

How can I calculate the percentage of the total within a group using Pandas?

Authored by
stats writer

June 29, 2024

RECOMMENDED CITATION

stats writer (2024). *How can I calculate the percentage of the total within a group using Pandas?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=159228>

Using the Pandas library, you can easily calculate the percentage of the total within a group by using the built-in functions and methods. First, you will need to group the data by a specific column using the "groupby" function. Then, you can use the "sum" function to calculate the total within each group. Finally, you can use the "apply" function to calculate the percentage by dividing the individual values by the total and multiplying by 100. This method allows you to efficiently calculate the percentage of the total within a group in a concise and accurate manner.

Pandas: Calculate Percentage of Total Within Group

You can use the following syntax to calculate the percentage of a total within groups in pandas:

```
df / df.groupby('group_var').transform('sum')
```

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

Example: Calculate Percentage of Total Within Group

Suppose we have the following pandas DataFrame that shows the points scored by basketball players on various teams:

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

```
#view DataFrame
```

```
print(df)
```

```
team points
```

```
0 A 12
```

```
1 A 29
```

```
2 A 34
```

```
3 A 14
```

```
4 A 10
```

```
5 B 11
```

```
6 B 7
```

```
7 B 36
```

```
8 B 34
```

```
9 B 22
```

We can use the following syntax to create a new column in the DataFrame that shows the percentage of total points scored, grouped by team:

```
#calculate percentage of total points scored grouped by team
```

```
df = df / df.groupby('team').transform('sum')
```

```
#view updated DataFrame
```

```
print(df)
```

```
team points team_percent
```

```
0 A 12 0.121212
```

```
1 A 29 0.292929
```

```
2 A 34 0.343434
```

```
3 A 14 0.141414
```

```
4 A 10 0.101010
```

```
5 B 11 0.100000
```

```
6 B 7 0.063636
```

```
7 B 36 0.327273
```

```
8 B 34 0.309091
```

```
9 B 22 0.200000
```

The `team_percent` column shows the percentage of total points scored by that player within their team.

For example, players on team A scored a total of 99 points.

Thus, the player in the first row of the DataFrame who scored 12 points scored a total of $12/99 = 12.12\%$ of the total points for team A.

Similarly, the player in the second row of the DataFrame

who scored 29 points scored a total of $29/99 = 29.29\%$ of the total points for team A.

And so on.

Note: You can find the complete documentation for the GroupBy function .

Additional Resources

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

ARABPSYCHOLOGY.COM