How can I replace values in a column in Pandas based on a condition?

Authored by

stats writer

July 2, 2024

Replacing values in a column in Pandas based on a condition involves using the built-in "where" method or the "loc" function to specify the condition and provide the new value to be replaced. This allows for efficient and targeted replacement of specific values in a column, based on a given condition. This process is commonly used in data manipulation and cleaning tasks to ensure accurate and consistent data.

Pandas: Replace Values in Column Based on Condition

You can use the following basic syntax to replace values in a column of a pandas DataFrame based on a condition:

#replace values in 'column1' that are greater than 10 with 20
df.loc > 10, 'column1'] = 20

The following examples show how to use this syntax in practice.

Example 1: Replace Values in Column Based on One Condition

Suppose we have the following pandas DataFrame:

import pandas as pd

#create DataFrame
df = pd.DataFrame({'team': ,
'position': ,

'points': ,
'assists': })

#view DataFrame
df

team position points assists
0 A G 5 3
1 A G 7 8
2 A F 7 2
3 A F 9 6
4 B G 12 6
5 B G 13 5
6 B F 9 9
7 B F 14 5

We can use the following code to replace every value in the 'points' column that is greater than 10 with a value of 20:

#replace any values in 'points' column greater than 10 with 20
df.loc > 10, 'points'] = 20

#view updated DataFrame

df

team position points assists
0 A G 5 3
1 A G 7 8
2 A F 7 2
3 A F 9 6
4 B G 20 6
5 B G 20 5
6 B F 9 9
7 B F 20 5

Notice that each of the three values in the 'points' column that were greater than 10 got replaced with the value 20.

Example 2: Replace Values in Column Based on Multiple Conditions

Suppose we have the following pandas DataFrame:

```
import pandas as pd

#create DataFrame
df = pd.DataFrame({'team': ,
'position': ,
'points': ,
```

'assists': })

#view DataFrame

df

team position points assists

0 A G 5 3

1 A G 7 8

2 A F 7 2

3 A F 9 6

4 B G 12 6

5 B G 13 5

6 B F 9 9

7 B F 14 5

We can use the following code to replace every value in the 'position' column where points is less than 10 *or* where assists is less than 5 with the string 'Bad':

#replace string in 'position' column with 'bad' if points < 10 or assists < 5

df.loc < 10) | (df < 5), 'position'] = 'Bad'

#view updated DataFrame

df

team position points assists
0 A Bad 5 3
1 A Bad 7 8
2 A Bad 7 2
3 A Bad 9 6
4 B G 20 6
5 B G 20 5
6 B Bad 9 9
7 B F 20 5

Similarly, we can use the following code to replace every value in the 'position' column where points is less than 10 and where assists is less than 5 with the string 'Bad':

```
#replace string in 'position' column with 'bad' if points < 10 and assists < 5
df.loc < 10) & (df < 5), 'position'] = 'Bad'

#view updated DataFrame
df
```

team position points assists
0 A Bad 5 3
1 A G 7 8

2 A Bad 7 2

3 A F 9 6

4 B G 12 6

5 B G 13 5

6 B F 9 9

7 B F 14 5

Notice that the two rows where points was less than 10 *and* assists was less than 5 had their 'position' value replaced with the string 'Bad'.

Additional Resources

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