How can I select columns in a Pandas DataFrame that contain a specific string?

Authored by

stats writer

June 27, 2024

Pandas is a popular Python library used for data manipulation and analysis. One common task when working with data is selecting specific columns from a DataFrame that contain a specific string. This can be achieved by using the "filter" function in Pandas, which allows you to specify a string and filter out columns that do not contain it. This function also has various parameters that allow for more specific filtering, such as ignoring case sensitivity or selecting columns that contain a substring. By using the filter function, you can efficiently select and work with columns of interest in your DataFrame, making data analysis and manipulation more streamlined.

Pandas: Select Columns Containing a Specific String

You can use the following methods to select columns that contain a particular string in a pandas DataFrame:

Method 1: Select Columns that Contain One Specific String

df.filter(regex='string1')

Method 2: Select Columns that Contain One of Several Strings

df.filter(regex='string1|string2|string3')

The following examples show how to use each of these methods in practice with the following pandas DataFrame:

import pandas as pd

```
#create DataFrame
df = pd.DataFrame({'mavs': ,
'cavs': ,
'hornets': ,
'spurs': ,
'nets': })

#view DataFrame
print(df)

mavs cavs hornets spurs nets
0 10 18 5 10 10
1 12 22 7 12 14
2 14 19 7 14 25
3 15 14 9 13 22
4 19 14 12 13 25
5 22 11 9 19 17
6 27 20 14 22 12
```


Example 1: Select Columns that Contain One Specific String

The following code shows how to use the filter() function to select only the columns that contain the string "avs" somewhere in their name:

```
#select columns that contain 'avs' in the name
df2 = df.filter(regex='avs')

#view DataFrame
print(df2)
```

```
  mavs cavs
0   10   18
1   12   22
2   14   19
3   15   14
4   19   14
5   22   11
6   27   20
```

Only the columns that contain "avs" in the name are returned.

In this case, "mavs" and "cavs" are the only columns that are returned.

Example 2: Select Columns that Contain One of Several Strings

The following code shows how to use the filter() function to select only the columns that contain "avs" or "ets" somewhere in their name:

```
#select columns that contain 'avs' in the name
df2 = df.filter(regex='avs|ets')

#view DataFrame
print(df2)

   mavs  cavs  hornets  nets
0    10    18        5    10
1    12    22        7    14
2    14    19        7    25
3    15    14        9    22
4    19    14       12    25
5    22    11        9    17
6    27    20       14    12
```

Only the columns that contain "avs" or "ets" in the name are returned.

Note that the vertical bar (|) is the "OR" operator in pandas.

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