How can I calculate the average of selected columns in Pandas?

Authored by

stats writer

July 1, 2024

Calculating the average of selected columns in Pandas is a process that involves using the built-in functions and methods of the Pandas library. With Pandas, one can easily select specific columns from a dataset and then use the mean function to calculate the average of those columns. This approach allows for a quick and efficient way to obtain the average of multiple columns in a dataset. Additionally, Pandas also offers various options for customizing the calculation, such as excluding null values and selecting specific rows. By utilizing the features of Pandas, one can easily and accurately calculate the average of selected columns in a dataset.

Calculate the Average of Selected Columns in Pandas

You can use the following methods to calculate the average row values for selected columns in a pandas DataFrame:

Method 1: Calculate Average Row Value for All Columns

df.mean(axis=1)

Method 2: Calculate Average Row Value for Specific Columns

df].mean(axis=1)

The following examples shows how to use each method in practice with the following pandas DataFrame:

import pandas as pd

```
#create DataFrame
df = pd.DataFrame({'points': ,
'assists': ,
'rebounds': })

#view DataFrame
df
```

```
points assists rebounds
0 14 5 11
1 19 7 8
2 9 7 10
3 21 9 6
4 25 12 6
5 29 9 5
6 20 9 9
7 11 4 12
```


Method 1: Calculate Average Row Value for All Columns

The following code shows how to create a new column in the DataFrame that displays the average row value for all columns:

```
#define new column that shows the average row value
```

```
for all columns
df = df.mean(axis=1)

#view updated DataFrame
df

points assists rebounds average_all
0 14 5 11 10.000000
1 19 7 8 11.333333
2 9 7 10 8.666667
3 21 9 6 12.000000
4 25 12 6 14.333333
5 29 9 5 14.333333
6 20 9 9 12.666667
7 11 4 12 9.000000
```

Here's how to interpret the output:

The average value of the first row is calculated as: (14+5+11) / 3 = 10.

The average value of the second row is calculated as: (19+7+8) / 3 = 11.33.

And so on.

Method 2: Calculate Average Row Value for Specific Columns

The following code shows how to calculate the average row value for just the "points" and "rebounds" columns:

```
#define new column that shows average of row values for points and rebounds columns
df = df].mean(axis=1)

#view updated DataFrame
df
```

```
points assists rebounds avg_points_rebounds
0 14 5 11 12.5
1 19 7 8 13.5
2 9 7 10 9.5
3 21 9 6 13.5
4 25 12 6 15.5
5 29 9 5 17.0
6 20 9 9 14.5
7 11 4 12 11.5
```

The average value of "points" and "rebounds" in the first row is calculated as: (14+11) / 2 = 12.5.

The average value of "points" and "rebounds" in the second row is calculated as: (19+8) / 2 = 13.5.

And so on.

Additional Resources

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