

How can I get the first row of each group in a Pandas DataFrame?

Authored by
stats writer

July 1, 2024

RECOMMENDED CITATION

stats writer (2024). *How can I get the first row of each group in a Pandas DataFrame?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=164689>

The process of obtaining the first row of each group in a Pandas DataFrame involves grouping the data by a certain criteria and then selecting the first row from each group. This can be done by using the "groupby" function in Pandas and then applying the "first" method to the grouped data. This will return a new DataFrame with the first row of each group as its index. This method is useful for analyzing and summarizing data by specific groups within a larger dataset.

Pandas: Get First Row of Each Group

You can use the following basic syntax to get the first row of each group in a pandas DataFrame:

```
df.groupby('column_name').nth(0)
```

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

Example: Get First Row of Each Group in Pandas

Suppose we have the following pandas DataFrame:

```
import pandas as pd

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

#view DataFrame
```

```
df
```

```
team points assists
```

```
0 A 18 5
```

```
1 A 22 19
```

```
2 B 19 14
```

```
3 B 14 8
```

```
4 B 14 9
```

```
5 C 11 12
```

```
6 C 20 13
```

```
7 C 29 8
```

We can use the following code to get the first row for each team:

```
#get first row for each team  
df.groupby('team').nth(0)
```

```
points assists
```

```
team
```

```
A 18 5
```

```
B 19 14
```

```
C 11 12
```

We can also specify `as_index=False` to keep the original

index values:

#get first row for each team, keep original index values

```
df.groupby('team', as_index=False).nth(0)
```

team points assists

0 A 18 5

2 B 19 14

5 C 11 12

Also note that you can pass a list of values to the nth() function if you'd like to get the first n rows for each group.

For example, the following code shows how to get the first two rows for each group:

#get first two rows for each team, keep original index values

```
df.groupby('team', as_index=False).nth((0, 1))
```

team points assists

0 A 18 5

1 A 22 19

2 B 19 14

3 B 14 8

5 C 11 12

6 C 20 13

Note: You can find the complete documentation for the `nth()` function .

Additional Resources

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

ARABPSYCHOLOGY.COM