How can I calculate percentile rank using Pandas, and could you provide some examples?”

Authored by

stats writer

June 27, 2024

The process of calculating percentile rank using Pandas involves using the built-in functions and methods provided by the Pandas library. This allows for efficient and accurate calculation of the percentile rank of a given set of data. To do so, one must first sort the data in ascending order, then use the "rank" function to assign a rank to each data point. The percentile rank can then be determined by dividing the assigned rank by the total number of data points and multiplying by 100. Examples of using this method can be seen in various statistical analysis tasks, such as determining the performance ranking of students or measuring the distribution of income in a population.

Calculate Percentile Rank in Pandas (With Examples)

The percentile rank of a value tells us the percentage of values in a dataset that rank equal to or below a given value.

You can use the following methods to calculate percentile rank in pandas:

Method 1: Calculate Percentile Rank for Column

df = df.rank(pct=True)

Method 2: Calculate Percentile Rank by Group

df = df.groupby('group_var').transform('rank', pct=True)

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

```
import pandas as pd

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

#view DataFrame
print(df)
```

team points
0 A 2
1 A 5
2 A 5
3 A 7
4 A 9
5 A 13
6 A 15
7 B 17
8 B 22
9 B 24
10 B 30
11 B 31
12 B 38
13 B 39

Example 1: Calculate Percentile Rank for Column

The following code shows how to calculate the percentile rank of each value in the points column:

```
#add new column that shows percentile rank of points
df = df.rank(pct=True)

#view updated DataFrame
print(df)
```

```
team points percent_rank
0 A 2 0.071429
1 A 5 0.178571
2 A 5 0.178571
3 A 7 0.285714
4 A 9 0.357143
5 A 13 0.428571
6 A 15 0.500000
7 B 17 0.571429
8 B 22 0.642857
9 B 24 0.714286
10 B 30 0.785714
11 B 31 0.857143
12 B 38 0.928571
```

13 B 39 1.000000

Here's how to interpret the values in the percent_rank column:

7.14% of the points values are equal to or less than 2.17.86% of the points values are equal to or less than 5.28.57% of the points values are equal to or less than 7.

And so on.

Example 2: Calculate Percentile Rank by Group

The following code shows how to calculate the percentile rank of each value in the points column, grouped by team:

```
#add new column that shows percentile rank of points, grouped by team
df = df.groupby('team').transform('rank', pct=True)

#view updated DataFrame
print(df)
```

team points percent_rank

0 A 2 0.142857

1 A 5 0.357143

2 A 5 0.357143

3 A 7 0.571429

4 A 9 0.714286

5 A 13 0.857143

6 A 15 1.000000

7 B 17 0.142857

8 B 22 0.285714

9 B 24 0.428571

10 B 30 0.571429

11 B 31 0.714286

12 B 38 0.857143

13 B 39 1.000000

14.3% of the points values for team A are equal to or less than 2.35.7% of the points values for team A are equal to or less than 5.57.1% of the points values for team A are equal to or less than 7.

And so on.

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