

How can I use Pandas to combine values from multiple columns into one column using the Coalesce function?

Authored by
stats writer

July 1, 2024

RECOMMENDED CITATION

stats writer (2024). *How can I use Pandas to combine values from multiple columns into one column using the Coalesce function?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=164847>

Pandas is a powerful Python library that allows for efficient data analysis and manipulation. One useful feature of Pandas is the ability to combine values from multiple columns into a single column using the Coalesce function. This function can be used to select the first non-null value from a group of columns and combine them into a new column. This is particularly useful when dealing with messy or incomplete data, as it allows for the consolidation of information from different columns into a single, more organized column. By using the Coalesce function in Pandas, users can efficiently merge data from multiple sources and create a more comprehensive dataset for further analysis.

Pandas: Coalesce Values from Multiple Columns into One

You can use the following methods to coalesce the values from multiple columns of a pandas DataFrame into one column:

Method 1: Coalesce Values by Default Column Order

```
df = df.bfill(axis=1).iloc
```

Method 2: Coalesce Values Using Specific Column Order

```
df = df[j].bfill(axis=1).iloc
```

The following examples show how to use each method in practice with the following pandas DataFrame:

```
import pandas as pd
import numpy as np

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

```
#view DataFrame
print(df)
```

```
points assists rebounds
0 NaN NaN 3.0
1 NaN 7.0 4.0
2 19.0 7.0 NaN
3 NaN 9.0 NaN
4 14.0 NaN 6.0
```

Method 1: Coalesce Values by Default Column Order

The following code shows how to coalesce the values in the points, assists, and rebounds columns into one column, using the first non-null value across the three columns as the coalesced value:

```
#create new column that contains first non-null value
```

from three existing columns

```
df = df.bfill(axis=1).iloc
```

```
#view updated DataFrame
```

```
print(df)
```

points assists rebounds coalesce

0 NaN NaN 3.0 3.0

1 NaN 7.0 4.0 7.0

2 19.0 7.0 NaN 19.0

3 NaN 9.0 NaN 9.0

4 14.0 NaN 6.0 14.0

Here's how the value in the coalesce column was chosen:

First row: The first non-null value was 3.0. Second row: The first non-null value was 7.0. Third row: The first non-null value was 19.0. Fourth row: The first non-null value was 9.0. Fifth row: The first non-null value was 14.0.

Method 2: Coalesce Values Using Specific Column Order

The following code shows how to coalesce the values in the three columns by analyzing the columns in the following order: assists, rebounds, points.

#coalesce values in specific column order

df = df].bfill(axis=1).iloc

#view updated DataFrame

print(df)

points assists rebounds coalesce

0 NaN NaN 3.0 3.0

1 NaN 7.0 4.0 7.0

2 19.0 7.0 NaN 7.0

3 NaN 9.0 NaN 9.0

4 14.0 NaN 6.0 6.0

Here's the logic that was used to decide which value to place in the coalesce column:

If the value in the assists column is non-null then use that value. Otherwise, if the value in the rebounds column is non-null then use that value. Otherwise, if the value in the points column is non-null then use that value.

Additional Resources

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