“How can I calculate the conditional mean using Pandas, and could you provide some examples?”

Authored by

stats writer

June 28, 2024

Pandas is a powerful data analysis library that allows users to easily calculate the conditional mean for their datasets. The conditional mean is a statistical measure that calculates the average value of a variable based on a specific condition or criteria. With Pandas, this can be achieved by using the groupby function to group the data by a certain category or feature, and then applying the mean function to the desired variable. This results in a new dataset with the conditional mean values for each group. For example, if we have a dataset of student grades and want to calculate the conditional mean for each grade level, we can use Pandas to group the data by grade level and then calculate the mean for the grades column. This allows for efficient and accurate analysis of data based on specific conditions.

Calculate Conditional Mean in Pandas (With Examples)

You can use the following syntax to calculate a conditional mean in pandas:

df.loc == 'A', 'points'].mean()

This calculates the mean of the 'points' column for every row in the DataFrame where the 'team' column is equal to 'A.'

The following examples show how to use this syntax in practice with the following pandas DataFrame:

import pandas as pd

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

'assists': })

#view DataFrame
print(df)

team points assists
0 A 99 33
1 A 90 28
2 A 93 31
3 B 86 39
4 B 88 34
5 B 82 30

Example 1: Calculate Conditional Mean for Categorical Variable

The following code shows how to calculate the mean of the 'points' column for only the rows in the DataFrame where the 'team' column has a value of 'A.'

#calculate mean of 'points' column for rows where team equals 'A'
df.loc == 'A', 'points'].mean()

94.0

The mean value in the 'points' column for the rows

where 'team' is equal to 'A' is 94.

We can manually verify this by calculating the average of the points values for only the rows where 'team' is equal to 'A':

Average of Points: (99 + 90 + 93) / 3 = 94

Example 2: Calculate Conditional Mean for Numeric Variable

The following code shows how to calculate the mean of the 'assists' column for only the rows in the DataFrame where the 'points' column has a value greater than or equal to 90.

```
#calculate mean of 'assists' column for rows where 'points' >= 90
df.loc >= 90, 'assists'].mean()
30.666666666666668
```

The mean value in the 'assists' column for the rows where 'points' is greater than or equal to 90 is 30.66667.

We can manually verify this by calculating the average of the points values for only the rows where 'team' is equal to 'A':

Average of Assists: (33 + 28 + 31) / 3 = 30.66667

Additional Resources

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