How can I use the aggfunc parameter in the crosstab() function in Pandas?

Authored by

stats writer

June 24, 2024

The aggfunc parameter in the crosstab() function in Pandas allows for the aggregation of data when creating a cross-tabulation table. This parameter allows the user to specify a function that will be applied to the values in each cell of the table, such as sum, mean, or count. This provides flexibility in analyzing and summarizing data in a cross-tabulation format. By using the aggfunc parameter, users can easily customize the output of their crosstab table to suit their specific needs.

Pandas: Use aggfunc in crosstab() Function

You can use the aggfunc argument within the pandas crosstab() function to create a crosstab that aggregates values using a specific metric:

pd.crosstab(index=df.col1, columns=df.col2, values=df.col3, aggfunc='count')

The default value for aggfunc is 'count' but you can specify other aggregation methods such as mean, median, sum, min, max, etc.

You can also specify multiple aggregation methods in the aggfunc argument:

pd.crosstab(index=df.col1, columns=df.col2, values=df.col3, aggfunc=)

The following examples show how to use each of these methods in practice with the following pandas

DataFrame:

import pandas as pd

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

#view DataFrame
print(df)

team position points
0 A G 22
1 A G 25
2 A F 24
3 B G 39
4 B F 34
5 B F 20
6 B F 18
7 C G 17
8 C G 20
9 C F 19
10 C F 22

Example 1: Create Crosstab with One Value in aggfunc

We can use the following crosstab() function with the argument aggfunc='mean' to create a crosstab that displays the mean value points for each combination of position and team:

#create crosstab that displays mean points value by team and position
pd.crosstab(index=df.team, columns=df.position, values=df.points, aggfunc='mean')

position F G
team
A 24.0 23.5
B 24.0 39.0
C 20.5 18.5

Here is how to interpret the output:

The average points for players on team A in position F is 24.The average points for players on team A in position G is 23.5.

And so on.

We can also use a different aggregation metric, such as the maximum value:

#create crosstab that displays max points value by team and position
pd.crosstab(index=df.team, columns=df.position, values=df.points, aggfunc='max')

position F G
team
A 24 25
B 34 39
C 22 20

Here is how to interpret the output:

The max points for players on team A in position F is 24.The max points for players on team A in position G is 25.

Example 2: Create Crosstab with Multiple Values in aggfunc

We can use the crosstab() function with multiple values in the aggfunc argument to aggregate the points values by multiple metrics for each combination of position and team:

#create crosstab that displays min and max points by team and position
pd.crosstab(df.team, df.position, df.points, aggfunc=)

min max
position F G F G
team
A 24 22 24 25
B 18 39 34 39
C 19 17 22 20

Here is how to interpret the output:

The minimum points value for players on team A in position F is 24.The minimum points value for players on team A in position G is 22.The maximum points value for players on team A in position F is 24.The maximum points value for players on team A in position G is 25.

And so on.

Note: You can find the complete documentation for the pandas crosstab() function .

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