

How can I calculate a weighted average in pandas?

Authored by
stats writer

April 29, 2024

RECOMMENDED CITATION

stats writer (2024). *How can I calculate a weighted average in pandas?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=141103>

Calculating a weighted average in pandas refers to the process of finding the average of a set of values by taking into consideration the weights assigned to each value. This can be achieved by using the weighted average formula, where the sum of the product of each value and its corresponding weight is divided by the sum of the weights. In pandas, this can be done by creating a new column with the product of the values and weights, and then using the sum function to find the total sum of these products. The final step is to divide this sum by the total sum of the weights, giving us the weighted average. This method is particularly useful in situations where certain values hold more significance or importance than others in the calculation of an average.

Calculate a Weighted Average in Pandas

You can use the following function to calculate a weighted average in Pandas:

```
def w_avg(df, values, weights):  
d = df  
w = df  
return (d * w).sum() / w.sum()
```

The following examples show how to use this syntax in practice.

Example 1: Weighted Average in Pandas

The following code shows how to use the weighted average function to calculate a weighted average for a given dataset, using "price" as the values and "amount" as the weights:

```
import pandas as pd

#create DataFrame
df = pd.DataFrame({'sales_rep': ,
'price': ,
'amount': })

#view DataFrame
df

sales_rep price amount
0 A 8 1
1 A 5 3
2 A 6 2
3 B 7 2
4 B 12 5
5 B 14 4

#find weighted average of price
w_avg(df, 'price', 'amount')

9.705882352941176
```

The weighted average of "price" turns out to be 9.706.

Example 2: Groupby and Weighted Average in Pandas

The following code shows how to use the weighted average function to calculate the weighted average of price, *grouped by* sales rep:

```
import pandas as pd

#create DataFrame
df = pd.DataFrame({'sales_rep': ,
'price':,
'amount': })

#find weighted average of price, grouped by sales rep
df.groupby('sales_rep').apply(w_avg, 'price', 'amount')

sales_rep
A 5.833333
B 11.818182
dtype: float64
```

We can see the following:

The weighted average of "price" for sales rep A is 5.833. The weighted average of "price" for sales rep B is 11.818.