How can I compare columns in two different DataFrames using Pandas?

Authored by

stats writer

June 27, 2024

Pandas is a popular Python library used for data analysis and manipulation. It offers a variety of functions and tools for comparing and analyzing data. One common task is comparing columns in two different DataFrames. This can be done easily using the "compare" function in Pandas, which allows for a side-by-side comparison of the columns in the two DataFrames. This function provides a clear overview of any similarities or differences between the columns, making it a useful tool for data comparison and analysis. By using Pandas, users can efficiently and accurately compare columns in different DataFrames, aiding in data analysis and decision making processes.

Pandas: Compare Columns in Two Different DataFrames

You can use the following methods to compare columns in two different pandas DataFrames:

Method 1: Count Matching Values Between Columns

df1.isin(df2).value_counts()

Method 2: Display Matching Values Between Columns

pd.merge(df1, df2, on=, how='inner')

The following examples show how to use each method with the following pandas DataFrames:

import numpy as np
import pandas as pd

```
#create first DataFrame
df1 = pd.DataFrame({'team': ,
'points': })

#view DataFrame
print(df1)

team points
0 Mavs 22
1 Rockets 30
2 Spurs 15
3 Heat 17
4 Nets 14

#create second DataFrame
df2 = pd.DataFrame({'team': ,
'points': })

#view DataFrame
print(df2)

team points
0 Mavs 25
1 Thunder 40
2 Spurs 31
```

3 Nets 32

4 Cavs 22

Example 1: Count Matching Values Between Columns

The following code shows how to count the number of matching values between the team columns in each DataFrame:

```
#count matching values in team columns
df1.isin(df2).value_counts()
```

```
True 3
False 2
Name: team, dtype: int64
```

We can see that the two DataFrames have 3 team names in common and 2 team names that are different.

Example 2: Display Matching Values Between Columns

The following code shows how to display the actual matching values between the team columns in each DataFrame:

```
#display matching values between team columns
```

pd.merge(df1, df2, on=, how='inner')

team points_x points_y

0 Mavs 22 25

1 Spurs 15 31

2 Nets 14 32

From the output we can see that the two DataFrames have the following values in common in the team columns:

MavsSpursNets

Related:

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