How can I use the groupby() function in Pandas to calculate the size of each group?

Authored by

stats writer

June 24, 2024

The groupby() function in Pandas allows for the grouping of data based on a specific column or set of columns. This function can be used to efficiently calculate the size of each group within a dataset. By using the groupby() function, the data can be organized into groups, and then the size of each group can be easily determined using the size() method. This allows for quick and accurate analysis of group sizes within a dataset, providing valuable insights into the underlying data. Overall, the groupby() function in Pandas is an essential tool for analyzing and understanding group sizes within a dataset.

Pandas: Use groupby() with size()

You can use the following methods with the groupby() and size() functions in pandas to count the number of occurrences by group:

Method 1: Count Occurrences Grouped by One Variable

df.groupby('var1').size()

Method 2: Count Occurrences Grouped by Multiple Variables

df.groupby().size()

Method 3: Count Occurrences Grouped by Multiple Variables and Sort by Count

df.groupby().size().sort_values(ascending=False)

The following examples show how to use each method in practice with the following pandas DataFrame:

```
import pandas as pd

#create DataFrame
df = pd.DataFrame({'team': ,
'position': ,
'points': })

#view DataFrame
print(df)
```

team position points
0 A G 15
1 A G 22
2 A F 24
3 A F 25
4 A F 20
5 B G 35
6 B G 34
7 B G 19
8 B G 14
9 B F 12

Example 1: Count Occurrences Grouped by One Variable

The following code shows how to use the groupby() and size() functions to count the occurrences of values in the team column:

#count occurrences of each value in team column
df.groupby('team').size()

team
A 5
B 5
dtype: int64

From the output we can see that the values A and B both occur 5 times in the team column.

Example 2: Count Occurrences Grouped by Multiple Variables

The following code shows how to use the groupby() and size() functions to count the occurrences of values for each combination of values in the team and position columns:

#count occurrences of values for each combination of team and position

df.groupby().size()

team position
A F 3
G 2
B F 1
G 4
dtype: int64

From the output we can see:

Team A and position F occurs 3 times.Team A and position G occurs 2 times.

And so on.

Example 3: Count Occurrences Grouped by Multiple Variables and Sort

The following code shows how to use the groupby() and size() functions to count the occurrences of values for each combination of values in the team and position columns, then sort by count:

#count occurrences for each combination of team and position and sort

```
df.groupby().size().sort_values(ascending=False)
```

```
team position
B G 4
A F 3
G 2
B F 1
dtype: int64
```

The output shows the count of each combination of team and position values, sorted by count in descending order.

Note: To sort by count in ascending order, simply remove ascending=False in the sort_values() function.

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