How can I create a crosstab in Pandas with percentages?

Authored by

stats writer

June 24, 2024

Creating a crosstab in Pandas with percentages allows for the visualization of the relationship between two variables, while also displaying the percentage distribution of the data. This can be achieved by using the built-in crosstab function in Pandas, which calculates the percentage of values in each column and row based on the total values in the table. This feature is particularly useful in data analysis and can provide valuable insights into the data.

Pandas: Create Crosstab with Percentages

You can use the normalize argument within the pandas crosstab() function to create a crosstab that displays percentage values instead of counts:

pd.crosstab(df.col1, df.col2, normalize='index')

The normalize argument accepts three different arguments:

all: Display percentage relative to all values.index: Display percentage as total of row values.columns: Display percentage as total of column values.

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

Here is what the default crosstab would look like for the count of players by team and position:

#create crosstab that displays count by team and positionpd.crosstab(df.team, df.position)

position F G
team
A 1 2
B 3 1
C 2 2

Example 1: Create Crosstab with Percentages Relative to All Values

We can use the crosstab() function with the argument normalize=all to create a crosstab that displays percentages of each value relative to the total count of all values:

#create crosstab that displays counts as percentage relative to total count
pd.crosstab(df.team, df.position, normalize='all')

position F G
team
A 0.090909 0.181818
B 0.272727 0.090909
C 0.181818 0.181818

Here is how to interpret the output:

Players on team A in position F account for 9.09% of total players.Players on team A in position G account for 18.18% of total players.

And so on.

Example 2: Create Crosstab with Percentages Relative to Row Totals

We can use the crosstab() function with the argument normalize=index to create a crosstab that displays percentages of each value relative to the row total:

```
#create crosstab that displays counts as percentage relative to row totals
pd.crosstab(df.team, df.position, normalize='index')
```

```
position F G
team
A 0.333333 0.666667
B 0.750000 0.250000
C 0.500000 0.500000
```

Players in position F account for 33.33% of total players on team A.Players in position F account for 75% of total players on team B.Players in position F account for 50%

of total players on team C.

And so on.

Example 3: Create Crosstab with Percentages Relative to Column Totals

We can use the crosstab() function with the argument normalize=columns to create a crosstab that displays percentages of each value relative to the column total:

```
#create crosstab that displays counts as percentage relative to column totals
pd.crosstab(df.team, df.position, normalize='columns')
```

```
position F G
team
A 0.166667 0.4
B 0.500000 0.2
C 0.333333 0.4
```

Here is how to interpret the output:

Players on team A account for 16.67% of total players with a position of F.Players on team B account for 50% of total players with a position of F.Players on team C

account for 33.33% of total players with a position of F.

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: