How can I perform a cross join in R, and can you provide an example?

Authored by

stats writer

June 29, 2024

A cross join in R is a method used to combine two data frames by creating all possible combinations of rows between them. This can be achieved by using the "merge" function with the "by = NULL" argument. An example of this would be merging two data frames with 3 and 4 rows respectively, resulting in a new data frame with 12 rows (3 x 4). This allows for a comprehensive analysis of data by including all possible combinations.

Do a Cross Join in R (With Example)

The easiest way to perform a cross join in R is to use the crossing() function from the package:

library(tidyr)

#perform cross join on df1 and df2
crossing(df1, df2)

The following example shows how to use this function in practice.

Example: Perform Cross Join in R

Suppose we have the following two data frames in R:

#define first data frame
df1 = data.frame(team1=c('A', 'B', 'C', 'D'),
points=c(18, 22, 19, 14))

df1

team1 points
1 A 18
2 B 22
3 C 19
4 D 14

#define second data frame
df2 = data.frame(team2=c('A', 'B', 'F'),
assists=c(4, 9, 8))

df2

team2 assists
1 A 4
2 B 9
3 F 8

We can use the crossing() function from the tidyr package to perform a cross join on these two data frames:

library(tidyr)

#perform cross join
cross <- crossing(df1, df2)

```
#view result
cross

# A tibble: 12 x 4
team1 points team2 assists

1 A 18 A 4
2 A 18 B 9
3 A 18 F 8
4 B 22 A 4
5 B 22 B 9
6 B 22 F 8
7 C 19 A 4
8 C 19 B 9
9 C 19 F 8
10 D 14 A 4
11 D 14 B 9
12 D 14 F 8
```


The result is a data frame that contains every possible combination of rows from each data frame.

For example, the first row of the first data frame contains team A and 18 points. This row is matched with every single row in the second data frame.

Next, the second row of the first data frame contains team B and 22 points. This row is also matched with every single row in the second data frame.

The end result is a data frame with 12 rows.

Additional Resources