

How can I use the groupby function in Pandas to count the number of occurrences in a DataFrame, while also applying a condition to the data?

Authored by
stats writer

June 28, 2024

RECOMMENDED CITATION

stats writer (2024). *How can I use the groupby function in Pandas to count the number of occurrences in a DataFrame, while also applying a condition to the data?*.

PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=155925>

The groupby function in Pandas allows for the grouping of data in a DataFrame based on a specific column or multiple columns. By using this function, it is possible to count the number of occurrences in the grouped data while also applying a condition to the data. This means that the data can be filtered and the count can be calculated based on the specified condition. This feature is particularly useful for data analysis and manipulation, as it allows for a more targeted and efficient way of counting occurrences in a DataFrame.

Pandas: Use Groupby and Count with Condition

You can use the following basic syntax to perform a groupby and count with condition in a pandas DataFrame:

```
df.groupby('var1').apply(lambda x: (x=='val').sum()).reset_index(name='count')
```

This particular syntax groups the rows of the DataFrame based on var1 and then counts the number of rows where var2 is equal to 'val.'

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

Example: Groupby and Count with Condition in Pandas

Suppose we have the following pandas DataFrame that contains information about various basketball players:

```
import pandas as pd

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

```
#view DataFrame
print(df)
```

```
team pos points
0 A Gu 18
1 A Fo 22
2 A Fo 19
3 A Fo 14
4 B Gu 14
5 B Gu 11
6 B Fo 20
7 B Fo 28
```

The following code shows how to group the `DataFrame` by the `team` variable and count the number of rows where the `pos` variable is equal to 'Gu':

```
#groupby team and count number of 'pos' equal to 'Gu'
```

```
df_count = df.groupby('team').apply(lambda x:
(x=='Gu').sum()).reset_index(name='count')
```

```
#view results
```

```
print(df_count)
```

```
team count
```

```
0 A 1
```

```
1 B 2
```

From the output we can see:

Team A has 1 row where the pos column is equal to 'Gu'
Team B has 2 rows where the pos column is equal to 'Gu'

We can use similar syntax to perform a groupby and count with some numerical condition.

For example, the following code shows how to group by the team variable and count the number of rows where the points variable is greater than 15:

```
#groupby team and count number of 'points' greater than 15
```

```
df_count = df.groupby('team').apply(lambda x:
```

```
(x>15).sum()).reset_index(name='count')
```

```
#view results
```

```
print(df_count)
```

```
team count
```

```
0 A 3
```

```
1 B 2
```

From the output we can see:

**Team A has 3 rows where the points column is greater than 15
Team B has 2 rows where the points column is greater than 15**

You can use similar syntax to perform a groupby and count with any specific condition you'd like.

Additional Resources

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