

SHETH L.U.J AND SIR M.V COLLEGE

PRACTICAL NO:15

AIM: Reshaping data using pivot_longer()/pivot_wider() (R).

INPUT:

```
library(dplyr)
```

```
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),
```

```
  In_Stock = c(TRUE, TRUE, FALSE, TRUE, FALSE, TRUE),
```

```
  Rating = c(4.5, 3.8, 4.9, 4.0, 3.5, 4.2)
```

```
)
```

```
print("--- Data Loaded ---")
```

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

```
str(retail_df)
```

```
print("--- OUTPUT OF summary() [Before Factor Conversion] ---")
```

```
summary(retail_df)
```

SHETH L.U.J AND SIR M.V COLLEGE

PRACTICAL NO:15

```
retail_df$Category <- as.factor(retail_df$Category)
```

```
print("--- OUTPUT OF summary() [After Factor Conversion] ---")
```

```
summary(retail_df)
```

```
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))
```

OUTPUT:

The screenshot shows the RStudio interface with the following details:

- Console:** Displays the R session history with the following output:

```
> 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
> library(tidyverse)
> library(tidyverse)
> retail_df <- data.frame(
+   ID : int,
+   Category : c("Electronics", "Home", "Electronics", "Clothing", "Home", "Clothing"),
+   Price = c(500.50, 45.00, 900.00, NA, 300.00, 25.00),
+   In_Stock = c(TRUE, TRUE, FALSE, TRUE, FALSE, TRUE),
+   Rating = c(4.5, 3.8, 4.9, 4.0, 3.5, 4.2)
+ )
> print("--- Data Loaded ---")
[1] "--- Data Loaded ---"
> 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" "Home" "Electronics" "Clothing" ...
 $ Price    : num 500 45 900 NA 300 ...
 $ In_Stock: logi TRUE TRUE FALSE TRUE FALSE TRUE
 $ Rating   : num 4.5 3.8 4.9 4 3.5 4.2
> 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. :3.500
1st Qu.:2.25 Class :character 1st Qu.:45.0   FALSE:12   1st Qu.:13.850
Median :3.50 Mode  :character Median :300.0   TRUE :4    Median :14.100
Mean   :3.50   Mean   :354.1   Mean   :14.150
3rd Qu.:4.75  3rd Qu.:500.5  3rd Qu.:14.425
```
- Environment:** Shows the global environment with two variables defined:

avg_rating	4.15
max_price	900
- File Explorer:** Shows the project structure with files like RData, Rhistory, Global_Health (1).csv, Global_Health (1).xlsx, and Visual Studio 2022.

SHETH L.U.J AND SIR M.V COLLEGE

PRACTICAL NO:15

The screenshot shows the RStudio interface with the following details:

- Source Pane:** Displays an R script with code related to data loading, printing, and summarizing a dataset named `retail_df`. The script includes calculations for average rating and maximum price.
- Environment Pane:** Shows a summary table for the `retail_df` dataset, which has 6 observations and 5 variables. The table includes columns for ID, Category, Price, In_Stock, and Rating, with detailed statistics for each.
- Files Pane:** A file browser showing various files and folders on the local machine, including RData, Rhistory, and several CSV and XLSX files related to data analysis.