

MVLU COLLEGE.

PRACTICAL NO :- 15

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

CODE :-

```
library(dplyr)
```

```
library(readr)
```

```
library(tidyr)
```

```
retail_df <- data.frame(  
  ID = 1:6,  
  Category = c("Electronics", "Home", "Electronics", "Clothing", "Home", "Clothing"),  
  Price = c(500.50, 45.00, 900.00, NA, 300.00, 25.00),  
  Rating = c(4.5, 3.8, 4.9, 4.0, 3.5, 4.2)  
)  
print("--- Data Loaded---")
```

```
#USING summary() (Statistical Summary)
```

```
print("--- OUTPUT OF str() ---")  
str(retail_df)
```

```
#IMPROVING summary() WITH FACTORS
```

```
retail_df$Category <- as.factor(retail_df$Category)  
print("--- OUTPUT OF summary() [After Factor Conversion] ---")
```

```
#Accessing Specific Summaries
```

```
avg_rating <- mean(retail_df$Rating)  
max_price <- max(retail_df$Price, na.rm = TRUE)  
print(paste("Average Rating:", avg_rating))  
print(paste("Highest Price:", max_price))
```

MVLU COLLEGE.



RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Console Terminal Background Jobs

```
[1] "Current Month: 12"
> events_df <- data.frame(
+   Eventcode = c("E101", "E102", "E103", "E104"),
+   Eventdate = c("2022-05-10", "2023-08-19", "2024-11-05", "2025-03-21")
+ )
> print(paste("Current Day:", day(current_time)))
[1] "Current Day: 8"
> print(paste("Current Hour:", hour(current_time)))
[1] "Current Hour: 11"
> print(paste("Current Minute:", minute(current_time)))
[1] "Current Minute: 12"
> library(dplyr)
> library(tidyverse)
> library(tidyR)
> retail_df <- data.frame(
+   ID = 1:6,
+   Category = c("electronics", "Home", "Electronics", "clothing", "Home", "clothing"),
+   Price = c(500, 50, 45.00, 900.00, NA, 300.00, 25.00),
+   Rating = c(4.5, 3.8, 4.9, 4.0, 3.5, 4.2)
+ )
> View(retail_df)
> print("--- Data Loaded---")
[1] "--- Data Loaded---"
> print("--- OUTPUT OF str() ---")
[1] "--- OUTPUT OF str() ---"
> str(retail_df)
'data.frame': 6 obs. of 4 variables:
 $ ID : int 1 2 3 4 5 6
 $ Category: chr "electronics" "Home" "Electronics" "clothing" ...
 $ Price : num 500 50 45 900 NA 300 ...
 $ Rating : num 4.5 3.8 4.9 4 3.5 4.2
> retail_df$category <- as.factor(retail_df$category)
> print("--- OUTPUT OF summary() [After Factor Conversion] ---")
[1] "--- OUTPUT OF summary() [After Factor Conversion] ---"
> #Accessing Specific Summaries
> avg_rating <- mean(retail_df$Rating)
> max_price <- max(retail_df$Price, na.rm = TRUE) # na.rm ignores the missing value
> print(paste("Average Rating:", avg_rating))
[1] "Average Rating: 4.15"
> print(paste("Highest Price:", max_price))
[1] "Highest Price: 900"
> |
```

Environment History Connections Tutorial

R Global Environment

Data

- clean_exact 5 obs. of 3 variables
- combined_data 10150 obs. of 2 variables
- df 497 obs. of 9 variables
- duplicates_report 2 obs. of 4 variables
- events_df 4 obs. of 2 variables
- flower_clean 10000 obs. of 2 variables
- flower_df 10000 obs. of 4 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
- iris_clean 150 obs. of 2 variables
- iris_df 150 obs. of 6 variables
- location_pivot 497 obs. of 10 variables
- long_df 2485 obs. of 6 variables
- low_sleep_low_mo_... 1226 obs. of 20 variables
- mental_health 101 obs. of 11 variables
- my_data 1000 obs. of 14 variables
- processed_data 4 obs. of 11 variables
- retail_df 6 obs. of 4 variables
- sales_data 1000 obs. of 14 variables
- sales_df 7 obs. of 3 variables
- school_type_file_ 0 obs. of 20 variables
- sleep_or_extracu_ 4363 obs. of 20 variables
- student 6607 obs. of 20 variables
- student_Mental.h... 101 obs. of 11 variables
- unique_customers 4 obs. of 3 variables
- wide_df 497 obs. of 9 variables

Values

avg_rating	4.15
current_time	2025-12-08 11:12:04 IST
max_price	900

Files Plots Packages Help Viewer Presentation

11:36 ENG IN 08-12-2025