Does the data frame contain a row that exists in another data frame?

Authored by

stats writer

June 23, 2024

The question of whether a data frame contains a row that exists in another data frame refers to the comparison of two data frames to determine if there are any identical rows shared between them. This comparison can be useful in identifying any overlap or similarities between the data sets. By checking for the existence of shared rows, it can provide insights on the relationships and connections between the data frames.

R: Check if Row in One Data Frame Exists in Another

You can use the following syntax to add a new column to a data frame in R that shows if each row exists in another data frame:

df1$exists <- do.call(paste0, df1) %in% do.call(paste0, df2)

This particular syntax adds a column called exists to the data frame called df1 that contains TRUE or FALSE to indicate if each row in df1 exists in another data frame called df2.

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

Example: Check if Row in One Data Frame Exists in Another in R

Suppose we have the following two data frames in R:

#create first data frame

```
df1 <- data.frame(team=c('A', 'B', 'C', 'D', 'E'),
points=c(12, 15, 22, 29, 24))

#view first data frame
df1

team points
1 A 12
2 B 15
3 C 22
4 D 29
5 E 24

#create second data frame
df2 <- data.frame(team=c('A', 'D', 'F', 'G', 'H'),
points=c(12, 29, 15, 19, 10))

#view second data frame
df2

team points
1 A 12
2 D 29
3 F 15
4 G 19
```

5 H 10

We can use the following syntax to add a column called exists to the first data frame that shows if each row exists in the second data frame:

```
#add new column to df1 that shows if row exists in df2
df1$exists <- do.call(paste0, df1) %in% do.call(paste0, df2)

#view updated data frame
df1
```

```
  team points exists
1    A     12   TRUE
2    B     15  FALSE
3    C     22  FALSE
4    D     29   TRUE
5    E     24  FALSE
```

The new exists column shows if each row in the first data frame exists in the second data frame.

From the output we can see:

The first row in df1 does exists in df2.The second row in df1 does not exist in df2.The third row in df1 does not exist in df2.

And so on.

Note that you can also use is.numeric() to display 1's and 0's instead of TRUE or FALSE in the exists column:

```
#add new column to df1 that shows if row exists in df2
df1$exists <- as.numeric(do.call(paste0, df1) %in% do.call(paste0, df2))

#view updated data frame
df1

team points exists
1 A 12 1
2 B 15 0
3 C 22 0
4 D 29 1
5 E 24 0
```

A value of 1 indicates that the row in the first data frame exists in the second.

Conversely, a value of 0 indicates that the row in the first data frame does not exist in the second.