

How can the rolling median be calculated in Pandas?

Authored by
stats writer

July 2, 2024

RECOMMENDED CITATION

stats writer (2024). *How can the rolling median be calculated in Pandas?*.

PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=165906>

The rolling median in Pandas is a statistical method used to calculate the median value of a dataset over a rolling window. This technique involves taking a fixed number of data points from the dataset and calculating the median value within that window. The window is then shifted by one data point and the process is repeated. This allows for a more dynamic and accurate representation of the overall median value as it takes into account changes in the data over time. The rolling median can be calculated in Pandas using the built-in function "rolling()" and specifying the size of the window. This method is commonly used for time series data analysis and can provide valuable insights into trends and patterns within the dataset.

Calculate Rolling Median in Pandas (With Examples)

A rolling median is the median of a certain number of previous periods in a time series.

To calculate the rolling median for a column in a pandas DataFrame, we can use the following syntax:

```
#calculate rolling median of previous 3 periods
```

```
df.rolling(3).median()
```

The following example shows how to use this function in practice.

Example: Calculate Rolling Median of Column

Suppose we have the following pandas DataFrame:

```
import pandas as pd
```

#create DataFrame

```
df = pd.DataFrame({'month': ,  
'leads': ,  
'sales': })
```

#view DataFrame

```
df
```

month leads sales

```
0 1 13 22
```

```
1 2 15 24
```

```
2 3 16 23
```

```
3 4 15 27
```

```
4 5 17 26
```

```
5 6 20 26
```

```
6 7 22 27
```

```
7 8 24 30
```

```
8 9 25 33
```

```
9 10 26 32
```

```
10 11 23 27
```

```
11 12 24 25
```

We can use the following syntax to create a new column that contains the rolling median of 'sales' for the

previous 3 periods:

#calculate 3-month rolling median

```
df = df.rolling(3).median()
```

#view updated data frame

```
df
```

```
month leads sales sales_rolling3
```

```
0 1 13 22 NaN
```

```
1 2 15 24 NaN
```

```
2 3 16 23 23.0
```

```
3 4 15 27 24.0
```

```
4 5 17 26 26.0
```

```
5 6 20 26 26.0
```

```
6 7 22 27 26.0
```

```
7 8 24 30 27.0
```

```
8 9 25 33 30.0
```

```
9 10 26 32 32.0
```

```
10 11 23 27 32.0
```

```
11 12 24 25 27.0
```

We can manually verify that the rolling median sales displayed for month 3 is the median of the previous 3 months:

Median of 22, 24, 23 = 23.0

Similarly, we can verify the rolling median sales of month 4:

Median of 24, 23, 27 = 24.0

We can use similar syntax to calculate the rolling 6-month median:

```
#calculate 6-month rolling median
df = df.rolling(6).median()#view updated data frame
df
```

```
month leads sales sales_rolling3 sales_rolling6
```

```
0 1 13 22 NaN NaN
1 2 15 24 NaN NaN
2 3 16 23 23.0 NaN
3 4 15 27 24.0 NaN
4 5 17 26 26.0 NaN
5 6 20 26 26.0 25.0
6 7 22 27 26.0 26.0
7 8 24 30 27.0 26.5
8 9 25 33 30.0 27.0
9 10 26 32 32.0 28.5
```

10 11 23 27 32.0 28.5

11 12 24 25 27.0 28.5

Additional Resources

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

ARABPSYCHOLOGY.COM