How can I count the number of unique values in a column in R?

Authored by

stats writer

June 25, 2024

To count the number of unique values in a column in R, you can use the function "length(unique(column_name))". This will return the total number of distinct values in the specified column.

Count Unique Values in Column in R

You can use the following methods to count the number of unique values in a column of a data frame in R:

Method 1: Using Base R

length(unique(df$my_column))

Method 2: Using dplyr

library(dplyr)

n_distinct(df$my_column)

The following examples show how to use each method in practice with the following data frame:

#create data frame
df <- data.frame(team=c('A', 'A', 'A', 'A', 'B', 'B', 'C', 'C', 'D'),
points=c(10, 13, 14, 14, 18, 19, 20, 20, 22))

#view data frame
df

team points
1 A 10
2 A 13
3 A 14
4 A 14
5 B 18
6 B 19
7 C 20
8 C 20
9 D 22

Method 1: Count Unique Values in Column Using Base R

The following code shows how to count the number of unique values in the points column of the data frame using functions from base R:

#count unique values in points column
length(unique(df$points))

7

There are 7 unique value in the points column.

To count the number of unique values in each column of the data frame, we can use the sapply() function:

```
#count unique values in each column
sapply(df, function(x) length(unique(x)))

team points
4 7
```

From the output we can see:

There are 7 unique values in the points column.There are 4 unique values in the team columm.

Method 2: Count Unique Values in Column Using dplyr

The following code shows how to count the number of distinct values in the points column using the n_distinct() function from the dplyr package:

```
library(dplyr)

#count unique values in points column
n_distinct(df$points)
```

7

There are 7 unique value in the points column.

To count the number of unique values in each column of the data frame, we can use the sapply() function:

```
library(dplyr)

#count unique values in each column
sapply(df, function(x) n_distinct(x))

team points
4 7
```

From the output we can see:

There are 7 unique values in the points column.There are 4 unique values in the team columm.

Notice that these results match the ones from the base R method.