

Pandas - Cleaning Empty Cells



Next >

Empty Cells

Empty cells can potentially give you a wrong result when you analyze data.

Remove Rows

One way to deal with empty cells is to remove rows that contain empty cells.

This is usually OK, since data sets can be very big, and removing a few rows will not have a big impact on the result.

Example

Get your own Python Server

Return a new Data Frame with no empty cells:

```
import pandas as pd

df = pd.read_csv('data.csv')

new_df = df.dropna()

print(new_df.to_string())
```

Try it Yourself »



If you want to change the original DataFrame, use the inplace = True argument:

Example

Remove all rows with NULL values:

```
import pandas as pd

df = pd.read_csv('data.csv')

df.dropna(inplace = True)

print(df.to_string())
```

Try it Yourself »

Note: Now, the dropna(inplace = True) will NOT return a new DataFrame, but it will remove all rows containing NULL values from the original DataFrame.

Replace Empty Values

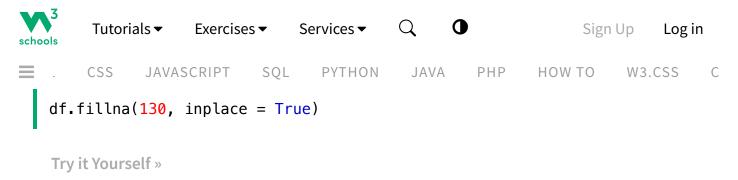
Another way of dealing with empty cells is to insert a *new* value instead.

This way you do not have to delete entire rows just because of some empty cells.

The fillna() method allows us to replace empty cells with a value:

Example

Replace NULL values with the number 130:



Replace Only For Specified Columns

The example above replaces all empty cells in the whole Data Frame.

To only replace empty values for one column, specify the *column name* for the DataFrame:

Example

Replace NULL values in the "Calories" columns with the number 130:

```
import pandas as pd

df = pd.read_csv('data.csv')

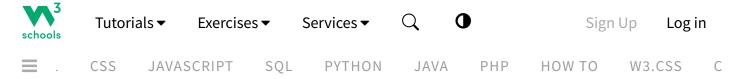
df["Calories"].fillna(130, inplace = True)
```

Try it Yourself »

Get Certified!



Complete the Pandas modules, do the exercises, take the exam, and you will become w3schools certified!



Replace Using Mean, Median, or Mode

A common way to replace empty cells, is to calculate the mean, median or mode value of the column.

Pandas uses the mean() median() and mode() methods to calculate the respective values for a specified column:

Example

Calculate the MEAN, and replace any empty values with it:

```
import pandas as pd

df = pd.read_csv('data.csv')

x = df["Calories"].mean()

df["Calories"].fillna(x, inplace = True)
```

Try it Yourself »

Mean = the average value (the sum of all values divided by number of values).

Example

Calculate the MEDIAN, and replace any empty values with it:

Median = the value in the middle, after you have sorted all values ascending.

Example

Calculate the MODE, and replace any empty values with it:

```
import pandas as pd

df = pd.read_csv('data.csv')

x = df["Calories"].mode()[0]

df["Calories"].fillna(x, inplace = True)

Try it Yourself »
```

Mode = the value that appears most frequently.



Log in to track progress

Next >



Tutorials **▼**

Exercises **▼**

Services **▼**



Sign Up

Log in

CSS **JAVASCRIPT** SQL

PYTHON

JAVA

PHP

HOW TO

W3.CSS



COLOR PICKER













SPACES

UPGRADE

AD-FREE

NEWSLETTER

GET CERTIFIED

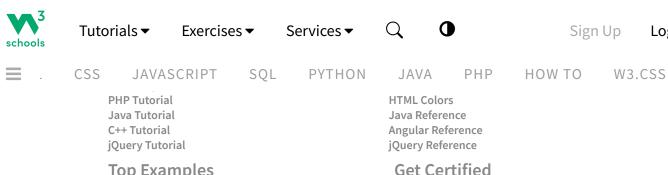
REPORT ERROR

Top Tutorials

Top References

HTML Tutorial CSS Tutorial

HTML Reference **CSS** Reference



Top Examples

HTML Examples CSS Examples JavaScript Examples How To Examples SQL Examples Python Examples W3.CSS Examples Bootstrap Examples PHP Examples Java Examples **XML Examples ¡Query Examples**

HTML Certificate CSS Certificate JavaScript Certificate Front End Certificate **SOL** Certificate **Python Certificate PHP Certificate ¡Query Certificate** Java Certificate C++ Certificate **C# Certificate XML Certificate**



W3Schools is optimized for learning and training. Examples might be simplified to improve reading and learning.

Tutorials, references, and examples are constantly reviewed to avoid errors, but we cannot warrant full correctness

of all content. While using W3Schools, you agree to have read and accepted our terms of use, cookie and privacy policy.

Copyright 1999-2023 by Refsnes Data. All Rights Reserved. W3Schools is Powered by W3.CSS.

Log in