Does Pandas have a method to check if a column exists in a DataFrame?

Authored by

stats writer

July 1, 2024

Pandas is a popular Python library used for data manipulation and analysis. It offers a wide range of functions and methods to efficiently work with tabular data, such as DataFrames. One common task when working with DataFrames is checking the existence of a specific column. To address this, Pandas provides a method called 'columns', which can be used to retrieve a list of all columns in a DataFrame. By comparing this list with the desired column name, users can easily determine whether a column exists in the DataFrame or not. This functionality allows for efficient and accurate data handling, making Pandas a valuable tool for data scientists and analysts.

Check if Column Exists in Pandas (With Examples)

You can use the following methods to check if a column exists in a pandas DataFrame:

Method 1: Check if One Column Exists

'column1' in df.columns

This will return True if 'column1' exists in the DataFrame, otherwise it will return False.

Method 2: Check if Multiple Columns Exist

{'column1', 'column2'}.issubset(df.columns)

This will return True if 'column1' *and* 'column2' exists in the DataFrame, otherwise it will return False.

The following examples shows 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 A 18 5 11
1 B 22 7 8
2 C 19 7 10
3 D 14 9 6
4 E 14 12 6
5 F 11 9 5
6 G 20 9 9
7 H 28 4 12
```

Example 1: Check if One Column Exists

We can use the following code to see if the column

'team' exists in the DataFrame:

#check if 'team' column exists in DataFrame
'team' in df.columns

True

The column 'team' does exist in the DataFrame, so pandas returns a value of True.

We can also use an if statement to perform some operation if the column 'team' exists:

#if 'team' exists, create new column called 'team_name'
if 'team' in df.columns:
df = df

#view updated DataFrame
print(df)

team points assists rebounds team_name
0 A 18 5 11 A
1 B 22 7 8 B
2 C 19 7 10 C
3 D 14 9 6 D
4 E 14 12 6 E

5 F 11 9 5 F

6 G 20 9 9 G

7 H 28 4 12 H

Example 2: Check if Multiple Columns Exist

We can use the following code to see if the columns 'team' *and* 'player' exist in the DataFrame:

```
#check if 'team' and 'player' columns both exist in DataFrame
{'team', 'player'}.issubset(df.columns)
```

False

The column 'team' exists in the DataFrame but 'player' does not, so pandas returns a value of False.

We could also use the following code to see if both 'points' and 'assists' exist in the DataFrame:

```
#check if 'points' and 'assists' columns both exist in DataFrame
{'points', 'assists'}.issubset(df.columns)
```

True

Both columns exist, so pandas returns a value of True.

We can then use an if statement to perform some operation if 'points' and 'assists' both exist:

```
#if both exist, create new column called 'total' that finds sum of points and assists
if {'points', 'assists'}.issubset(df.columns):
df = df + df

#view updated DataFrame
print(df)
```

```
team points assists rebounds total
0 A 18 5 11 23
1 B 22 7 8 29
2 C 19 7 10 26
3 D 14 9 6 23
4 E 14 12 6 26
5 F 11 9 5 20
6 G 20 9 9 29
7 H 28 4 12 32
```

Since 'points' and 'assists' both exist in the DataFrame, pandas went ahead and created a new column called

'total' that shows the sum of the 'points' and 'assists' columns.

Additional Resources

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