

How can I group a Pandas DataFrame by its index and perform a calculation on the grouped data?

Authored by
stats writer

July 2, 2024

RECOMMENDED CITATION

stats writer (2024). *How can I group a Pandas DataFrame by its index and perform a calculation on the grouped data?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=165794>

To group a Pandas DataFrame by its index and perform a calculation on the grouped data, first use the "groupby" function to group the data by the desired index. Then, specify the desired calculation to be performed on the grouped data using built-in Pandas functions such as "sum", "mean", or "count". This will return a new DataFrame with the calculated values for each group based on the specified index. This method is useful for analyzing and summarizing data based on specific categories within the index.

Pandas: Group By Index and Perform Calculation

You can use the following methods to group by one or more index columns in pandas and perform some calculation:

Method 1: Group By One Index Column

```
df.groupby('index1').max()
```

Method 2: Group By Multiple Index Columns

```
df.groupby().sum()
```

Method 3: Group By Index Column and Regular Column

```
df.groupby().nunique()
```

The following examples show how to use each method with the following pandas DataFrame that has a

Multindex:

```
import pandas as pd
```

```
#create DataFrame
```

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

```
#set 'team' column to be index column
```

```
df.set_index(, inplace=True)
```

```
#view DataFrame
```

```
df
```

```
points rebounds
```

```
team position
```

```
A G 7 8
```

```
G 7 8
```

```
G 7 8
```

```
F 19 10
```

```
F 16 11
```

```
B G 9 12
```

```
G 10 13
```

F 10 13

F 8 15

F 8 11

Method 1: Group By One Index Column

The following code shows how to find the max value of the 'points' column, grouped by the 'position' index column:

```
#find max value of 'points' grouped by 'position' index column
```

```
df.groupby('position').max()
```

```
position
```

```
F 19
```

```
G 10
```

```
Name: points, dtype: int64
```

Method 2: Group By Multiple Index Columns

The following code shows how to find the sum of the 'points' column, grouped by the 'team' and 'position' index columns:

```
#find max value of 'points' grouped by 'position' index
```

column

df.groupby().sum()

team position

A F 35

G 21

B F 26

G 19

Name: points, dtype: int64

Method 3: Group By Index Column & Regular Column

The following code shows how to find the number of unique values in the 'rebounds' column, grouped by the index column 'team' and the ordinary column 'points':

#find max value of 'points' grouped by 'position index column

df.groupby().nunique()

team points

A 7 1

16 1

19 1

B 8 2

9 1

10 1

Name: rebounds, dtype: int64

Additional Resources

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

ARABPSYCHOLOGY.COM