

How do I filter rows in a Pandas DataFrame that contain a specific string?

Authored by
stats writer

July 2, 2024

RECOMMENDED CITATION

stats writer (2024). *How do I filter rows in a Pandas DataFrame that contain a specific string?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=165590>

Pandas is a Python library used for data manipulation and analysis. It includes a function called "filter" which allows users to select specific rows in a DataFrame based on a certain criteria. To filter rows containing a specific string, users can use the "str.contains()" method, which searches for the specified string in a specific column of the DataFrame. The "str.contains()" method can be combined with other filtering methods to further refine the selection of rows. This allows users to easily extract and analyze data that contains a specific string in a Pandas DataFrame.

Pandas: Filter Rows that Contain a Specific String

You can use the following syntax to filter for rows that contain a certain string in a pandas DataFrame:

```
df.str.contains("this string")]
```

This tutorial explains several examples of how to use this syntax in practice with the following DataFrame:

```
import pandas as pd  
  
#create DataFrame  
df = pd.DataFrame({'team': ,  
'conference': ,  
'points': })
```

```
#view DataFrame  
df
```

```
team conference points
```

0 A East 11

1 A East 8

2 A East 10

3 B West 6

4 B West 6

5 C East 5

Example 1: Filter Rows that Contain a Specific String

The following code shows how to filter for rows in the DataFrame that contain 'A' in the team column:

```
df[df.team.str.contains("A")]
```

team conference points

0 A East 11

1 A East 8

2 A East 10

Only the rows where the team column contains 'A' are kept.

Example 2: Filter Rows that Contain a String in a List

The following code shows how to filter for rows in the DataFrame that contain 'A' or 'B' in the team column:

```
df.str.contains("A|B")]
```

team conference points

0 A East 11

1 A East 8

2 A East 10

3 B West 6

4 B West 6

Only the rows where the team column contains 'A' or 'B' are kept.

Example 3: Filter Rows that Contain a Partial String

In the previous examples, we filtered based on rows that exactly matched one or more strings.

However, if we'd like to filter for rows that contain a partial string then we can use the following syntax:

```
#identify partial string to look for  
keep=
```

```
#filter for rows that contain the partial string "Wes" in  
the conference column
```

```
df
```

team conference points

3 B West 6

4 B West 6

Only the rows where the conference column contains "Wes" are kept.

Additional Resources

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