

How can I combine rows in Pandas with the same column values?

Authored by
stats writer

June 27, 2024

RECOMMENDED CITATION

stats writer (2024). *How can I combine rows in Pandas with the same column values?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=154637>

Pandas is a popular data analysis library in Python, widely used for manipulating and organizing large datasets. One common task in data analysis is combining rows with the same column values, which can be easily achieved using the `groupby` function in Pandas. This function groups rows based on their shared column values and allows for aggregation of data, such as summing or averaging values in the same group. By combining rows with the same column values, analysts can efficiently explore and analyze their datasets, gaining insights and making informed decisions. Overall, Pandas provides a simple and effective solution for combining rows with the same column values, making it a valuable tool for data analysis.

Pandas: Combine Rows with Same Column Values

You can use the following basic syntax to combine rows with the same column values in a pandas DataFrame:

```
#define how to aggregate various fields
```

```
agg_functions = {'field1': 'first', 'field2': 'sum', 'field':  
'sum'}
```

```
#create new DataFrame by combining rows with same  
id values
```

```
df_new = df.groupby(df).aggregate(agg_functions)
```

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

Example: Combine Rows with Same Column Values in Pandas

Suppose we have the following pandas DataFrame that

contains information about sales and returns made by various employees at a company:

```
import pandas as pd
```

```
#create dataframe
```

```
df = pd.DataFrame({'id': ,  
'employee': ,  
'sales': ,  
'returns': })
```

```
#view DataFrame
```

```
print(df)
```

```
id employee sales returns
```

```
0 101 Dan 4 1
```

```
1 101 Dan 1 2
```

```
2 102 Rick 3 2
```

```
3 103 Ken 2 1
```

```
4 103 Ken 5 3
```

```
5 103 Ken 3 2
```

We can use the following syntax to combine rows that have the same value in the id column and then aggregate the remaining columns:

```
#define how to aggregate various fields
```

```
agg_functions = {'employee': 'first', 'sales': 'sum',  
'returns': 'sum'}
```

```
#create new DataFrame by combining rows with same  
id values
```

```
df_new = df.groupby(df).aggregate(agg_functions)
```

```
#view new DataFrame
```

```
print(df_new)
```

```
employee sales returns
```

```
id
```

```
101 Dan 5 3
```

```
102 Rick 3 2
```

```
103 Ken 10 6
```

The new DataFrame combined all of the rows in the previous DataFrame that had the same value in the id column and then calculated the sum of the values in the sales and returns columns.

Note: Refer to the for a complete list of aggregations available to use with the GroupBy() function.

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

ARABPSYCHOLOGY.COM