

How can I perform an inner join using Pandas, and can you provide an example of how it works?

Authored by
stats writer

June 29, 2024

RECOMMENDED CITATION

stats writer (2024). *How can I perform an inner join using Pandas, and can you provide an example of how it works?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=159302>

Inner join is a method of combining data from two or more data sets based on a shared key or column. This can be easily performed using the Pandas library in Python. The inner join function in Pandas allows for the merging of data frames by matching values in a specified column, creating a new data frame with only the matching rows. This can be useful for combining data sets that share common information, such as customer or product data. An example of how this works is if we have two data frames, one with customer information and another with transaction information, we can use the inner join function to merge the two data frames based on the shared customer ID column, creating a new data frame with all the relevant information in one place. This allows for easy analysis and manipulation of the data.

Do an Inner Join in Pandas (With Example)

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

```
import pandas as pd
```

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

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

Example: How to Do Inner 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 an inner join, which only keeps the rows where the team name appears in both DataFrames:

```
#perform left join  
df1.merge(df2, on='team', how='inner')
```

team points assists

0 A 18 4

1 B 22 9

2 C 19 14

3 D 14 13

4 G 20 10

5 H 28 8

The only rows contained in the merged DataFrame are the ones where the team name appears in both DataFrames.

Notice that two teams were dropped (teams E and F)

because they didn't appear in both DataFrames.

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='inner')
```

```
team points assists
```

```
0 A 18 4
```

```
1 B 22 9
```

```
2 C 19 14
```

```
3 D 14 13
```

```
4 G 20 10
```

```
5 H 28 8
```

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: