How can I merge data frames by their column names using the “R” programming language?

Authored by

stats writer

June 29, 2024

Merging data frames in the "R" programming language refers to the process of combining two or more data frames based on their common column names. This can be achieved using the "merge()" function, which allows the user to specify the columns to be used as the merging criteria. The resulting merged data frame will contain all the columns from both original data frames, with matching values in the specified columns merged into a single row. This process is useful for integrating and analyzing data from multiple sources, providing a comprehensive view of the data.

R: Merge Data Frames by Column Names

You can use the following methods to merge data frames by column names in R:

Method 1: Merge Based on One Matching Column Name

merge(df1, df2, by='var1')

Method 2: Merge Based on One Unmatched Column Name

merge(df1, df2, by.x='var1', by.y='variable1')

Method 3: Merge Based on Multiple Matching Column Names

merge(df1, df2, by=c('var1', 'var2'))

Method 4: Merge Based on Multiple Unmatched Column

Names

merge(df1, df2, by.x=c('var1', 'var2'), by.y=c('variable1', 'variable2'))

The following examples show how to use each method in practice.

Example 1: Merge Based on One Matching Column Name

The following code shows how to merge two data frames in R based on one matching column name:

```
#define data frames
df1 <- data.frame(team=c('A', 'B', 'C', 'D'),
points=c(88, 98, 104, 100))

df2 <- data.frame(team=c('A', 'B', 'C', 'D'),
rebounds=c(22, 31, 29, 20))

#merge based on one column with matching name
merge(df1, df2, by='team')

team points rebounds
1 A 88 22
2 B 98 31
```

3 C 104 29

4 D 100 20

The result is one data frame that matched rows in each data frame using the team column.

Example 2: Merge Based on One Unmatched Column Name

The following code shows how to merge two data frames in R based on one unmatched column name:

```
#define data frames
df1 <- data.frame(team=c('A', 'B', 'C', 'D'),
points=c(88, 98, 104, 100))

df2 <- data.frame(team_name=c('A', 'B', 'C', 'D'),
rebounds=c(22, 31, 29, 20))

#merge based on one column with unmatched name
merge(df1, df2, by.x='team', by.y='team_name')
```

team points rebounds

1 A 88 22

2 B 98 31

3 C 104 29

4 D 100 20

Example 3: Merge Based on Multiple Matching Column Names

The following code shows how to merge two data frames in R based on multiple matching column names:

```
#define data frames
df1 <- data.frame(team=c('A', 'A', 'B', 'B'),
position=c('G', 'F', 'G', 'F'),
points=c(88, 98, 104, 100))

df2 <- data.frame(team=c('A', 'A', 'B', 'B'),
position=c('G', 'F', 'G', 'F'),
rebounds=c(22, 31, 29, 20))

#merge based on multiple columns with matching names
merge(df1, df2, by=c('team', 'position'))

team position points rebounds
1 A F 98 31
2 A G 88 22
3 B F 100 20
4 B G 104 29
```

The result is one data frame that matched rows in each

data frame using the team and position column in each data frame.

Example 4: Merge Based on Multiple Unmatched Column Names

The following code shows how to merge two data frames in R based on multiple unmatched column names:

```
#define data frames
df1 <- data.frame(team=c('A', 'A', 'B', 'B'),
position=c('G', 'F', 'G', 'F'),
points=c(88, 98, 104, 100))

df2 <- data.frame(team_name=c('A', 'A', 'B', 'B'),
position_name=c('G', 'F', 'G', 'F'),
rebounds=c(22, 31, 29, 20))

#merge based on multiple columns with matching names
merge(df1, df2, by.x=c('team', 'position'),
by.y=c('team_name', 'position_name'))

team position points rebounds
1 A F 98 31
2 A G 88 22
```

3 B F 100 20

4 B G 104 29

The result is one data frame that matched rows using the team and position columns in the first data frame and the team_name and position_name columns in the second data frame.

Additional Resources

The following tutorials explain how to perform other common functions related to data frames in R: