

What should I do when columns overlap but no suffix is specified?

Authored by
stats writer

May 1, 2024

RECOMMENDED CITATION

stats writer (2024). *What should I do when columns overlap but no suffix is specified?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=141774>

When columns in a table overlap but no specific suffix is specified, it is recommended to use a unique identifier or key to distinguish between the columns. This can be achieved by adding a suffix or prefix to the column name, or by creating a new column with a unique name. This will prevent confusion and ensure accurate data representation in the table. It is important to maintain consistency in naming conventions to avoid any potential errors or misinterpretation of the data.

Fix: columns overlap but no suffix specified

One error you may encounter when using pandas is:

**ValueError: columns overlap but no suffix specified:
Index(, dtype='object')**

This error occurs when you attempt to join together two data frames that share at least one common column name and a suffix is not provided for either the left or right data frame to distinguish between the columns in the new data frame.

There are two ways to fix this error:

Solution 1: Provide suffix names.

```
df1.join(df2, how = 'left', lsuffix='left', rsuffix='right')
```

Solution 2: Use the merge function instead.

```
df1.merge(df2, how = 'left')
```

The following example shows how to fix this error in practice.

How to Reproduce the Error

Suppose we attempt to join together the following two data frames:

```
import pandas as pd
```

```
#create first data frame
```

```
df1 = pd.DataFrame({'player': ,  
'points': ,  
'assists': })
```

```
#create second data frame
```

```
df2 = pd.DataFrame({'player': ,  
'rebounds': ,  
'steals': })
```

```
#attempt to perform left join on data frames
```

```
df1.join(df2, how = 'left')
```

ValueError: columns overlap but no suffix specified:

Index(, dtype='object')

We receive an error because the two data frames both share the "player" column, but there is no suffix provided for either the left or right data frame to distinguish between the columns in the new data frame.

How to Fix the Error

One way to fix this error is to provide a suffix name for either the left or right data frame:

```
#perform left join on data frames with suffix provided  
df1.join(df2, how = 'left', lsuffix='left', rsuffix='right')
```

```
playerleft points assists playerright rebounds steals
```

```
0 A 5 11 A 4 2
```

```
1 B 7 8 B 4 2
```

```
2 C 7 10 C 6 1
```

```
3 D 9 6 D 9 4
```

```
4 E 12 6 E 13 3
```

```
5 F 9 5 F 16 2
```

Another way to fix this error is to simply use the `merge()` function, which doesn't encounter this problem

when joining two data frames together:

#merge two data frames

df1.merge(df2, how = 'left')

player points assists rebounds steals

0 A 5 11 4 2

1 B 7 8 4 2

2 C 7 10 6 1

3 D 9 6 9 4

4 E 12 6 13 3

5 F 9 5 16 2

Notice that the merge() function simply drops any names from the second data frame that already belong to the first data frame.