen_Time(hrs. sleep_quality.1.10. Stress_Level.1.10. Happiness_Index.1.10.
1  U001           3.1           7           6           10
2  U002           5.1           7           8           10
3  U003           7.4           6           7           6
4  U004           5.7           7           8           8
5  U005           7.0           4           7           8
6  U006           6.6           5           7           8

R - R4.5.2 - ~/RStudio
Source
Console Terminal Background Jobs
R - R4.5.2 - ~/RStudio
  User_ID Daily_Screen_Time(hrs. sleep_quality.1.10. Stress_Level.1.10. Happiness_Index.1.10.
1  U001           3.1           7           6           10
2  U002           5.1           7           8           10
3  U003           7.4           6           7           6
4  U004           5.7           7           8           8
5  U005           7.0           4           7           8
6  U006           6.6           5           7           8
> print(head(merged_data))
Error: object 'merged_data' not found

> # 3. MERGE (Joining Columns)
> merged_data <- merge(data_A, data_B, by = "User_ID")
> print(head(merged_data))
  User_ID Age Gender Daily_Screen_Time(hrs. sleep_quality.1.10. Stress_Level.1.10. Happiness_Index.1.10.
1  U001  44  Male           3.1           7           6           10
2  U002  30  Other           5.1           7           8           10
3  U003  23  Other           7.4           6           7           6
4  U004  36  Female          5.7           7           8           8
5  U005  34  Female           7.0           4           7           8
6  U006  38  Male           6.6           5           7           8
> data_new_users <- data.frame(
+   User_ID = c("U151", "U152"),
+   Age = c(29, 33),
+   Gender = c("Female", "Male"),
+   Daily_Screen_Time(hrs = c(4.5, 6.2),
+   sleep_quality.1.10 = c(7, 6),
+   Stress_Level.1.10 = c(5, 7),
+   Happiness_Index.1.10 = c(8, 7)
+ )
> final_dataset <- bind_rows(merged_data, data_new_users)
> print(head(final_dataset))
  User_ID Age Gender Daily_Screen_Time(hrs. sleep_quality.1.10. Stress_Level.1.10. Happiness_Index.1.10.
1  U001  44  Male           3.1           7           6           10
2  U002  30  Other           5.1           7           8           10
3  U003  23  Other           7.4           6           7           6
4  U004  36  Female          5.7           7           8           8
5  U005  34  Female           7.0           4           7           8
6  U006  38  Male           6.6           5           7           8
  daily_Screen_Time(hrs sleep_quality.1.10 Stress_Level.1.10 Happiness_Index.1.10
1                NA                NA                NA                NA
2                NA                NA                NA                NA

Environment History Connections Tutorial
R - Global Environment
data_main      500 obs. of 10 variables
data_new_users  2 obs. of 7 variables
final_dataset  502 obs. of 11 variables
high_exam_score 43 obs. of 20 variables
high_study_high 727 obs. of 20 variables
high_study_subset 3063 obs. of 20 variables
low_sleep_low_mo 1226 obs. of 20 variables
Mental_health  101 obs. of 11 variables
merged_data    500 obs. of 7 variables

Files Plots Packages Help Viewer Presentation
New Folder New File Delete Rename More
Home
```

# MVLU COLLEGE

The screenshot shows the RStudio interface with the following content:

**Console:**

```
R > R 4.5.2 ~ /  
> merged_data <- merge(data_A, data_B, by = "user_ID")  
> print(head(merged_data))  
Error: object 'merged_data' not found  
  
> # 3. MERGE (joining columns)  
> merged_data <- merge(data_A, data_B, by = "user_ID")  
> print(head(merged_data))  
  User_ID Age Gender Daily_Screen_Time.hrs Sleep_Quality.1.10 Stress_Level.1.10 Happiness_Index.1.10  
1  U001  44   Male           3.1             7                6                10  
2  U002  30   other           5.1             7                8                10  
3  U003  23   other           7.4             6                7                6  
4  U004  36   Female         5.7             7                8                8  
5  U005  34   Female         7.0             4                7                8  
6  U006  38   Male           6.6             5                7                8  
  
> data_new_users <- data.frame(  
+   user_ID = c("U151", "U152"),  
+   Age = c(29, 33),  
+   Gender = c("Female", "Male"),  
+   Daily_Screen_Time.hrs = c(4.5, 6.2),  
+   Sleep_Quality.1.10 = c(7, 6),  
+   Stress_Level.1.10 = c(5, 7),  
+   Happiness_Index.1.10 = c(8, 7)  
+ )  
> final_dataset <- bind_rows(merged_data, data_new_users)  
> print(head(final_dataset))  
  User_ID Age Gender Daily_Screen_Time.hrs Sleep_Quality.1.10 Stress_Level.1.10 Happiness_Index.1.10  
1  U001  44   Male           3.1             7                6                10  
2  U002  30   other           5.1             7                8                10  
3  U003  23   other           7.4             6                7                6  
4  U004  36   Female         5.7             7                8                8  
5  U005  34   Female         7.0             4                7                8  
6  U006  38   Male           6.6             5                7                8  
7    NA    NA    NA           NA             NA                NA                NA  
8    NA    NA    NA           NA             NA                NA                NA  
9    NA    NA    NA           NA             NA                NA                NA  
10   NA    NA    NA           NA             NA                NA                NA  
11   NA    NA    NA           NA             NA                NA                NA  
12   NA    NA    NA           NA             NA                NA                NA  
  
>
```

**Environment:**

Object	Class	Attributes
data_main	data.frame	500 obs. of 10 variables
data_new_users	data.frame	2 obs. of 7 variables
final_dataset	data.frame	502 obs. of 11 variables
high_exam_score	data.frame	43 obs. of 20 variables
high_study_high	data.frame	727 obs. of 20 variables
high_study_subset	data.frame	3063 obs. of 20 variables
low_sleep_low_mo	data.frame	1226 obs. of 20 variables
Mental_health	data.frame	101 obs. of 11 variables
merged_data	data.frame	500 obs. of 7 variables