Tutorials ▾   Exercises ▾   Services ▾   🔍   ◐   Sign Up   Log in

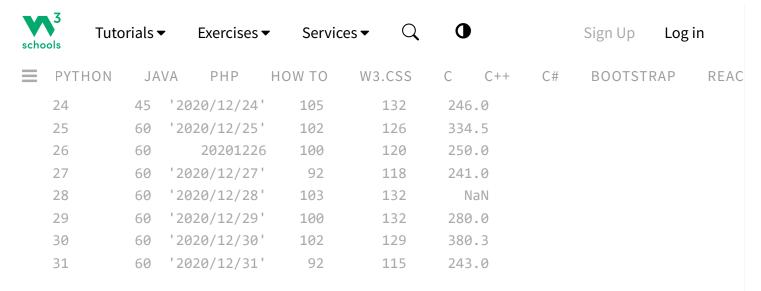☰   PYTHON   JAVA   PHP   HOW TO   W3.CSS   C   C++   C#   BOOTSTRAP   REAC

Pandas - Removing Duplicates

[‹ Previous]   [Next ›]

Discovering Duplicates

Duplicate rows are rows that have been registered more than one time.

```
     Duration          Date  Pulse  Maxpulse   Calories
0          60  '2020/12/01'    110       130      409.1
1          60  '2020/12/02'    117       145      479.0
2          60  '2020/12/03'    103       135      340.0
3          45  '2020/12/04'    109       175      282.4
4          45  '2020/12/05'    117       148      406.0
5          60  '2020/12/06'    102       127      300.0
6          60  '2020/12/07'    110       136      374.0
7         450  '2020/12/08'    104       134      253.3
8          30  '2020/12/09'    109       133      195.1
9          60  '2020/12/10'     98       124      269.0
10         60  '2020/12/11'    103       147      329.3
11         60  '2020/12/12'    100       120      250.7
12         60  '2020/12/12'    100       120      250.7
13         60  '2020/12/13'    106       128      345.3
14         60  '2020/12/14'    104       132      379.3
15         60  '2020/12/15'     98       123      275.0
16         60  '2020/12/16'     98       120      215.2
17         60  '2020/12/17'    100       120      300.0
18         45  '2020/12/18'     90       112        NaN
19         60  '2020/12/19'    103       123      323.0
```

```
24      45   '2020/12/24'    105      132      246.0
25      60   '2020/12/25'    102      126      334.5
26      60       20201226    100      120      250.0
27      60   '2020/12/27'     92      118      241.0
28      60   '2020/12/28'    103      132        NaN
29      60   '2020/12/29'    100      132      280.0
30      60   '2020/12/30'    102      129      380.3
31      60   '2020/12/31'     92      115      243.0
```

By taking a look at our test data set, we can assume that row 11 and 12 are duplicates.

To discover duplicates, we can use the `duplicated()` method.

The `duplicated()` method returns a Boolean values for each row:

Example

Get your own Python Server

Returns `True` for every row that is a duplicate, otherwise `False`:

```
print(df.duplicated())
```

Try it Yourself »

Removing Duplicates

To remove duplicates, use the `drop_duplicates()` method.

Example

Remove all duplicates:

Remember: The `(inplace = True)` will make sure that the method does NOT return a *new* DataFrame, but it will remove all duplicates from the *original* DataFrame.

Exercise [?]

What are duplicate rows in a DataFrame?

○   Rows with similar content

○   Identical rows

○   Rows where all columns of that row have the same value

Submit Answer »

Track your progress - it's free!

Pandas
Learning
Resources
For Teachers
CHECK IT OUT!

COLOR PICKER

Sign Up   Log in

Top Tutorials

HTML Tutorial
CSS Tutorial
JavaScript Tutorial
How To Tutorial
SQL Tutorial
Python Tutorial
W3.CSS Tutorial
Bootstrap Tutorial
PHP Tutorial
Java Tutorial
C++ Tutorial
jQuery Tutorial

Top References

HTML Reference
CSS Reference
JavaScript Reference
SQL Reference
Python Reference
W3.CSS Reference
Bootstrap Reference
PHP Reference
HTML Colors
Java Reference
Angular Reference
jQuery 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
jQuery Examples

Get Certified

HTML Certificate
CSS Certificate
JavaScript Certificate
Front End Certificate
SQL Certificate
Python Certificate
PHP Certificate
jQuery Certificate
Java Certificate
C++ Certificate
C# Certificate
XML Certificate

FORUM   ABOUT   ACADEMY

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

Tutorials ▾   Exercises ▾   Services ▾

Sign Up   Log in

☰   PYTHON   JAVA   PHP   HOW TO   W3.CSS   C   C++   C#   BOOTSTRAP   REAC