How can I use ffill in Pandas based on a specific condition?

Authored by

stats writer

June 25, 2024

FFill, or forward filling, is a function in Pandas that can be used to fill missing values in a dataset by copying the previous non-null value. This can be especially useful when dealing with time series data. In order to use ffill in Pandas based on a specific condition, one can first create a conditional statement using the "where" function to identify the specific values that need to be filled. Then, the "ffill" function can be applied to only those values that meet the condition, leaving the rest of the dataset unchanged. This allows for targeted and efficient filling of missing values in a Pandas dataframe.

Pandas: Use ffill Based on Condition

You can use the following basic syntax to use the ffill() function in pandas to forward fill values based on a condition in another column:

df = df.groupby('store').ffill()

This particular example will forward fill values in the sales column only if the previous value in the store column is equal to the current value in the store column.

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

Example: Use ffill Based on Condition in Pandas

Suppose we have the following pandas DataFrame that contains information about the total sales made by two

different retail stores during four business quarters:

```
import pandas as pd
import numpy as np

#create DataFrame
df = pd.DataFrame({'store': ,
'quarter': ,
'sales': })

#view DataFrame
print(df)

store quarter sales
0 A 1 12.0
1 A 2 22.0
2 B 1 30.0
3 A 3 NaN
4 B 2 24.0
5 A 4 NaN
6 B 3 NaN
7 B 4 NaN
```

Notice that there are multiple NaN values in the sales column.

Suppose we would like to fill in these NaN values using the previous value in the sales column but we want to make sure that values correspond to the correct store.

We can use the following syntax to do so:

```
#group by store and forward fill values in sales column
df = df.groupby('store').ffill()

#view updated DataFrameprint(df)
```

```
store quarter sales
0 A 1 12.0
1 A 2 22.0
2 B 1 30.0
3 A 3 22.0
4 B 2 24.0
5 A 4 22.0
6 B 3 24.0
7 B 4 24.0
```

Notice that the NaN values in the sales column have been replaced by the previous sales value and that the values correspond to the correct store.

For example:

The NaN value in row index position 3 has been replaced by the value 22, which was the most recent value in the sales column that corresponded to store A.The NaN value in row index position 6 has been replaced by the value 24, which was the most recent value in the sales column that corresponded to store B.

And so on.

Note: You can find the complete documentation for the pandas ffill() function .