

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.

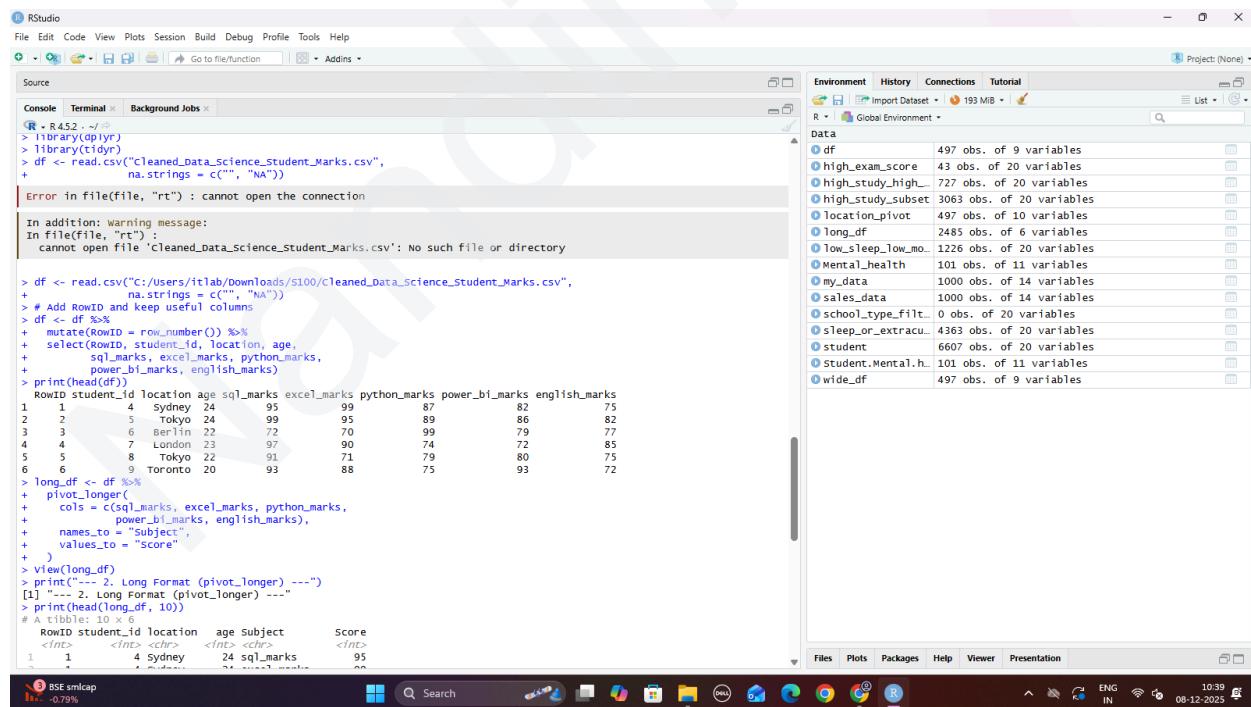
```
pivot_wider(  
  names_from = Subject,  
  values_from = Score  
)  
  
print("--- 3. Wide Format (Back to Original) ---")  
print(head(wide_df))
```

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 :-



```
RStudio  
File Edit Code View Plots Session Build Debug Profile Tools Help  
Console Terminal > Background Jobs  
Go to file/function Project: (None)  
Source  
Console Terminal > Background Jobs  
> R 4.2.2 ~/  
> 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("~/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_b1_marks, english_marks)  
> print(head(df))  
RowID student_id location age sql_marks excel_marks python_marks power_b1_marks english_marks  
1 1 4 Sydney 24 95 99 87 82 75  
2 2 5 Tokyo 24 99 95 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  
  
> long_df <- df %>%  
+   pivot_longer(  
+     cols = c(sql_marks, excel_marks, python_marks,  
+               power_b1_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 × 6  
  RowID student_id location age Subject Score  
    <int> <chr> <chr> <dbl> <chr> <dbl>  
1 1 4 Sydney 24 sql_marks 95  
2 2 5 Tokyo 24 excel_marks 99  
3 3 6 Berlin 22 python_marks 72  
4 4 7 London 23 power_b1_marks 97  
5 5 8 Tokyo 22 english_marks 91  
6 6 9 Toronto 20 sql_marks 93  
7 7 10 Paris 25 excel_marks 88  
8 8 11 New_York 26 python_marks 75  
9 9 12 Sydney 27 power_b1_marks 99  
10 10 13 Berlin 28 english_marks 71  
# ... with 10 more rows  
Files Plots Packages Help Viewer Presentation
```

MVLU COLLEGE.

RStudio

```
> print(head(long_df, 10))
#> # A tibble: 10 × 6
#>   RowID student_id location  age subject    score
#>   <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 × 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      95      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      91      71      79      80
#> 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 × 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     NA    99    NA    NA    NA    NA    NA    NA
#> 3     3     NA    NA    72    NA    NA    NA    NA    NA
#> 4     4     NA    NA    97    NA    NA    NA    NA    NA
#> 5     5     NA    91    NA    NA    NA    NA    NA    NA
#> 6     6     NA    NA    NA    93    NA    NA    NA    NA
```

BSE smicap

RStudio

```
> print(head(long_df, 10))
#> # A tibble: 10 × 6
#>   RowID student_id location  age subject    score
#>   <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 × 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      95      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      91      71      79      80
#> 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 × 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     NA    99    NA    NA    NA    NA    NA    NA
#> 3     3     NA    NA    72    NA    NA    NA    NA    NA
#> 4     4     NA    NA    97    NA    NA    NA    NA    NA
#> 5     5     NA    91    NA    NA    NA    NA    NA    NA
#> 6     6     NA    NA    NA    93    NA    NA    NA    NA
```

BSE smicap