

MVLU COLLEGE

R PRACTICAL 15

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

Output :

The screenshot shows the RStudio interface with the following code in the Console tab:

```

> # 1. SETUP: Create Sample Data
> # we create a data frame with mixed data types (Numeric, Character, Logical, NA)
> retail_df <- data.frame(
+   ID = 1:6,
+   Category = c("Electronics", "food", "Computer", "bike accessories", "Home", "Clothing"),
+   Price = c(500.50, 45.00, 900.00, NA, 300.00, 25.00), # Note the NA
+   In_Stock = c(TRUE, TRUE, FALSE, TRUE, FALSE, TRUE),
+   Rating = c(4.7, 4.0, 5.0, 3.1, 2.9, 3.9),
+   stringsAsFactors = FALSE
+ )
> print("--- Data Loaded ---")
[1] "--- Data Loaded ---"
> print(retail_df)
#>   ID Category Price In_Stock Rating
#> 1  1 Electronics 500.5 TRUE    4.7
#> 2  2 food      45.0 TRUE    4.0
#> 3  3 Computer   900.0 FALSE   5.0
#> 4  4 bike accessories NA     TRUE    3.1
#> 5  5 Home      300.0 FALSE   2.9
#> 6  6 Clothing   25.0 TRUE    3.9
#> 
#> # 2. USING str() (Structure)
#> # Purpose: Compactly display the internal structure of the R object.
> print("--- OUTPUT OF str() ---")
[1] "--- OUTPUT OF str() ---"
> str(retail_df)
#> 'data.frame': 6 obs. of 5 variables:
#> #>   $ ID : int 1 2 3 4 5 6
#> #>   $ Category: chr "Electronics" "food" "Computer" "bike accessories" ...
#> #>   $ Price : num 500.45 900 NA 300 ...
#> #>   $ In_Stock: logi TRUE TRUE FALSE TRUE FALSE TRUE
#> #>   $ Rating : num 4.7 4.5 3.1 2.9 3.9
#> 
#> # 3. USING summary() (Statistical Summary)
#> # Purpose: detailed summary statistics for each column.
> print("--- OUTPUT OF summary() [Before Factor Conversion] ---")
[1] "--- OUTPUT OF summary() [Before Factor Conversion] ---"

```

The Environment pane shows the following objects:

- runs_by_opponent: 8 obs. of 5 variables
- selected_cols: 106 obs. of 9 variables
- starts_with_p: 106 obs. of 9 variables
- team_lookup: 8 obs. of 2 variables
- tidy_cricket: 106 obs. of 11 variables

The Files pane shows the following files:

- tuel_sales.csv
- GIS DataBase
- ISExpress
- My Music
- My Pictures
- My Videos
- My Web Sites
- R
- SM PRACTICAL - II 20.txt
- SM PRACTICAL - II.txt
- total_bills.csv
- Virtual Machines
- Visual Studio 2022
- WindowsPowerShell

The screenshot shows the RStudio interface with the following code in the Console tab:

```

'data.frame': 6 obs. of 5 variables:
$ ID : int 1 2 3 4 5 6
$ Category: chr "Electronics" "food" "Computer" "bike accessories" ...
$ Price : num 500.45 900 NA 300 ...
$ In_Stock: logi TRUE TRUE FALSE TRUE FALSE TRUE
$ Rating : num 4.7 4.5 3.1 2.9 3.9
#> 
#> # 3. USING summary() (Statistical Summary)
#> # Purpose: detailed summary statistics for each column.
> print("--- OUTPUT OF summary() [Before Factor Conversion] ---")
[1] "--- OUTPUT OF summary() [Before Factor Conversion] ---"
> summary(retail_df)
#>   ID       Category      Price     In_Stock   Rating
#> Min.   :1.00   Length:6   Min.   :25.0   Mode :logical  Min.  :2.900
#> 1st Qu.:2.25  Class  :character 1st Qu.:45.0   FALSE:2   1st Qu.:3.300
#> Median :3.50  Mode   :character Median :300.0   TRUE :4    Median :3.950
#> Mean   :3.50                           Mean  :354.1   Mean  :3.933
#> 3rd Qu.:4.75                          3rd Qu.:500.5  3rd Qu.:4.525
#> Max.   :6.00                           Max.  :900.0   Max.  :5.000
#> 
#> NA's   :1
#> 
#> # 4. IMPROVING summary() WITH FACTORS
#> # by default category is character here; convert to factor to get counts per level.
> print("--- Category counts (before factor conversion) ---")
[1] "--- Category counts (before factor conversion) ---"
> print(table(retail_df$category))
#> 
#> bike accessories      Clothing      Computer      Electronics      food
#>                  1              1              1              1              1
#> 
#> retail_df$category <- as.factor(retail_df$category)
> print("--- OUTPUT OF summary() [After Factor Conversion] ---")
[1] "--- OUTPUT OF summary() [After Factor Conversion] ---"
> summary(retail_df)
#>   ID       Category      Price     In_Stock   Rating
#> Min.   :1.00   bike.accessories:1   Min.   :25.0   Mode :logical  Min.  :2.900
#> 1st Qu.:2.25
#> Median :3.50
#> Mean   :3.50
#> 3rd Qu.:4.75
#> Max.   :6.00
#> 
```

The Environment pane shows the same objects as the previous screenshot.

MVLU COLLEGE R PRACTICAL 15

The screenshot shows the RStudio interface with the following details:

- File Menu:** File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, Help.
- Toolbar:** Includes icons for file operations like Open, Save, Print, and a search bar labeled "Go to file/function".
- Source Tab:** Shows the R code being run in the console.
- Console Tab:** Displays the R session output, including summary statistics for categories like bike accessories, clothing, computer, electronics, and food.
- Background Jobs:** Shows no active jobs.
- Environment Tab:** Lists global environment objects and their details.
- History Tab:** Shows the history of the R session.
- Connections Tab:** Shows connections to databases and other services.
- Tutorial Tab:** Provides access to R documentation.
- File Browser:** Shows the local file system structure under the "Home" directory.

Hariprasad Vishwakarma
S126
SYCS