

How can I drop columns in a Pandas dataframe if the column name contains a specific string?

Authored by
stats writer

June 25, 2024

RECOMMENDED CITATION

stats writer (2024). *How can I drop columns in a Pandas dataframe if the column name contains a specific string?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=151616>

To drop columns in a Pandas dataframe based on a specific string in their names, the user can use the "contains" method in conjunction with the "drop" method. This will allow for the removal of all columns containing the specified string, providing a more efficient and streamlined approach to data manipulation.

Pandas: Drop Columns if Name Contains Specific String

You can use the following methods to drop columns from a pandas DataFrame whose name contains specific strings:

Method 1: Drop Columns if Name Contains Specific String

```
df.drop(list(df.filter(regex='this_string')), axis=1, inplace=True)
```

Method 2: Drop Columns if Name Contains One of Several Specific Strings

```
df.drop(list(df.filter(regex='string1|string2|string3')), axis=1, inplace=True)
```

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

```
import pandas as pd
```

```
#create DataFrame
```

```
df = pd.DataFrame({'team_name': ,  
'team_location': ,  
'player_name': ,  
'points': })
```

```
#view DataFrame
```

```
print(df)
```

```
team_name team_location player_name points  
0 A AU Andy 22  
1 B AU Bob 29  
2 C EU Chad 35  
3 D EU Dan 30  
4 E AU Ed 18  
5 F EU Fran 12
```

Example 1: Drop Columns if Name Contains Specific String

We can use the following syntax to drop all columns in the DataFrame that contain 'team' anywhere in the column name:

```
#drop columns whose name contains 'team'
```

```
df.drop(list(df.filter(regex='team')), axis=1, inplace=True)
```

```
#view updated DataFrame
```

```
print(df)
```

```
player_name points
```

```
0 Andy 22
```

```
1 Bob 29
```

```
2 Chad 35
```

```
3 Dan 30
```

```
4 Ed 18
```

```
5 Fran 12
```

Notice that both columns that contained 'team' in the name have been dropped from the DataFrame.

Example 2: Drop Columns if Name Contains One of Several Specific Strings

We can use the following syntax to drop all columns in the DataFrame that contain 'player' or 'points' anywhere in the column name:

```
#drop columns whose name contains 'player' or 'points'  
df.drop(list(df.filter(regex='player|points')), axis=1,  
inplace=True)
```

```
#view updated DataFrame
```

```
print(df)
```

```
team_name team_location
```

```
0 A AU
```

```
1 B AU
```

```
2 C EU
```

```
3 D EU
```

```
4 E AU
```

```
5 F EU
```

Notice that both columns that contained either 'player' or 'points' in the name have been dropped from the DataFrame.

Note: The | symbol in pandas is used as an "OR" operator.