How can I use groupby() to resample a time series in Pandas?

Authored by

stats writer

June 25, 2024

Groupby() is a function in the Pandas library that allows for grouping and aggregating data based on a specified column or index. By using groupby(), it is possible to resample a time series in Pandas by grouping the data by a specific time period (e.g. daily, weekly, monthly) and performing an operation on the grouped data (e.g. mean, sum, count). This allows for the transformation of a time series into a different time frequency, providing a more concise and organized representation of the data. Resampling with groupby() is a useful tool for analyzing and visualizing time series data in a more manageable and meaningful way.

Pandas: Resample Time Series with groupby()

To resample time series data means to aggregate the data by a new time period.

If you'd like to resample a time series in pandas while using the groupby operator, you can use the following basic syntax:

grouper = df.groupby()

result = grouper.sum().unstack('store').fillna(0)

This particular example groups the rows in the DataFrame by the store column, then resamples the time series by week (freq='W'), then calculates the sum of values in the sales column.

Note that we can resample the time series data by various time periods, including:

S: Secondsmin: MinutesH: HoursD: DayW: WeekM: MonthQ: QuarterA: Year

The following example shows how to resample time series data with a groupby operation in practice.

Example: Resample Time Series with groupby in Pandas

Suppose we have the following pandas DataFrame that shows the total sales made each day at two different stores:

```
import pandas as pd

#create DataFrame
df = pd.DataFrame({'sales': ,
'store': },
index=pd.date_range('2023-01-06',    '2023-01-16',
freq='d'))

#view DataFrame
print(df)

sales store
2023-01-06 13 A
2023-01-07 14 A
```

2023-01-08 17 A

2023-01-09 17 A

2023-01-10 16 A

2023-01-11 22 B

2023-01-12 28 B

2023-01-13 10 B

2023-01-14 17 B

2023-01-15 10 B

2023-01-16 11 B

Suppose we would like to group the rows by store, then resamples the time series by week, then calculates the sum of values in the sales column.

We can use the following syntax to do so:

```
#group by store and resample time series by week
grouper = df.groupby()

#calculate sum of sales each week by store
result = grouper.sum().unstack('store').fillna(0)

#view resultsprint(result)
```

store A B

2023-01-08 14.0 0.0

2023-01-15 16.5 17.0

2023-01-22 0.0 11.0

From the output we cans see:

The sum of sales on the week ending 1/8/2023 at store A is 14.The sum of sales on the week ending 1/8/2023 at store B is 0.

And so on.

Note that in this example we chose to calculate the sum of values in the sales column.

Simply replace sum() in the code above with count(), mean(), median(), etc. to calculate whatever metric you'd like.

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