How can we group data by month in a Pandas DataFrame?

Authored by

stats writer

June 29, 2024

Pandas is a popular data analysis library in Python that offers powerful tools for manipulating and organizing data. One essential task in data analysis is grouping data by a specific time frame, such as by month. This can be easily achieved in Pandas by using the groupby() function, which allows us to group data by a specific variable, in this case, the month. By specifying the month column as the grouping variable, we can then perform various operations on the data, such as calculating the mean or sum of each month's values. This allows for efficient and comprehensive analysis of data in a monthly format, making it easier to identify trends and patterns over time.

Group by Month in Pandas DataFrame (With Example)

You can use the following basic syntax to group rows by month in a pandas DataFrame:

df.groupby(df.your_date_column.dt.month).sum()

This particular formula groups the rows by date in your_date_column and calculates the sum of values for the values_column in the DataFrame.

Note that the dt.month() function extracts the month from a date column in pandas.

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

Example: How to Group by Month in Pandas

Suppose we have the following pandas DataFrame that shows the sales made by some company on various

dates:

```
import pandas as pd

#create DataFrame
df = pd.DataFrame({'date':
pd.date_range(start='1/1/2020', freq='W', periods=10),
'sales': ,
'returns': })

#view DataFrame
print(df)
```

date sales returns
0 2020-01-05 6 0
1 2020-01-12 8 3
2 2020-01-19 9 2
3 2020-01-26 11 2
4 2020-02-02 13 1
5 2020-02-09 8 3
6 2020-02-16 8 2
7 2020-02-23 15 4
8 2020-03-01 22 1
9 2020-03-08 9 5

Related:

We can use the following syntax to calculate the sum of sales grouped by month:

```
#calculate sum of sales grouped by month
df.groupby(df.date.dt.month).sum()

date
1 34
2 44
3 31
Name: sales, dtype: int64
```

Here's how to interpret the output:

The total sales made during month 1 (January) was 34.The total sales made during month 2 (February) was 44.The total sales made during month 3 (March) was 31.

We can use similar syntax to calculate the max of the sales values grouped by month:

```
#calculate max of sales grouped by month
df.groupby(df.date.dt.month).max()
```

date

1 11

2 15

3 22

Name: sales, dtype: int64

We can use similar syntax to calculate any value we'd like grouped by the month value of a date column.

Note: You can find the complete documentation for the GroupBy operation in pandas .

Additional Resources

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