Does Pandas have a method to check if a string contains multiple substrings?

Authored by

stats writer

June 26, 2024

Pandas is a popular Python library used for data manipulation and analysis. It offers a wide range of methods to efficiently handle and process large datasets. One common question that arises when working with text data is whether Pandas has a method to check if a string contains multiple substrings. The answer is yes, Pandas does provide a method called `str.contains()` which allows users to check if a string contains one or more specified substrings. This method can be used to filter and manipulate data based on the presence of certain substrings, making it a useful tool for text data analysis.

Pandas: Check if String Contains Multiple Substrings

You can use the following methods to check if a string in a pandas DataFrame contains multiple substrings:

Method 1: Check if String Contains One of Several Substrings

df.str.contains('|'.join())

Method 2: Check if String Contains Several Substrings

df.str.contains(r'^(?=.*string1)(?=.*string2)')

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' : })

#view DataFrame
print(df)

team points
0 Good East Team 93
1 Good West Team 99
2 Great East Team 105
3 Great West Team 110
4 Bad East Team 85
5 Bad West Team 88
```

Example 1: Check if String Contains One of Several Substrings

We can use the following syntax to check if each string in the team column contains either the substring "Good" or "East":

```
#create new column that checks if each team name contains 'Good' or 'East'
df = df.str.contains('|'.join())

#view updated DataFrame
```

```
print(df)

team points good_or_east
0 Good East Team 93 True
1 Good West Team 99 True
2 Great East Team 105 True
3 Great West Team 110 False
4 Bad East Team 85 True
5 Bad West Team 88 False
```

The new good_or_east column returns the following values:

True if team contains "Good" or "East"False if team contains neither "Good" nor "East"

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

Example 2: Check if String Contains Several Substrings

We can use the following syntax to check if each string in the team column contains the substring "Good" and "East":

```
#create new column that checks if each team name contains 'Good' and 'East'
```

```
df = df.str.contains(r'^(?=.*Good)(?=.*East)')

#view updated DataFrame
print(df)

team points good_and_east
0 Good East Team 93 True
1 Good West Team 99 False
2 Great East Team 105 False
3 Great West Team 110 False
4 Bad East Team 85 False
5 Bad West Team 88 False
```

The new good_and_east column returns the following values:

True if team contains "Good" and "East"False if team doesn't contain "Good" and "East"

Notice that only one True value is returned since there is only one team name that contains the substring "Good" and the substring "East."

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