How can I use groupby with multiple aggregations in Pandas?

Authored by

stats writer

June 27, 2024

Groupby with multiple aggregations in Pandas refers to the process of organizing data into groups based on a specific column or set of columns and then applying multiple functions or operations to those groups simultaneously. This allows for a more efficient and streamlined analysis of data, as it eliminates the need for multiple separate groupby operations. By using this method, one can easily summarize and compare data within different groups, gaining valuable insights and making data analysis more efficient and accurate.

Pandas: Use Groupby with Multiple Aggregations

You can use the following basic syntax to use a groupby with multiple aggregations in pandas:

```
df.groupby('team').agg(
mean_points=('points', np.mean),
sum_points=('points', np.sum),
std_points=('points', np.std))
```

This particular formula groups the rows of the DataFrame by the variable called team and then calculates several summary statistics for the variable called points.

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

Example: Using Groupby with Multiple Aggregations 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': ,
'points': ,
'assists': })

#view DataFrame
print(df)

team points assists
0 Mavs 18 5
1 Mavs 22 7
2 Mavs 19 7
3 Heat 14 9
4 Heat 14 12
5 Heat 11 9
```

We can use the following syntax to group the rows of the DataFrame by team and then calculate the mean, sum, and standard deviation of points for each team:

```
import numpy as np
```

```
#group by team and calculate mean, sum, and standard deviation of points
df.groupby('team').agg(
mean_points=('points', np.mean),
sum_points=('points', np.sum),
std_points=('points', np.std))
```

```
mean_points sum_points std_points
team
Heat 13.000000 39 1.732051
Mavs 19.666667 59 2.081666
```

The output displays the mean, sum, and standard deviation of the points variable for each team.

You can use similar syntax to perform a groupby and calculate as many aggregations as you'd like.

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