How can I perform a left join in R? Can you provide some examples?

Authored by

stats writer

May 2, 2024

A left join in R is a type of data merging operation that combines two datasets based on a common key variable while retaining all the rows from the first (left) dataset. This means that all the observations from the left dataset will be included in the resulting dataset, while only matching observations from the second dataset will be added. This type of join is useful when you want to add additional information from one dataset to another without losing any data. To perform a left join in R, you can use the "merge" function or the "left_join" function from the "dplyr" package. Here are some examples of how to perform a left join in R:

1. Using the "merge" function:
merged_data <- merge(data1, data2, by = "key_variable", all.x = TRUE)

2. Using the "left_join" function:
merged_data <- left_join(data1, data2, by = "key_variable")

In both cases, "data1" and "data2" represent the two datasets to be merged, and "key_variable" is the common variable used for merging. The resulting dataset, "merged_data", will contain all the rows from "data1" and only the matching rows from "data2".

Do a Left Join in R (With Examples)

You can use the merge() function to perform a left join in base R:

#left join using base R
merge(df1,df2, all.x=TRUE)

You can also use the left_join() function from the package to perform a left join:

#left join using dplyr
dplyr::left_join(df2, df1)

Note: If you're working with extremely large datasets, the left_join() function will tend to be faster than the merge() function.

The following examples show how to use each of these functions in practice with the following data frames:

```
#define first data frame
df1 <- data.frame(team=c('Mavs', 'Hawks', 'Spurs', 'Nets'),
points=c(99, 93, 96, 104))

df1

team points
1 Mavs 99
2 Hawks 93
3 Spurs 96
4 Nets 104

#define second data frame
df2 <- data.frame(team=c('Mavs', 'Hawks', 'Spurs', 'Nets'),
rebounds=c(25, 32, 38, 30),
assists=c(19, 18, 22, 25))
```

df2

team rebounds assists
1 Mavs 25 19
2 Hawks 32 18
3 Spurs 38 22
4 Nets 30 25

Example 1: Left Join Using Base R

We can use the merge() function in base R to perform a left join, using the 'team' column as the column to join on:

```
#perform left join using base R
merge(df1, df2, by='team', all.x=TRUE)
```

team points rebounds assists
1 Hawks 93 32 18
2 Mavs 99 25 19
3 Nets 104 30 25
4 Spurs 96 38 22

Example 2: Left Join Using dplyr

We can use the left_join() function from the dplyr

package to perform a left join, using the 'team' column as the column to join on:

```
library(dplyr)

#perform left join using dplyr
left_join(df1, df2, by='team')
```

```
team points rebounds assists
1 Mavs 99 25 19
2 Hawks 93 32 18
3 Spurs 96 38 22
4 Nets 104 30 25
```

One difference you'll notice between these two functions is that the merge() function automatically sorts the rows alphabetically based on the column you used to perform the join.

Conversely, the left_join() function preserves the original order of the rows from the first data frame.