

How can I search for a string in all columns of a Pandas DataFrame?

Authored by
stats writer

June 26, 2024

RECOMMENDED CITATION

stats writer (2024). *How can I search for a string in all columns of a Pandas DataFrame?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=153329>

Pandas DataFrame is a powerful tool for data analysis and manipulation in Python. One common task in data analysis is searching for a specific string within a DataFrame. This can be achieved by using the "str.contains()" method, which allows the user to search for a string in all columns of the DataFrame. This method takes in the string to be searched for as a parameter and returns a boolean value indicating whether the string was found in any of the columns. By utilizing this method, users can efficiently search for strings within large datasets and extract relevant information for further analysis.

Pandas: Search for String in All Columns of DataFrame

You can use the following syntax to search for a particular string in each column of a pandas DataFrame and filter for rows that contain the string in at least one column:

```
#define filter
```

```
mask = np.column_stack(.str.contains(r"my_string",  
na=False) for col in df])
```

```
#filter for rows where any column contains 'my_string'  
df.loc
```

The following example shows how to use this syntax in practice.

Example: Search for String in All Columns of Pandas DataFrame

Suppose we have the following pandas DataFrame that contains information about the first role and second

role of various basketball players on a team:

```
import pandas as pd
```

```
#create DataFrame
```

```
df = pd.DataFrame({'player': ,  
'first_role': ,  
'second_role': })
```

```
#view DataFrame
```

```
print(df)
```

```
player first_role second_role
```

```
0 A P Guard S Guard
```

```
1 B P Guard S Guard
```

```
2 C S Guard Forward
```

```
3 D S Forward S Guard
```

```
4 E P Forward S Guard
```

```
5 F Center S Forward
```

```
6 G Center P Forward
```

```
7 H Center P Forward
```

The following code shows how to filter the pandas DataFrame for rows where the string "Guard" occurs in any column:

```
import numpy as np
```

```
#define filter
```

```
mask = np.column_stack([df[col].str.contains(r"Guard",  
na=False) for col in df])
```

```
#filter for rows where any column contains 'Guard'  
df.loc
```

```
player first_role second_role
```

```
0 A P Guard S Guard
```

```
1 B P Guard S Guard
```

```
2 C S Guard Forward
```

```
3 D S Forward S Guard
```

```
4 E P Forward S Guard
```

Notice that each row in the resulting DataFrame contains the string "Guard" in at least one column.

You could also filter for rows where one of several strings occurs in at least one column by using the "OR" (|) operator in pandas.

For example, the following code shows how to filter for rows where either "P Guard" or "Center" occurs in at

least one column:

```
import numpy as np
```

```
#define filter
```

```
mask = np.column_stack(.str.contains(r"P  
Guard|Center", na=False) for col in df])
```

```
#filter for rows where any column contains 'P Guard' or  
'Center'
```

```
df.loc
```

```
player first_role second_role
```

```
0 A P Guard S Guard
```

```
1 B P Guard S Guard
```

```
5 F Center S Forward
```

```
6 G Center P Forward
```

```
7 H Center P Forward
```

Notice that each row in the resulting DataFrame contains "P Guard" or Center" in at least one column.

Note: It's important to include the argument na=False within the contains() function or else you will encounter if NaN values are present in the DataFrame.

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

ARABPSYCHOLOGY.COM