How can I add a count column to a Pandas DataFrame?

Authored by

stats writer

June 28, 2024

Adding a count column to a Pandas DataFrame can be achieved by using the built-in function "count" which calculates the number of non-null values in each row or column. This function can be applied to a specific column or the entire DataFrame, and the resulting count values can be stored in a new column using the "assign" method. This process allows for easy tracking and analysis of the number of values present in a DataFrame, providing useful insights and aiding in data manipulation.

Add a Count Column to a Pandas DataFrame

You can use the following basic syntax to add a 'count' column to a pandas DataFrame:

df = df.groupby('var1').transform('count')

This particular syntax adds a column called var1_count to the DataFrame that contains the count of values in the column called var1.

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

Example: Add Count Column in Pandas

Suppose we have the following pandas DataFrame that contains information about various basketball players:

import pandas as pd

#create DataFrame

df = pd.DataFrame({'team': ,
'pos': ,
'points': })

#view DataFrame
print(df)

team pos points
0 A Gu 18
1 A Fo 22
2 A Fo 19
3 B Fo 14
4 B Gu 14
5 B Gu 11
6 B Fo 20
7 B Fo 28

We can use the following code to add a column called team_count that contains the count of each team:

#add column that shows total count of each team
df = df.groupby('team').transform('count')

#view updated DataFrame
print(df)

team pos points team_count

0 A Gu 18 3

1 A Fo 22 3

2 A Fo 19 3

3 B Fo 14 5

4 B Gu 14 5

5 B Gu 11 5

6 B Fo 20 5

7 B Fo 28 5

There are 3 rows with a team value of A and 5 rows with a team value of B.

Thus:

For each row where the team is equal to A, the value in the team_count column is 3.For each row where the team is equal to B, the value in the team_count column is 5.

You can also add a 'count' column that groups by multiple variables.

For example, the following code shows how to add a 'count' column that groups by the team and pos

variables:

#add column that shows total count of each team and position
df = df.groupby(.transform('count')

#view updated DataFrame
print(df)

team pos points team_pos_count
0 A Gu 18 1
1 A Fo 22 2
2 A Fo 19 2
3 B Fo 14 3
4 B Gu 14 2
5 B Gu 11 2
6 B Fo 20 3
7 B Fo 28 3

From the output we can see:

There is 1 row that contains A in the team column and Gu in the pos column.There are 2 rows that contain A in the team column and Fo in the pos column.There are 3 rows that contain B in the team column and Fo in the

pos column.There are 2 rows that contain B in the team column and Gu in the pos column.

Additional Resources

The following tutorials explain how to perform other common tasks in pandas: