How do I merge two Pandas DataFrames on their index?

Authored by

stats writer

April 18, 2024

"Merging two Pandas DataFrames on their index is a process of combining two data structures based on the common index values. This can be achieved using the 'merge' function in Pandas, where the index of both DataFrames becomes the key for joining the data. The resulting merged DataFrame will have a new index, with rows from both original DataFrames paired up and any missing values filled with NaN. This technique is useful for consolidating and analyzing data from multiple sources, allowing for efficient and accurate data manipulation and analysis."

Merge Two Pandas DataFrames on Index

Often you may want to merge two pandas DataFrames by their indexes. There are three ways to do so in pandas:

1. Use join: By default, this performs a left join.

df1.join(df2)

2. Use merge. By default, this performs an inner join.

pd.merge(df1, df2, left_index=True, right_index=True)

3. Use concat. By default, this performs an outer join.

pd.concat(, axis=1)

The following examples illustrate how to use each of these functions on the following two pandas

DataFrames:

```
import pandas as pd

#create first DataFrame
df1 = pd.DataFrame({'rating': ,
'points': },
index=list('abcdefgh'))

print(df1)

rating points
a 90 25
b 85 20
c 82 14
d 88 16
e 94 27
f 90 20
g 76 12
h 75 15

#create second DataFrame
df2 = pd.DataFrame({'assists': ,
'rebounds': },
index=list('acdgmn'))
```

print(df2)

assists rebounds
a 5 11
c 7 8
d 7 10
g 8 6
m 5 6
n 7 9

Example 1: Merge DataFrames Using Join

The following code shows how to use join() to merge the two DataFrames:

df1.join(df2)

rating points assists rebounds
a 90 25 5.0 11.0
b 85 20 NaN NaN
c 82 14 7.0 8.0
d 88 16 7.0 10.0
e 94 27 NaN NaN
f 90 20 NaN NaN
g 76 12 8.0 6.0

h 75 15 NaN NaN

The join() function performs a left join by default, so each of the indexes in the first DataFrame are kept.

Example 2: Merge DataFrames Using Merge

The following code shows how to use merge() to merge the two DataFrames:

```
pd.merge(df1, df2, left_index=True, right_index=True)
  rating points assists rebounds
a     90     25       5       11
c     82     14       7        8
d     88     16       7       10
g     76     12       8        6
```

The merge() function performs an inner join by default, so only the indexes that appear in both DataFrames are kept.

Example 3: Merge DataFrames Using Concat

The following code shows how to use concat() to merge the two DataFrames:

pd.concat(, axis=1)

rating points assists rebounds
a 90.0 25.0 5.0 11.0
b 85.0 20.0 NaN NaN
c 82.0 14.0 7.0 8.0
d 88.0 16.0 7.0 10.0
e 94.0 27.0 NaN NaN
f 90.0 20.0 NaN NaN
g 76.0 12.0 8.0 6.0
h 75.0 15.0 NaN NaN
m NaN NaN 5.0 6.0
n NaN NaN 7.0 9.0

The concat() function performs an outer join by default, so each index value from each DataFrame is kept.

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

How to Stack Multiple Pandas DataFrames
How to Insert a Column Into a Pandas DataFrame