How can I drop duplicate columns in Pandas, and what are some examples of doing so?

Authored by

stats writer

May 6, 2024

Pandas is a popular Python library used for data analysis and manipulation. One common task in data cleaning is removing duplicate columns from a dataset. This can be achieved using the "drop_duplicates()" function in Pandas. This function allows users to drop columns that have identical values, keeping only one of them.
Examples of using this function include removing columns with misspelled names, columns with duplicate information, or columns with redundant data. The function can also be used to drop columns based on specific criteria, such as dropping columns with a certain percentage of missing values. By dropping duplicate columns, the resulting dataset becomes more concise and easier to work with.

Drop Duplicate Columns in Pandas (With Examples)

You can use the following basic syntax to drop duplicate columns in pandas:

df.T.drop_duplicates().T

The following examples show how to use this syntax in practice.

Example: Drop Duplicate Columns in Pandas

Suppose we have the following pandas DataFrame:

import pandas as pd

#create DataFrame with duplicate columns
df = pd.DataFrame({'team': ,
'points': ,
'assists': ,

'rebounds': })

df.columns =

#view DataFrame
df

team points points rebounds
0 A 25 25 11
1 A 12 12 8
2 A 15 15 10
3 A 14 14 6
4 B 19 19 6
5 B 23 23 5
6 B 25 25 9
7 B 29 29 12

We can use the following code to remove the duplicate 'points' column:

#remove duplicate columns
df.T.drop_duplicates().T

team points rebounds
0 A 25 11

1 A 12 8

2 A 15 10

3 A 14 6

4 B 19 6

5 B 23 5

6 B 25 9

7 B 29 12

Notice that the 'points' column has been removed while all other columns remained in the DataFrame.

It's also worth noting that this code will remove duplicate columns even if the columns have different names, yet contain identical values.

For example, suppose we have the following pandas DataFrame:

```
import pandas as pd

#create DataFrame with duplicate columns
df = pd.DataFrame({'team': ,
'points': ,
'points2': ,
'rebounds': })
```

#view DataFrame
df

team points points2 rebounds
0 A 25 25 11
1 A 12 12 8
2 A 15 15 10
3 A 14 14 6
4 B 19 19 6
5 B 23 23 5
6 B 25 25 9
7 B 29 29 12

Notice that the 'points' and 'points2' columns contain identical values.

We can use the following code to remove the duplicate 'points2' column:

#remove duplicate columns
df.T.drop_duplicates().T

team points rebounds
0 A 25 11
1 A 12 8

2 A 15 10

3 A 14 6

4 B 19 6

5 B 23 5

6 B 25 9

7 B 29 12

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