How can I calculate quantiles by group in Pandas?

Authored by

stats writer

May 5, 2024

Pandas is a popular Python library for data analysis and manipulation. One of its useful functionalities is the ability to calculate quantiles by group. This means that you can find the quantiles (i.e. values that divide a dataset into equal portions) for each unique group in a given dataset. This can be achieved by using the "groupby" function in Pandas, which allows you to group your data by a specific column or set of columns. Then, you can use the "quantile" function to calculate the desired quantiles for each group. This feature is particularly useful for exploring the distribution of data within different subgroups and can provide valuable insights in data analysis tasks.

Calculate Quantiles by Group in Pandas

You can use the following basic syntax to calculate quantiles by group in Pandas:

df.groupby('grouping_variable').quantile(.5)

The following examples show how to use this syntax in practice.

Example 1: Calculate Quantile by Group

Suppose we have the following pandas DataFrame:

import pandas as pd

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

#view first five rows
df.head()

team score
0 1 3
1 1 4
2 1 4
3 1 5
4 1 5

The following code shows how to calculate the 90th percentile of values in the 'points' column, grouped by the 'team' column:

df.groupby('team').quantile(.90)

score
team
1 6.5
2 4.0

Here's how to interpret the output:

The 90th percentile of 'points' for team 1 is 6.5.The 90th percentile of 'points' for team 2 is 4.0.

Example 2: Calculate Several Quantiles by Group

The following code shows how to calculate several quantiles at once by group:

```
import pandas as pd

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

#create functions to calculate 1st and 3rd quartiles
def q1(x):
return x.quantile(0.25)

def q3(x):
return x.quantile(0.75)

#calculate 1st and 3rd quartiles by group
vals = {'score': }

df.groupby('team').agg(vals)

score
q1 q3
team
```

1 4.0 5.0

2 2.0 3.0

Here's how to interpret the output:

The first and third quartile of scores for team 1 is 4.0 and 5.0, respectively.The first and third quartile of scores for team 2 is 2.0 and 3.0, respectively.

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