

How can I read an Excel file with merged cells using Pandas?

Authored by
stats writer

June 25, 2024

RECOMMENDED CITATION

stats writer (2024). *How can I read an Excel file with merged cells using Pandas?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=152306>

Pandas is a popular open-source library for data analysis in Python. It provides powerful tools for reading, manipulating, and analyzing data from various sources, including Excel files. When working with Excel files, one common challenge is dealing with merged cells, where two or more cells are combined into a single larger cell. This can cause issues when trying to read the data into a Pandas dataframe. However, Pandas offers a simple solution for reading Excel files with merged cells. By using the "header=None" argument in the read_excel() function, Pandas will automatically detect and handle merged cells, ensuring that the data is correctly read into the dataframe. This feature makes working with Excel files with merged cells much more efficient and convenient, allowing users to easily extract and analyze data without having to manually handle merged cells.

Pandas: Read Excel File with Merged Cells

When you read an Excel file with merged cells into a pandas DataFrame, the merged cells will automatically be filled with NaN values.

The easiest way to fill in these NaN values after importing the file is to use the pandas fillna() function as follows:

```
df = df.fillna(method='ffill', axis=0)
```

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

Example: Read Excel File with Merged Cells in Pandas

Suppose we have the following Excel file called merged_data.xlsx that contains information about

various basketball players:

	A	B	C	D	E	F
1	Team	Player	Points	Assists		
2	Mavericks	A	22	4		
3		B	29	4		
4		C	45	3		
5		D	30	7		
6	Rockets	E	29	8		
7		F	16	6		
8		G	25	9		
9		H	20	12		
10						
11						
12						
13						
14						
15						
16						
17						
18						
19						
20						
21						

Notice that the values in the Team column are merged.

Players A through D belong to the Mavericks while players E through H belong to the Rockets.

Suppose we use the `read_excel()` function to read this Excel file into a pandas DataFrame:

```
import pandas as pd
```

```
#import Excel file
df = pd.read_excel('merged_data.xlsx')#view DataFrame
print(df)
```

Team Player Points Assists

0 Mavericks A 22 4

1 NaN B 29 4

2 NaN C 45 3

3 NaN D 30 7

4 Rockets E 29 8

5 NaN F 16 6

6 NaN G 25 9

7 NaN H 20 12

By default, pandas fills in the merged cells with NaN values.

To fill in each of these NaN values with the team names instead, we can use the fillna() function as follows:

```
#fill in NaN values with team names
```

```
df = df.fillna(method='ffill', axis=0)
```

```
#view updated DataFrame
```

```
print(df)
```

Team Player Points Assists

0 Mavericks A 22 4

1 Mavericks B 29 4

2 Mavericks C 45 3

3 Mavericks D 30 7

4 Rockets E 29 8

5 Rockets F 16 6

6 Rockets G 25 9

7 Rockets H 20 12

Notice that each of the NaN values has been filled in with the appropriate team name.

Note that the argument `axis=0` tells pandas to fill in the NaN values vertically.

To instead fill in NaN values horizontally across columns, you can specify `axis=1`.

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