How can I use Pandas to calculate the mean and standard deviation of one column within a groupby operation?

Authored by

stats writer

June 25, 2024

Pandas, a popular data analysis library in Python, offers the functionality to calculate the mean and standard deviation of a specific column within a groupby operation. This allows for efficient and accurate analysis of data grouped by certain categories. By utilizing the groupby function, the data can be grouped based on a specific column, and then the mean and standard deviation can be calculated for that particular column within each group. This provides valuable insights and statistical measures for understanding the data and identifying patterns or trends within the groups. Overall, Pandas provides a convenient and powerful tool for performing groupby operations and calculating descriptive statistics for efficient data analysis.

Pandas: Calculate Mean & Std of One Column in groupby

You can use the following syntax to calculate the mean and standard deviation of a column after using the groupby() operation in pandas:

df.groupby(, as_index=False).agg({'points':})

This particular example groups the rows of a pandas DataFrame by the value in the team column, then calculates the mean and standard deviation of values in the points column.

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

Example: Calculate Mean & Std of One Column in Pandas groupby

Suppose we have the following pandas DataFrame that

contains information about basketball players on various teams:

```
import pandas as pd

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

#view DataFrame
print(df)
```

```
team points assists
0 A 12 5
1 A 15 5
2 A 17 7
3 A 17 9
4 B 19 10
5 B 14 14
6 B 15 13
7 C 20 8
8 C 24 2
9 C 28 7
```

We can use the following syntax to calculate the mean and standard deviation of values in the points column, grouped by the team column:

#calculate mean and standard deviation of points, grouped by team
output = df.groupby(, as_index=False).agg({'points':})

#view resultsprint(output)

team points
mean std
0 A 15.25 2.362908
1 B 16.00 2.645751
2 C 24.00 4.000000

From the output we can see:

The mean points value for team A is 15.25.The standard deviation of points for team A is 2.362908.

And so on.

We can also rename the columns so that the output is easier to read:

#rename columns

output.columns =

#view updated resultsprint(output)

team points_mean points_std

0 A 15.25 2.362908

1 B 16.00 2.645751

2 C 24.00 4.000000

Note: You can find the complete documentation for the pandas groupby() operation .

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