

MVLU COLLEGE.

Practical No :- 11

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

CODE :-

```
library(dplyr)
library(tidyr)

df <-
read.csv("C:/Users/itlab/Downloads/S100/Cleaned_Data_Science_Student_Marks.csv",
         na.strings = c("", "NA"))

# Add RowID and keep useful columns
df <- df %>%
  mutate(RowID = row_number()) %>%
  select(RowID, student_id, location, age,
         sql_marks, excel_marks, python_marks,
         power_bi_marks, english_marks)

print("--- 1. Original Wide Data ---")
print(head(df))

# 2. PIVOT_LONGER (Wide → Long)
# Convert all marks columns into Subject + Score format

long_df <- df %>%
  pivot_longer(
    cols = c(sql_marks, excel_marks, python_marks,
             power_bi_marks, english_marks),
    names_to = "Subject",
    values_to = "Score"
  )

print("--- 2. Long Format (pivot_longer) ---")
print(head(long_df, 10))

# 3. PIVOT_WIDER (Long → Wide)

wide_df <- long_df %>%
```

NANDINI PANDIT S100
DATA ANALYSIS WITH SAS/SPSS/R.

MVLU COLLEGE.

```
pivot_wider(  
  names_from = Subject,  
  values_from = Score  
)
```

```
print("--- 3. Wide Format (Back to Original) ---")  
print(head(wide_df))
```

Example: show SQL marks for every RowID, pivoting locations as columns

```
location_pivot <- df %>%  
  select(RowID, location, sql_marks) %>%  
  pivot_wider(  
    names_from = location,  
    values_from = sql_marks  
  )
```

```
print("--- 4. Location Pivot (SQL Marks by Location) ---")  
print(head(location_pivot))
```

OUTPUT :-

The screenshot shows the RStudio interface with the following components:

- Source Pane:** Contains R code for data manipulation. It includes loading libraries (dplyr, tidyr), reading a CSV file, and performing pivot operations. The code is as follows:

```
> library(dplyr)  
> library(tidyr)  
> df <- read.csv("Cleaned_Data_Science_Student_Marks.csv",  
+               na.strings = c("", "NA"))  
  
Error in file(file, "rt") : cannot open the connection  
In addition: warning message:  
In file(file, "rt") :  
cannot open file 'Cleaned_Data_Science_Student_Marks.csv': No such file or directory  
  
> df <- read.csv("C:/Users/itlab/Downloads/S100/Cleaned_Data_Science_Student_Marks.csv",  
+               na.strings = c("", "NA"))  
> # Add RowID and keep useful columns  
> df <- df %>%  
+   mutate(RowID = row_number()) %>%  
+   select(RowID, student_id, location, age,  
+         sql_marks, excel_marks, python_marks,  
+         power_bi_marks, english_marks)  
> print(head(df))  
# A tibble: 10 x 6  
#   RowID student_id location age  sql_marks  
#   <int> <int> <chr> <int> <chr>  
1     1         4 Sydney   24 sql_marks  
2     2         5 Tokyo    24 sql_marks  
3     3         6 Berlin   22 sql_marks  
4     4         7 London   23 sql_marks  
5     5         8 Tokyo    22 sql_marks  
6     6         9 Toronto  20 sql_marks  
  
> long_df <- df %>%  
+   pivot_longer(  
+     cols = c(sql_marks, excel_marks, python_marks,  
+             power_bi_marks, english_marks),  
+     names_to = "Subject",  
+     values_to = "Score"  
+   )  
> view(long_df)  
> print("--- 2. Long Format (pivot_longer) ---")  
[1] "--- 2. Long Format (pivot_longer) ---"  
> print(head(long_df, 10))  
# A tibble: 10 x 6  
#   RowID student_id location age  Subject      Score  
#   <int> <int> <chr> <int> <chr> <int>  
1     1         4 Sydney   24 sql_marks    95  
2     2         5 Tokyo    24 excel_marks   99  
3     3         6 Berlin   22 python_marks 99  
4     4         7 London   23 power_bi_marks 74  
5     5         8 Tokyo    22 english_marks 72  
6     6         9 Toronto  20 sql_marks    88  
7     7         1 Sydney   24 excel_marks   87  
8     8         2 Tokyo    24 python_marks 86  
9     9         3 Berlin   22 power_bi_marks 79  
10    10         4 London   23 english_marks 85
```
- Environment Pane:** Lists the objects in the global environment, including 'df' (497 obs. of 9 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), '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), 'sales_data' (1000 obs. of 14 variables), 'school_type_filt' (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), and 'wide_df' (497 obs. of 9 variables).

MVLU COLLEGE.

The image displays two screenshots of the RStudio interface, illustrating data manipulation steps using the `tidyverse` packages.

