How can I rename columns in a groupby function using Pandas?

Authored by

stats writer

June 27, 2024

The process of renaming columns in a groupby function using Pandas involves first grouping a dataset by a specific column or set of columns, and then using the "rename" function to assign new names to the grouped columns. This allows for more organized and clear labeling of data within the grouped dataset. Renaming columns in a groupby function can be useful for data analysis and visualization, as it provides a more descriptive and meaningful presentation of the data.

Pandas: Rename Columns in Groupby Function

You can use the following basic syntax to rename columns in a groupby() function in pandas:

df.groupby('group_col').agg(sum_col1=('col1', 'sum'), mean_col2=('col2', 'mean'), max_col3=('col3', 'max'))

This particular example calculates three aggregated columns and names them sum_col1, mean_col2, and max_col3.

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

Example: Rename Columns in Groupby Function in Pandas

Suppose we have the following pandas DataFrame:

import pandas as pd

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

#view DataFrame
print(df)

team points assists rebounds
0 A 30 5 4
1 A 22 6 13
2 A 19 6 15
3 A 14 5 10
4 B 14 8 7
5 B 11 7 7
6 B 20 7 5
7 B 28 9 11
```


We can use the following syntax to group the rows by the team column, then calculate three aggregated columns while providing specific names to the aggregated columns:

```
#calculate several aggregated columns by group and
```

rename aggregated columns

df.groupby('team').agg(sum_points=('points', 'sum'),
mean_assists=('assists', 'mean'),
max_rebounds=('rebounds', 'max'))

sum_points mean_assists max_rebounds
team
A 85 5.50 15
B 73 7.75 11

Notice that the three aggregated columns have the custom names that we provided in the agg() function.

Also note that we could use NumPy functions to calculate the sum, mean, and max values within the agg() function if we'd like.

import numpy as np

#calculate several aggregated columns by group and rename aggregated columns
df.groupby('team').agg(sum_points=('points', np.sum),
mean_assists=('assists', np.mean),
max_rebounds=('rebounds', np.max))

sum_points mean_assists max_rebounds
team
A 85 5.50 15
B 73 7.75 11

These results match the ones from the previous example.

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