

How do I create a duplicate column in a Pandas DataFrame?

Authored by
stats writer

June 27, 2024

RECOMMENDED CITATION

stats writer (2024). *How do I create a duplicate column in a Pandas DataFrame?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=154585>

Creating a duplicate column in a Pandas DataFrame is a simple process that involves using the built-in function "assign" to create a new column with the same values as an existing column. This can be achieved by specifying the name of the new column and the name of the existing column as parameters in the "assign" function. The new column will be added to the DataFrame, creating a duplicate column with identical values. This can be useful for making changes to the data in one column without affecting the original column, or for performing calculations on the same data in two separate columns. Overall, using the "assign" function allows for easy creation of duplicate columns in a Pandas DataFrame.

Create a Duplicate Column in Pandas DataFrame

You can use the following basic syntax to create a duplicate column in a pandas DataFrame:

```
df = df.loc
```

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

Example: Create Duplicate Column in Pandas DataFrame

Suppose we have the following pandas DataFrame:

```
import pandas as pd

#create DataFrame
df = pd.DataFrame({'points': ,
'assists': ,
'rebounds': })
```

```
#view DataFrame
```

```
print(df)
```

```
points assists rebounds
```

```
0 25 5 11
```

```
1 12 7 8
```

```
2 15 7 10
```

```
3 14 9 6
```

```
4 19 12 6
```

```
5 23 9 5
```

```
6 25 9 9
```

```
7 29 4 12
```

```
8 32 5 8
```

We can use the following code to create a duplicate of the points column and name it points_duplicate:

```
#create duplicate points column
```

```
df = df.loc
```

```
#view updated DataFrame
```

```
print(df)
```

```
points assists rebounds points_duplicate
```

```
0 25 5 11 25
```

```
1 12 7 8 12
2 15 7 10 15
3 14 9 6 14
4 19 12 6 19
5 23 9 5 23
6 25 9 9 25
7 29 4 12 29
8 32 5 8 32
```

Notice that the `points_duplicate` column contains the exact same values as the `points` column.

Note that the duplicate column must have a different column name than the original column, otherwise a duplicate column will not be created.

For example, if we attempt to use the following code to create a duplicate column, it won't work:

```
#attempt to create duplicate points column
df = df.loc

#view updated DataFrame
print(df)
```

points assists rebounds

0 25 5 11

1 12 7 8

2 15 7 10

3 14 9 6

4 19 12 6

5 23 9 5

6 25 9 9

7 29 4 12

8 32 5 8

No duplicate column was created.

The duplicate column must have a different column name than the original column.

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