

How can quartiles be calculated using Pandas?

Authored by
stats writer

June 25, 2024

RECOMMENDED CITATION

stats writer (2024). *How can quartiles be calculated using Pandas?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=151688>

Quartiles can be calculated using Pandas by utilizing the built-in function "describe()" which generates descriptive statistics for a given dataset. This function provides the minimum, maximum, mean, standard deviation, and quartile values (25%, 50%, and 75%) for each column in the dataset. Additionally, the "quantile()" function can be used to specifically calculate the quartile values for a particular column by specifying the desired percentile. These Pandas functions make it easy to quickly obtain quartile information for data analysis and visualization purposes.

Calculate Quartiles in Pandas (With Example)

In statistics, quartiles are values that split up a dataset into four equal parts.

When analyzing a distribution, we're typically interested in the following quartiles:

First Quartile (Q1): The value located at the 25th percentile
Second Quartile (Q2): The value located at the 50th percentile
Third Quartile (Q3): The value located at the 75th percentile

You can use the following methods to calculate the quartiles for columns in a pandas DataFrame:

Method 1: Calculate Quartiles for One Column

df.quantile()

Method 2: Calculate Quartiles for Each Numeric Column

df.quantile(q=, axis=0, numeric_only=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': ,  
'assists': })
```

```
#view DataFrame
```

```
print(df)
```

```
team points assists
```

```
0 A 12 2
```

```
1 B 14 2
```

```
2 C 14 3
```

```
3 D 16 3
```

```
4 E 24 4
```

```
5 F 26 6
```

```
6 G 28 7
```

```
7 H 30 8
```

```
8 I 31 10
```

9 J 35 15

Example 1: Calculate Quartiles for One Column

The following code shows how to calculate the quartiles for the points column only:

```
#calculate quartiles for points column  
df.quantile()
```

```
0.25 14.5
```

```
0.50 25.0
```

```
0.75 29.5
```

```
Name: points, dtype: float64
```

From the output we can see:

The first quartile is located at 14.5. The second quartile is located at 25. The third quartile is located at 29.5.

By only knowing these three values, we have a pretty good idea of how the values are distributed in the points column.

Example 2: Calculate Quartiles for Each Numeric Column

#calculate quartiles for each numeric column in DataFrame

```
df.quantile(q=[0.25, 0.50, 0.75], axis=0, numeric_only=True)
```

points assists

0.25 14.5 3.00

0.50 25.0 5.00

0.75 29.5 7.75

The output displays the quartiles for the two numeric columns in the DataFrame.

Note that there is more than one way to calculate quartiles for a distribution.

Refer to the pandas documentation to see the various methods that the pandas quantile() function uses to calculate quartiles.

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