How can we combine rows with the same column values in R?

Authored by

stats writer

June 27, 2024

In R, rows with the same column values can be combined using the "aggregate" function. This function allows users to group the rows by their column values and then perform a specific operation, such as sum or mean, on the remaining columns. This results in a single row for each unique combination of column values. Alternatively, the "merge" function can be used to combine rows with the same column values from two different data frames, resulting in a new data frame with all the columns from both original data frames. Overall, these methods allow for efficient and organized merging of rows with the same column values in R.

Combine Rows with Same Column Values in R

You can use the following basic syntax to combine rows with the same column values in a data frame in R:

```
library(dplyr)

df %>%
group_by(group_var1, group_var2) %>%
summarise(across(c(values_var1, values_var2), sum))
```

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

Example: Combine Rows with Same Column Values in R

Suppose we have the following data frame that contains information about sales and returns made by various employees at a company:

```
#create data frame
```

```
df <- data.frame(id=c(101, 101, 102, 103, 103, 103),
employee=c('Dan', 'Dan', 'Rick', 'Ken', 'Ken', 'Ken'),
sales=c(4, 1, 3, 2, 5, 3),
returns=c(1, 2, 2, 1, 3, 2))

#view data frame
df
```

id employee sales returns
1 101 Dan 4 1
2 101 Dan 1 2
3 102 Rick 3 2
4 103 Ken 2 1
5 103 Ken 5 3
6 103 Ken 3 2

We can use the following syntax to combine rows that have the same value in the id and employee columns and then aggregate the remaining columns:

```
library(dplyr)

#combine rows with same value for id and employee
and aggregate remaining columns
df %>%
```

```
group_by(id, employee) %>%
summarise(across(c(sales, returns), sum))

# A tibble: 3 x 4
# Groups: id
id employee sales returns

1 101 Dan 5 3
2 102 Rick 3 2
3 103 Ken 10 6
```

The result is a data frame that combines all of the rows in the original data frame that had the same value in the id and employee columns and then calculates the sum of values in the sales and returns columns.

Note: We chose to aggregate the sales and returns columns using the sum function, but you can aggregate by another metric such as the mean if you'd like.

Related: