How can an outer join be performed in Pandas?

Authored by

stats writer

June 25, 2024

An outer join in Pandas is a way of merging two data sets together by combining all rows from both data sets, regardless of whether they have matching values in a specific column. This can be achieved using the "merge" function in Pandas, where the "how" parameter is set to "outer". This will result in a new data set that includes all rows from both original data sets, with missing values filled in with "NaN". This type of join is useful for analyzing data sets with missing or incomplete information.

Perform an Outer Join in Pandas (With Example)

An outer join is a type of join that returns all rows from two pandas DataFrames.

You can use the following basic syntax to perform an outer join in pandas:

import pandas as pd

df1.merge(df2, on='some_column', how='outer')

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

Example: How to Perform an Outer Join in Pandas

Suppose we have the following two pandas DataFrames that contain information about various basketball teams:

import pandas as pd

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

df2 = pd.DataFrame({'team': ,
'assists': })

#view DataFrames
print(df1)

team points
0 A 18
1 B 22
2 C 19
3 D 14
4 E 14
5 F 11
6 G 20
7 H 28

print(df2)

team assists
0 A 4
1 B 9
```

2 C 14

3 D 13

4 J 10

5 K 8

We can use the following code to perform an outer join, matching the rows between the DataFrames based on the values in the team column and keeping all rows from both DataFrames:

#perform outer join
df1.merge(df2, on='team', how='outer')

team points assists
0 A 18.0 4.0
1 B 22.0 9.0
2 C 19.0 14.0
3 D 14.0 13.0
4 E 14.0 NaN
5 F 11.0 NaN
6 G 20.0 NaN
7 H 28.0 NaN
8 J NaN 10.0
9 K NaN 8.0

The result is a DataFrame that contains all rows from each DataFrame.

Notice that NaN values have been filled in for each row where the value in the team column did not exist in both DataFrames.

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

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