Top Screenshot:

- Console:** Shows the initial data structure and the first transformation.

```
R - R452 ~/  
> print(head(long_df, 10))  
# A tibble: 10 x 6  
  RowID student_id location age subject      Score  
  <int>   <int> <chr>   <int> <chr>    <int>  
1     1       4 Sydney    24 sql_marks 95  
2     1       4 Sydney    24 excel_marks 99  
3     1       4 Sydney    24 python_marks 87  
4     1       4 Sydney    24 power_bi_marks 82  
5     1       4 Sydney    24 english_marks 75  
6     2       5 Tokyo     24 sql_marks 99  
7     2       5 Tokyo     24 excel_marks 95  
8     2       5 Tokyo     24 python_marks 89  
9     2       5 Tokyo     24 power_bi_marks 86  
10    2       5 Tokyo     24 english_marks 82  
  
> wide_df <- long_df %>%  
+ pivot_wider(  
+   names_from = subject,  
+   values_from = score  
+ )  
> view(wide_df)  
> print("---- 3. Wide Format (back to original) ----")  
[1] "---- 3. Wide Format (back to original) ----"  
> print(head(wide_df))  
# A tibble: 6 x 9  
  RowID student_id location age sql_marks excel_marks python_marks power_bi_marks english_marks  
  <int>   <int> <chr>   <int>   <int>   <int>   <int>   <int>   <int>  
1     1       4 Sydney    24     95     99     87     82     75  
2     2       5 Tokyo     24     99     89     86     82  
3     3       6 Berlin    22     72     70     99     79     77  
4     4       7 London    23     97     90     74     72     85  
5     5       8 Tokyo     22     91     71     79     80     75  
6     6       9 Toronto    20     93     88     75     93     72  
  
> location_pivot <- df %>%  
+ select(RowID, location, sql_marks) %>%  
+ pivot_wider(  
+   names_from = location,  
+   values_from = sql_marks  
+ )  
> print("---- 4. Location Pivot (SQL Marks by Location) ----")  
[1] "---- 4. Location Pivot (SQL Marks by Location) ----"  
> print(head(location_pivot))  
# A tibble: 6 x 10  
  RowID Sydney Tokyo Berlin London Toronto Melbourne Paris 'Los Angeles' 'New York'  
  <int>   <int> <int>   <int>   <int>   <int>   <int>   <int>   <int>  
1     1     95    NA    NA    NA    NA    NA    NA    NA  
2     2     99    NA    NA    NA    NA    NA    NA    NA  
3     3     72    NA    NA    NA    NA    NA    NA    NA  
4     4     97    NA    NA    NA    NA    NA    NA    NA  
5     5     91    NA    NA    NA    NA    NA    NA    NA  
6     6     93    NA    NA    NA    NA    NA    NA    NA
```
- Environment:** Lists the data frames in the workspace: `df` (497 obs. of 9 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), `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), `sales_data` (1000 obs. of 14 variables), `school_type_filt` (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), and `wide_df` (497 obs. of 9 variables).

Bottom Screenshot:

- Console:** Shows the second transformation, pivoting the wide format back to a long format.

```
R - R452 ~/  
6     2       5 Tokyo     24 sql_marks 99  
7     2       5 Tokyo     24 excel_marks 89  
8     2       5 Tokyo     24 python_marks 86  
9     2       5 Tokyo     24 power_bi_marks 86  
10    2       5 Tokyo     24 english_marks 82  
  
> wide_df <- long_df %>%  
+ pivot_wider(  
+   names_from = subject,  
+   values_from = score  
+ )  
> view(wide_df)  
> print("---- 3. Wide Format (back to original) ----")  
[1] "---- 3. Wide Format (back to original) ----"  
> print(head(wide_df))  
# A tibble: 6 x 9  
  RowID student_id location age sql_marks excel_marks python_marks power_bi_marks english_marks  
  <int>   <int> <chr>   <int>   <int>   <int>   <int>   <int>   <int>  
1     1       4 Sydney    24     95     99     87     82     75  
2     2       5 Tokyo     24     99     89     86     82  
3     3       6 Berlin    22     72     70     99     79     77  
4     4       7 London    23     97     90     74     72     85  
5     5       8 Tokyo     22     91     71     79     80     75  
6     6       9 Toronto    20     93     88     75     93     72  
  
> location_pivot <- df %>%  
+ select(RowID, location, sql_marks) %>%  
+ pivot_wider(  
+   names_from = location,  
+   values_from = sql_marks  
+ )  
> print("---- 4. Location Pivot (SQL Marks by Location) ----")  
[1] "---- 4. Location Pivot (SQL Marks by Location) ----"  
> print(head(location_pivot))  
# A tibble: 6 x 10  
  RowID Sydney Tokyo Berlin London Toronto Melbourne Paris 'Los Angeles' 'New York'  
  <int>   <int> <int>   <int>   <int>   <int>   <int>   <int>   <int>  
1     1     95    NA    NA    NA    NA    NA    NA    NA  
2     2     99    NA    NA    NA    NA    NA    NA    NA  
3     3     72    NA    NA    NA    NA    NA    NA    NA  
4     4     97    NA    NA    NA    NA    NA    NA    NA  
5     5     91    NA    NA    NA    NA    NA    NA    NA  
6     6     93    NA    NA    NA    NA    NA    NA    NA  
> |
```
- Environment:** The same list of data frames is shown, with `wide_df` still present.