

# How can I count duplicates in Pandas?

Authored by  
**stats writer**

June 27, 2024

## RECOMMENDED CITATION

stats writer (2024). *How can I count duplicates in Pandas?*. PSYCHOLOGICAL SCALES.  
Retrieved from <https://scales.arabpsychology.com/?p=155552>

Counting duplicates in Pandas is a process of identifying and quantifying the number of duplicate values present in a dataset using the Pandas library in Python. This can be achieved by applying various methods such as the "duplicated()" function, which flags duplicate values as True and unique values as False, and the "value\_counts()" function, which returns a count of each unique value in a specified column. These methods provide a convenient and efficient way to analyze and handle duplicate values in a Pandas dataframe.

## Count Duplicates in Pandas (With Examples)

You can use the following methods to count duplicates in a pandas DataFrame:

### Method 1: Count Duplicate Values in One Column

```
len(df)-len(df.drop_duplicates())
```

### Method 2: Count Duplicate Rows

```
len(df)-len(df.drop_duplicates())
```

### Method 3: Count Duplicates for Each Unique Row

```
df.groupby(df.columns.tolist(), as_index=False).size()
```

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': ,  
'position': ,  
'points': })
```

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

```
team position points
```

```
0 A G 5
```

```
1 A G 5
```

```
2 A G 8
```

```
3 A F 10
```

```
4 B G 5
```

```
5 B G 7
```

```
6 B F 10
```

```
7 B F 10
```

**Example 1: Count Duplicate Values in One Column**

**The following code shows how to count the number of duplicate values in the points column:**

```
#count duplicate values in points column  
len(df)-len(df.drop_duplicates())
```

**4**

**We can see that there are 4 duplicate values in the points column.**

### **Example 2: Count Duplicate Rows**

**The following code shows how to count the number of duplicate rows in the DataFrame:**

```
#count number of duplicate rows  
len(df)-len(df.drop_duplicates())
```

**2**

**We can see that there are 2 duplicate rows in the DataFrame.**

```
#display duplicated rowsdf
```

```
team position points
```

```
1 A G 5
```

```
7 B F 10
```

### **Example 3: Count Duplicates for Each Unique Row**

The following code shows how to count the number of duplicates for each unique row in the DataFrame:

```
#display number of duplicates for each unique row  
df.groupby(df.columns.tolist(), as_index=False).size()
```

```
team position points size
```

```
0 A F 10 1
```

```
1 A G 5 2
```

```
2 A G 8 1
```

```
3 B F 10 2
```

```
4 B G 5 1
```

```
5 B G 7 1
```

The size column displays the number of duplicates for each unique row.

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