

How can I create a column in Pandas only if it does not already exist?

Authored by
stats writer

June 27, 2024

RECOMMENDED CITATION

stats writer (2024). *How can I create a column in Pandas only if it does not already exist?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=154662>

Creating a column in Pandas only if it does not already exist involves using the "if not in" statement to check if the column already exists in the dataframe. If the column is not present, then a new column can be created using the "df = " method. This approach ensures that the column is only created if it does not already exist in the dataframe, avoiding any potential errors or duplicates.

Pandas: Create Column If It Doesn't Exist

You can use the following basic syntax to create a column in a pandas DataFrame if it doesn't already exist:

```
df = df.get('my_column', df * df)
```

This particular syntax creates a new column called `my_column` if it doesn't already exist in the DataFrame and it is defined as the product of the existing columns `col1` and `col2`.

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

Example: Create Column in Pandas If It Doesn't Exist

Suppose we have the following pandas DataFrame:

```
import pandas as pd
```

```
#create DataFrame
```

```
df = pd.DataFrame({'day': ,  
'sales': ,  
'price': })
```

```
#view DataFrame  
print(df)
```

```
day sales price
```

```
0 1 4 1  
1 2 6 2  
2 3 5 2  
3 4 8 1  
4 5 14 2  
5 6 13 4  
6 7 13 4  
7 8 12 3  
8 9 9 3  
9 10 8 2  
10 11 19 2  
11 12 14 3
```

Now suppose we attempt to add a column called price if it doesn't already exist and define it as a column in which each value is equal to 100:

```
#attempt to add column called 'price'
```

```
df = df.get('price', 100)
```

```
#view updated DataFrame
```

```
print(df)
```

```
day sales price
```

```
0 1 4 1
```

```
1 2 6 2
```

```
2 3 5 2
```

```
3 4 8 1
```

```
4 5 14 2
```

```
5 6 13 4
```

```
6 7 13 4
```

```
7 8 12 3
```

```
8 9 9 3
```

```
9 10 8 2
```

```
10 11 19 2
```

```
11 12 14 3
```

Since a column called price already exists, pandas simply doesn't add it to the DataFrame.

However, suppose we attempt to add a new column called revenue if it doesn't already exist and define it as

a column in which the values are the product of the sales and price columns:

```
#attempt to add column called 'revenue'
```

```
df = df.get('revenue', df * df)
```

```
#view updated DataFrame
```

```
print(df)
```

```
day sales price revenue
```

```
0 1 4 1 4
```

```
1 2 6 2 12
```

```
2 3 5 2 10
```

```
3 4 8 1 8
```

```
4 5 14 2 28
```

```
5 6 13 4 52
```

```
6 7 13 4 52
```

```
7 8 12 3 36
```

```
8 9 9 3 27
```

```
9 10 8 2 16
```

```
10 11 19 2 38
```

```
11 12 14 3 42
```

This revenue column is added to the DataFrame because it did not already exist.

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

ARABPSYCHOLOGY.COM