

How can I use the where() function in Pandas to filter data? Can you provide some examples?

Authored by
stats writer

May 5, 2024

RECOMMENDED CITATION

stats writer (2024). *How can I use the where() function in Pandas to filter data? Can you provide some examples?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=142940>

The where() function in Pandas is a useful tool for filtering data based on certain conditions. It allows users to specify a condition and return only the rows where that condition is met. This function can be applied to a Pandas DataFrame or Series to extract the desired subset of data. For example, it can be used to filter out all rows with missing values, or to select only rows with values above a certain threshold. The where() function also supports chaining multiple conditions together. Overall, it provides a convenient way to manipulate and extract specific data from a larger dataset. Some examples of using the where() function include filtering sales data to only show sales above a certain amount, or filtering a customer database to only include customers in a specific age range.

Use where() Function in Pandas (With Examples)

The where() function can be used to replace certain values in a pandas DataFrame.

This function uses the following basic syntax:

```
df.where(cond, other=nan)
```

For every value in a pandas DataFrame where cond is True, the original value is retained.

For every value where cond is False, the original value is replaced by the value specified by the other argument.

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

```
import pandas as pd

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

```
#view DataFrame
df
```

```
points assists rebounds
0 25 5 11
1 12 7 8
2 15 7 10
3 14 9 6
4 19 12 6
5 23 9 5
6 25 9 9
7 29 4 12
```

Example 1: Replace Values in Entire DataFrame

The following code shows how to use the where() function to replace all values that don't meet a certain condition in an entire pandas DataFrame with a NaN

value.

#keep values that are greater than 7, but replace all others with NaN

df.where(df>7)

points assists rebounds

0 25 NaN 11.0

1 12 NaN 8.0

2 15 NaN 10.0

3 14 9.0 NaN

4 19 12.0 NaN

5 23 9.0 NaN

6 25 9.0 9.0

7 29 NaN 12.0

We can also use the other argument to replace values with something other than NaN.

#keep values that are greater than 7, but replace all others with 'low'

df.where(df>7, other='low')

points assists rebounds

0 25 low 11

```
1 12 low 8
2 15 low 10
3 14 9 low
4 19 12 low
5 23 9 low
6 25 9 9
7 29 low 12
```

Example 2: Replace Values in Specific Column of DataFrame

The following code shows how to use the where() function to replace all values that don't meet a certain condition in a specific column of a DataFrame.

```
#keep values greater than 15 in 'points' column, but
replace others with 'low'
df = df.where(df>15, other='low')
```

```
#view DataFrame
```

```
df
```

```
points assists rebounds
```

```
0 25 5 11
```

```
1 low 7 8
```

```
2 low 7 10
```

3 low 9 6

4 19 12 6

5 23 9 5

6 25 9 9

7 29 4 12

You can find the complete online documentation for the pandas where() function .

ARABPSYCHOLOGY.COM