How can I find the difference between two columns in Pandas?

Authored by

stats writer

April 28, 2024

RECOMMENDED CITATION

stats writer (2024). *How can I find the difference between two columns in Pandas?*. PSYCHOLOGICAL SCALES. Retrieved from https://scales.arabpsychology.com/?p=140444

ARABPSYCHOLOGY.COM

Pandas is a popular library in Python used for data analysis and manipulation. It offers various functions and methods to efficiently handle data, including the ability to find the difference between two columns. This can be achieved by using the "subtract" function, which takes the two columns as inputs and returns a new column with the difference values. Alternatively, the "diff" function can also be used to find the difference between two columns, which provides more options for handling missing values and the direction of subtraction. Overall, by utilizing these functions in Pandas, one can easily find the difference between two columns and analyze the data effectively.

Pandas: Find the Difference Between Two Columns

To find the difference between any two columns in a pandas DataFrame, you can use the following syntax:

df = df - df

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

Example 1: Find Difference Between Two Columns

Suppose we have the following pandas DataFrame that shows the total sales for two regions (A and B) during eight consecutive sales periods:

import pandas as pd

#create DataFrame
df = pd.DataFrame({'period': ,
'A_sales': ,

ARABPSYCHOLOGY.COM

'B_sales': })

#view DataFrame
df

period A_sales B_sales
0 1 12 14
1 2 14 19
2 3 15 20
3 4 13 22
4 5 18 24
5 6 20 20
6 7 19 17
7 8 24 23

The following code shows how calculate the difference between the sales in region B and region A for each sales period:

#add new column to represent difference between B sales and A sales
df = df - df

#view DataFrame
df

ARABPSYCHOLOGY.COM

period A_sales B_sales diff

0 1 12 14 2

1 2 14 19 5

2 3 15 20 5

3 4 13 22 9

4 5 18 24 6

5 6 20 20 0

6 7 19 17 -2

7 8 24 23 -1

We could also calculate the absolute difference in sales by using the pandas.Series.abs() function:

```
#add new column to represent absolute difference between B sales and A sales
df = pd.Series.abs(df - df)

#view DataFrame
df
```

period A_sales B_sales diff

0 1 12 14 2

1 2 14 19 5

2 3 15 20 5

3 4 13 22 9

ARABPSYCHOLOGY.COM

4 5 18 24 6

5 6 20 20 0

6 7 19 17 2

7 8 24 23 1

Example 2: Find Difference Between Columns Based on Condition

We can also filter the DataFrame to only show rows where the difference between the columns is less than or greater than some value.

For example, the following code returns only the rows where the the sales in region A is greater than the sales in region B:

```
#add new column to represent difference between B sales and A sales
df = df - df

#display rows where sales in region A is greater than sales in region B
df<0]
```

period A_sales B_sales diff

6 7 19 17 -2

7 8 24 23 -1

ARABPSYCHOLOGY.COM

ARABPSYCHOLOGY.COM