How can I calculate the cumulative count in Pandas?

Authored by

stats writer

June 26, 2024

Calculating the cumulative count in Pandas refers to the process of determining the total number of occurrences of a specific value in a given dataset, up to a certain point. This can be achieved by using the "cumcount()" function in Pandas, which returns a cumulative count for each row in a specified column. This function is useful for analyzing and visualizing the frequency of certain values in a dataset, and can be used in various data analysis tasks. By understanding how to use the "cumcount()" function, one can efficiently calculate and interpret cumulative counts in Pandas.

Calculate Cumulative Count in Pandas

You can use the following methods to calculate a cumulative count in a pandas DataFrame:

Method 1: Cumulative Count by Group

df = df.groupby('col1').cumcount()

Method 2: Cumulative Count by Multiple Groups

df = df.groupby().cumcount()

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({'team': ,
'position': ,

'points': })

#view DataFrame
print(df)

team position points
0 A G 14
1 A G 22
2 A G 25
3 A F 34
4 B G 30
5 B G 12
6 B F 10
7 B F 18

Example 1: Cumulative Count by Group in Pandas

We can use the following syntax to create a new column called team_cum_count that displays the cumulative count for each team in the DataFrame:

#calculate cumulative count by team
df = df.groupby('team').cumcount()

#view updated DataFrame

```
print(df)

team position points team_cum_count
0 A G 14 0
1 A G 22 1
2 A G 25 2
3 A F 34 3
4 B G 30 0
5 B G 12 1
6 B F 10 2
7 B F 18 3
```

The new column called team_cum_count contains the cumulative count of each team, starting with a value of zero.

If you'd like the count to start at one instead, simply add one to the end of the line:

```
#calculate cumulative count (starting at 1) by team
df = df.groupby('team').cumcount() + 1

#view updated DataFrame
print(df)
```

team position points team_cum_count
0 A G 14 1
1 A G 22 2
2 A G 25 3
3 A F 34 4
4 B G 30 1
5 B G 12 2
6 B F 10 3
7 B F 18 4

The new column called team_cum_count contains the cumulative count of each team, starting with a value of one.

Example 2: Calculate Cumulative Count by Group in Pandas

We can use the following syntax to create a new column called team_pos_cum_count that displays the cumulative count for each team and position in the DataFrame:

```
#calculate cumulative count by team
df = df.groupby().cumcount()

#view updated DataFrame
```

print(df)

team position points team_pos_cum_count
0 A G 14 0
1 A G 22 1
2 A G 25 2
3 A F 34 0
4 B G 30 0
5 B G 12 1
6 B F 10 0
7 B F 18 1

Note: You can find the complete documentation for the cumcount function in pandas .

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