How can I drop rows in a Pandas DataFrame based on a condition?

Authored by

stats writer

July 2, 2024

The process of dropping rows in a Pandas DataFrame based on a condition involves using the .drop() method to remove rows that do not meet a specified criteria. This can be achieved by using conditional statements or by using the .loc function to select rows based on a specific condition. By dropping rows, the DataFrame can be filtered to only include the desired data, making it easier to work with and analyze. This method is useful for cleaning and organizing data in a DataFrame.

Drop Rows in Pandas DataFrame Based on Condition

We can use the following syntax to drop rows in a pandas DataFrame based on condition:

Method 1: Drop Rows Based on One Condition

df = df

Method 2: Drop Rows Based on Multiple Conditions

df = df

Note: We can also use the drop() function to drop rows from a DataFrame, but this function has been shown to be much slower than just assigning the DataFrame to a filtered version of itself.

The following examples show how to use this syntax in practice with the following pandas DataFrame:

```
import pandas as pd

#create DataFrame
df = pd.DataFrame({'team': ,
'pos': ,
'assists': ,
'rebounds': })

#view DataFrame
df
```

team pos assists rebounds
0 A G 5 11
1 A G 7 8
2 A F 7 10
3 A F 9 6
4 B G 12 6
5 B G 9 5
6 B F 9 9
7 B F 4 12

Method 1: Drop Rows Based on One Condition

The following code shows how to drop rows in the DataFrame based on one condition:

#drop rows where value in 'assists' column is less than or equal to 8
df = df

#view updated DataFrame
df

team pos assists rebounds
3 A F 9 6
4 B G 12 6
5 B G 9 5
6 B F 9 9

Any row that had a value less than or equal to 8 in the 'assists' column was dropped from the DataFrame.

Method 2: Drop Rows Based on Multiple Conditions

The following code shows how to drop rows in the DataFrame based on multiple conditions:

#only keep rows where 'assists' is greater than 8 and rebounds is greater than 5
df = df

#view updated DataFrame

df

team pos assists rebounds
3 A F 9 6
4 B G 12 6
5 B G 9 5
6 B F 9 9

The only rows that we kept in the DataFrame were the ones where the assists value was greater than 8 *and* the rebounds value was greater than 5.

Note that we can also use the | operator to apply an "or" filter:

#only keep rows where 'assists' is greater than 8 or rebounds is greater than 10
df = df

#view updated DataFrame
df

team pos assists rebounds
0 A G 5 11
3 A F 9 6

4 B G 12 6

5 B G 9 5

6 B F 9 9

7 B F 4 12

The only rows that we kept in the DataFrame were the ones where the assists value was greater than 8 or the rebounds value was greater than 10.

Any rows that didn't meet one of these conditions was dropped.

Additional Resources

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