How can I select rows in Pandas where two columns have equal values?

Authored by

stats writer

June 26, 2024

In order to select rows in Pandas where two columns have equal values, one can use the boolean indexing method. This involves creating a condition where the values in the two columns are compared using the "==" operator, and then passing this condition as an argument in the dataframe's indexing function. The result will be a new dataframe with only the rows where the specified columns have equal values. This method can be useful for data analysis and manipulation tasks, such as identifying duplicates or finding correlations between variables.

Pandas: Select Rows where Two Columns Are Equal

You can use the following methods to select rows in a pandas DataFrame where two columns are (or are not) equal:

Method 1: Select Rows where Two Columns Are Equal

df.query('column1 == column2')

Method 2: Select Rows where Two Columns Are Not Equal

df.query('column1 != column2')

The following examples show how to use each method in practice with the following pandas DataFrame:

import pandas as pd

#create DataFrame

df = pd.DataFrame({'painting': ,

'rater1': ,

'rater2': })

#view DataFrame

print(df)

painting rater1 rater2

0 A Good Good

1 B Good Bad

2 C Bad Bad

3 D Bad Good

4 E Good Good

5 F Good Good

Example 1: Select Rows where Two Columns Are Equal

We can use the following syntax to select only the rows in the DataFrame where the values in the rater1 and rater2 column are equal:

#select rows where rater1 is equal to rater2

df.query('rater1 == rater2')

painting rater1 rater2

0 A Good Good

2 C Bad Bad

4 E Good Good

5 F Good Good

Notice that only the rows where rater1 and rater2 are equal are selected.

We could also use the len() function if we simply want to count how many rows have equal values in the rater1 and rater2 columns:

```
#count the number of rows where rater1 is equal to rater2
len(df.query('rater1 == rater2'))
```

4

This tells us that there are 4 rows where the values in the rater1 and rater2 column are equal.

Example 2: Select Rows where Two Columns Are Not Equal

We can use the following syntax to select only the rows in the DataFrame where the values in the rater1 and rater2 column are not equal:

#select rows where rater1 is not equal to rater2

df.query('rater1 != rater2')

painting rater1 rater2

1 B Good Bad

3 D Bad Good

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