

How can I filter a pandas dataframe for values that do not contain a certain string or character?

Authored by
stats writer

June 26, 2024

RECOMMENDED CITATION

stats writer (2024). *How can I filter a pandas dataframe for values that do not contain a certain string or character?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=153706>

Filtering a pandas dataframe for values that do not contain a certain string or character can be done by using the "str.contains()" method. This method allows us to search for a specific string or character within a column of the dataframe and return a boolean series indicating whether each value contains the specified string or character. By using the "~" operator, we can then invert the boolean series and filter out the values that contain the specified string or character. This allows us to easily exclude certain values from our dataframe and obtain a subset of data that meets our desired criteria.

Pandas: Filter for "Not Contains"

You can use the following methods to perform a "Not Contains" filter in a pandas DataFrame:

Method 1: Filter for Rows that Do Not Contain Specific String

```
filtered_df = df.str.contains('some_string') == False]
```

Method 2: Filter for Rows that Do Not Contain One of Several Specific Strings

```
filtered_df = df.str.contains('string1|string2|string3')  
==False]
```

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': ,  
'points': ,  
'assists': ,  
'rebounds': })
```

#view DataFrame

```
print(df)
```

```
team points assists rebounds
```

```
0 Nets 18 5 11
```

```
1 Rockets 22 7 8
```

```
2 Mavs 19 7 10
```

```
3 Spurs 14 9 6
```

```
4 Kings 14 12 6
```

```
5 Nuggets 11 9 5
```

Example 1: Filter for Rows that Do Not Contain Specific String

The following code shows how to filter the pandas DataFrame for rows where the team column does not contain "ets" in the name:

```
#filter for rows that do not contain 'ets' in the 'team'  
column
```

```
filtered_df = df.str.contains('ets') ==False]
```

```
#view filtered DataFrame
```

```
print(filtered_df)
```

```
team points assists rebounds
```

```
2 Mavs 19 7 10
```

```
3 Spurs 14 9 6
```

```
4 Kings 14 12 6
```

Notice that the resulting DataFrame does not contain any rows where the value in the team column contains "ets" in the name.

In particular, the following teams were filtered out of the DataFrame:

```
NetsRocketsNuggets
```

Notice that each of these team names contained "ets" in the name.

Example 2: Filter for Rows that Do Not Contain One of Several Specific Strings

The following code shows how to filter the pandas

DataFrame for rows where the team column does not contain "ets" in the name:

#filter for rows that do not contain 'ets' or 'urs' in the 'team' column

```
filtered_df = df.str.contains('ets|urs') == False]
```

#view filtered DataFrame

```
print(filtered_df)
```

team points assists rebounds

2 Mavs 19 7 10

4 Kings 14 12 6

Note: The | operator stands for "OR" in pandas.

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