

How can I select columns in Pandas based on a partial match?

Authored by
stats writer

June 25, 2024

RECOMMENDED CITATION

stats writer (2024). *How can I select columns in Pandas based on a partial match?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=152552>

Pandas is a popular library in Python used for data manipulation and analysis. One common task in data analysis is selecting columns based on a partial match. This can be achieved in Pandas using the "filter" function, which allows the user to specify a partial match string and returns columns that contain that string. This feature is useful for selecting columns with similar names or columns that follow a certain naming convention. It simplifies the process of selecting specific columns and allows for efficient data analysis. Overall, the "filter" function in Pandas provides a convenient and powerful way to select columns based on a partial match, making it a valuable tool for data analysts and scientists.

Pandas: Select Columns Based on Partial Match

You can use the following methods to select columns in a pandas DataFrame based on partial matching:

Method 1: Select Columns Based on One Partial Match

```
#select columns that contain 'team'df.loc
```

Method 2: Select Columns Based on Multiple Partial Matches

```
#select columns that contain 'team' or 'rebounds'  
df.loc
```

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

```
import pandas as pd
```

#create DataFrame

```
df = pd.DataFrame({'team_name': ,  
'team_points': ,  
'assists': ,  
'rebounds': })
```

#view DataFrame

```
print(df)
```

```
team_name team_points assists rebounds  
0 A 5 11 6  
1 A 7 8 7  
2 A 7 10 7  
3 A 9 6 6  
4 B 12 6 10  
5 B 9 5 12  
6 B 9 9 10  
7 B 4 12 9
```

Example 1: Select Columns Based on One Partial Match

The following code shows how to select all columns in the pandas DataFrame that contain 'team' in the column name:

```
#select columns that contain 'team'df_team_cols =  
df.loc
```

```
#view results
```

```
print(df_team_cols)
```

```
team_name team_points
```

```
0 A 5
```

```
1 A 7
```

```
2 A 7
```

```
3 A 9
```

```
4 B 12
```

```
5 B 9
```

```
6 B 9
```

```
7 B 4
```

Notice that both columns that contain 'team' in the name are returned.

Example 2: Select Columns Based on Multiple Partial Matches

The following code shows how to select all columns in the pandas DataFrame that contain 'team' or 'rebounds' in the column name:

```
#select columns that contain 'team' or
```

```
'rebounds'df_team_rebs = df.loc
```

```
#view results
```

```
print(df_team_rebs)
```

```
team_name team_points rebounds
```

```
0 A 5 6
```

```
1 A 7 7
```

```
2 A 7 7
```

```
3 A 9 6
```

```
4 B 12 10
```

```
5 B 9 12
```

```
6 B 9 10
```

```
7 B 4 9
```

All columns that contain either 'team' or 'rebounds' in the name are returned.

Note: The | operator represents "OR" in pandas.

Feel free to use as many of these operators as you'd like to search for as many partial string matches as you'd like.

The following tutorials explain how to perform other

common operations in pandas:

ARABPSYCHOLOGY.COM