

How can I perform a left join in Pandas? Can you provide an example?

Authored by
stats writer

June 29, 2024

RECOMMENDED CITATION

stats writer (2024). *How can I perform a left join in Pandas? Can you provide an example?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=159281>

A left join is a type of data merging operation in pandas that combines two dataframes based on a common key column, while retaining all rows from the left dataframe and filling in the missing values from the right dataframe. This can be achieved by using the "merge" function in pandas with the parameter "how='left'". An example of this would be merging a dataframe of customer information with a dataframe of their purchase history, using the "customer_id" column as the common key. This would result in a new dataframe with all customer information and their corresponding purchase history, with null values for customers who have not made any purchases.

Do a Left Join in Pandas (With Example)

You can use the following basic syntax to perform a left join in pandas:

```
import pandas as pd
```

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

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

Example: How to Do Left Join in Pandas

Suppose we have the following two pandas DataFrames that contains 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 G 10

5 H 8

We can use the following code to perform a left join, keeping all of the rows from the first DataFrame and adding any columns that match based on the team column in the second DataFrame:

```
#perform left join
```

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

```
team points assists
```

```
0 A 18 4.0
```

```
1 B 22 9.0
```

```
2 C 19 14.0
```

```
3 D 14 13.0
```

```
4 E 14 NaN
```

```
5 F 11 NaN
```

```
6 G 20 10.0
```

```
7 H 28 8.0
```

Every team from the left DataFrame (df1) is returned in the merged DataFrame and only the rows in the right

DataFrame (df2) that match a team name in the left DataFrame are returned.

Notice that the two teams in df2 (teams E and F) that do not match a team name in df1 simply return a NaN value in the assists column of the merged DataFrame.

Note that you can also use `pd.merge()` with the following syntax to return the exact same result:

```
#perform left join
```

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

```
team points assists
```

```
0 A 18 4.0
```

```
1 B 22 9.0
```

```
2 C 19 14.0
```

```
3 D 14 13.0
```

```
4 E 14 NaN
```

```
5 F 11 NaN
```

```
6 G 20 10.0
```

```
7 H 28 8.0
```

Notice that this merged DataFrame matches the one from the previous example.

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

Additional Resources

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

ARABPSYCHOLOGY.COM