

How can I create a new column in a Pandas dataframe based on a condition or criteria?

Authored by
stats writer

April 17, 2024

RECOMMENDED CITATION

stats writer (2024). *How can I create a new column in a Pandas dataframe based on a condition or criteria?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=136463>

To create a new column in a Pandas dataframe based on a condition or criteria, one can use the "df.loc" function and specify the condition or criteria within brackets. This will allow for the creation of a new column with values that meet the specified condition or criteria. The process involves using boolean indexing to select rows that meet the condition and assigning a new column name and values to those rows. By following this method, users can efficiently add new columns to their dataframes based on specific conditions or criteria.

Create a New Column Based on a Condition in Pandas

Often you may want to create a new column in a pandas DataFrame based on some condition.

This tutorial provides several examples of how to do so using the following DataFrame:

```
import pandas as pd
import numpy as np

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

#view DataFrame
df
rating points assists rebounds
0 90 25 5 11
```

1 85 20 7 8
2 82 14 7 10
3 88 16 8 6
4 94 27 5 6
5 90 20 7 9
6 76 12 6 6
7 75 15 9 10
8 87 14 9 10
9 86 19 5 7

Example 1: Create a New Column with Binary Values

The following code shows how to create a new column called 'Good' where the value is 'yes' if the points in a given row is above 20 and 'no' if not:

```
#create new column titled 'Good'  
df = np.where(df>20, 'yes', 'no')
```

```
#view DataFrame
```

```
df
```

```
rating points assists rebounds Good
```

```
0 90 25 5 11 yes
```

```
1 85 20 7 8 no
```

2 82 14 7 10 no
3 88 16 8 6 no
4 94 27 5 6 yes
5 90 20 7 9 no
6 76 12 6 6 no
7 75 15 9 10 no
8 87 14 9 10 no
9 86 19 5 7 no

Example 2: Create a New Column with Multiple Values

The following code shows how to create a new column called 'Good' where the value is:

'Yes' if the points ≥ 25 'Maybe' if $15 \leq \text{points} < 25$ 'No' if points < 15

```
#define function for classifying players based on points
def f(row):
    if row < 15:
        val = 'no'
    elif row < 25:
        val = 'maybe'
    else:
        val = 'yes'
```

return val

#create new column 'Good' using the function above

df = df.apply(f, axis=1)

#view DataFrame

df

rating points assists rebounds Good

0 90 25 5 11 yes

1 85 20 7 8 maybe

2 82 14 7 10 no

3 88 16 8 6 maybe

4 94 27 5 6 yes

5 90 20 7 9 maybe

6 76 12 6 6 no

7 75 15 9 10 maybe

8 87 14 9 10 no

9 86 19 5 7 maybe

Example 3: Create a New Column Based on Comparison with Existing Column

The following code shows how to create a new column called 'assist_more' where the value is:

'Yes' if assists > rebounds.'No' otherwise.

```
#create new column titled 'assist_more'df =  
np.where(df>df, 'yes', 'no')
```

```
#view DataFrame
```

```
df
```

```
rating points assists rebounds assist_more
```

```
0 90 25 5 11 no
```

```
1 85 20 7 8 no
```

```
2 82 14 7 10 no
```

```
3 88 16 8 6 yes
```

```
4 94 27 5 6 no
```

```
5 90 20 7 9 no
```

```
6 76 12 6 6 no
```

```
7 75 15 9 10 no
```

```
8 87 14 9 10 no
```

```
9 86 19 5 7 no
```

You can find more Python tutorials .