

How can I use GroupBy and value counts in Pandas to analyze a dataset?

Authored by
stats writer

July 1, 2024

RECOMMENDED CITATION

stats writer (2024). *How can I use GroupBy and value counts in Pandas to analyze a dataset?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=165329>

GroupBy and value counts are useful functions in the Pandas library that can be used to analyze datasets. GroupBy allows you to group data based on a specific column or variable, while value counts calculates the frequency of unique values within that group. By combining these two functions, you can gain insights into the distribution of your data and identify patterns or trends within different groups. This can be particularly helpful in identifying relationships between variables or identifying outliers. Overall, using GroupBy and value counts in Pandas can provide valuable information for data analysis and decision making.

Pandas: Use GroupBy and Value Counts

You can use the following basic syntax to count the frequency of unique values by group in a pandas DataFrame:

```
df.groupby().size().unstack(fill_value=0)
```

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

Example: Use GroupBy and Value Counts in Pandas

Suppose we have 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 8
```

```
1 A G 8
```

```
2 A F 10
```

```
3 A F 10
```

```
4 A C 11
```

```
5 B G 8
```

```
6 B F 9
```

```
7 B F 10
```

```
8 B F 10
```

```
9 B F 10
```

We can use the following syntax to count the frequency of the points values, grouped by the team and position columns:

```
#count frequency of points values, grouped by team  
and position
```

```
df.groupby().size().unstack(fill_value=0)
```

points 8 9 10 11

team position

A C 0 0 0 1

F 0 0 2 0

G 2 0 0 0

B F 0 1 3 0

G 1 0 0 0

Here's how to interpret the output:

The value 8 occurred in the points column 0 times for players on team A and position C. The value 9 occurred in the points column 0 times for players on team A and position C. The value 10 occurred in the points column 0 times for players on team A and position C. The value 11 occurred in the points column 1 time for players on team A and position C.

And so on.

We could also use the following syntax to count the frequency of the positions, grouped by team:

#count frequency of positions, grouped by team

df.groupby().size().unstack(fill_value=0)

position C F G

team

A 1 2 2

B 0 4 1

Here's how to interpret the output:

The value 'C' occurred 1 time on team A. The value 'F' occurred 2 times on team A. The value 'G' occurred 2 times on team A. The value 'C' occurred 0 times on team B. The value 'F' occurred 4 times on team B. The value 'G' occurred 1 time on team B.

And so on.

Additional Resources