How can I use Pandas GroupBy on a MultiIndex?

Authored by

stats writer

June 29, 2024

Pandas GroupBy on a MultiIndex is a powerful tool that allows for efficient and organized processing of data in a Pandas DataFrame with multiple index levels. This feature enables users to group and aggregate data based on specific criteria within each level of the index, providing a comprehensive view of the data. By utilizing Pandas GroupBy on a MultiIndex, users can easily analyze and manipulate complex datasets, making it a valuable tool for data analysis and management.

Pandas: Use GroupBy on a MultiIndex

You can use the following basic syntax to use GroupBy on a pandas DataFrame with a multiindex:

#calculate sum by level 0 and 1 of multiindex
df.groupby(level=).sum()

#calculate count by level 0 and 1 of multiindex
df.groupby(level=).count()

#calculate max value by level 0 and 1 of multiindex
df.groupby(level=).max()

...

Each of these examples calculate some metric grouped by two levels of a multiindex pandas DataFrame.

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

Example: Use GroupBy on MultiIndex in pandas

Suppose we have the following pandas DataFrame with a multiindex:

```
import pandas as pd

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

#define multiindex
df.set_index(, inplace=True)

#view DataFrame
print(df)
```

```
points
team position
A G 6
G 8
F 9
F 11
B G 13
G 8
```

F 8
F 15

We can use the following syntax to calculate the sum of the points values grouped by both levels of the multiindex:

#calculate sum of points grouped by both levels of the multiindex:
df.groupby(level=).sum()

points
team position
A F 20
G 14
B F 23
G 21

We can use similar syntax to calculate the max of the points values grouped by both levels of the multiindex:

#calculate max of points grouped by both levels of the multiindex:
df.groupby(level=).max()

points

team position

A F 11

G 8

B F 15

G 13

We can use similar syntax to calculate any value we'd like grouped by several levels of a multiindex.

Note: You can find the complete documentation for the GroupBy operation in pandas .

Additional Resources

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