

How can I fill NaN values in a Pandas dataframe with the mode value?

Authored by
stats writer

June 28, 2024

RECOMMENDED CITATION

stats writer (2024). *How can I fill NaN values in a Pandas dataframe with the mode value?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=156007>

The process of filling NaN (Not a Number) values in a Pandas dataframe with the mode value involves identifying the most frequently occurring value in a particular column or row and replacing the NaN values with that value. This technique is useful for handling missing data and maintaining the integrity of the dataset. By using the mode value to fill in the missing values, the overall distribution and patterns in the data can be preserved. This can be achieved by using the built-in "fillna" function in Pandas, which allows for the replacement of NaN values with the mode value of the respective column or row. This approach ensures that the data remains accurate and can be effectively used for analysis and other data manipulation tasks.

Pandas: Fill NaN Values with Mode

You can use the following syntax to replace NaN values in a column of a pandas DataFrame with the mode value of the column:

```
df = df.fillna(df.mode())
```

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

Example: Replace Missing Values with Mode in Pandas

Suppose we have the following pandas DataFrame with some missing values:

```
import numpy as np
```

```
import pandas as pd
```

```
#create DataFrame with some NaN values
```

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

```
#view DataFrame
```

```
df
```

```
rating points assists rebounds
```

```
0 NaN 25.0 5.0 11
```

```
1 85.0 NaN 7.0 8
```

```
2 NaN 14.0 7.0 10
```

```
3 88.0 16.0 NaN 6
```

```
4 94.0 27.0 5.0 6
```

```
5 90.0 20.0 7.0 9
```

```
6 75.0 12.0 6.0 6
```

```
7 75.0 15.0 9.0 10
```

```
8 87.0 14.0 9.0 10
```

```
9 86.0 19.0 7.0 7
```

We can use the `fillna()` function to fill the NaN values in the rating column with the mode value of the rating column:

```
#fill NaNs with column mode in 'rating' column df =
```

```
df.fillna(df.mode())
```

```
#view updated DataFrame
```

```
df
```

```
rating points assists rebounds
```

```
0 75.0 25.0 5.0 11
```

```
1 85.0 NaN 7.0 8
```

```
2 75.0 14.0 7.0 10
```

```
3 88.0 16.0 NaN 6
```

```
4 94.0 27.0 5.0 6
```

```
5 90.0 20.0 7.0 9
```

```
6 75.0 12.0 6.0 6
```

```
7 75.0 15.0 9.0 10
```

```
8 87.0 14.0 9.0 10
```

```
9 86.0 19.0 7.0 7
```

The mode value in the rating column was 75 so each of the NaN values in the rating column were filled with this value.

Note: You can find the complete online documentation for the fillna() function .

Additional Resources

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

ARABPSYCHOLOGY.COM