PANDAS: group by date and other fields

Groupby, resample and pd.Grouper

Your problem...

You have a dataframe with the following fields:
- date
- name
-amount spent
-category of spending

	spending	category_spending	name	date
325	61.504729	0	Barbara	2020-01-01
1609	17.744582	4	Richard	2020-01-01
506	8.552533	9	Michael	2020-01-01
1807	0.543128	0	Susan	2020-01-01
794	38.950612	9	Robert	2020-01-01
...	...	...	...	...
1064	82.815693	3	Susan	2020-12-31
291	74.328212	4	Michael	2020-12-31
857	69.219972	6	Barbara	2020-12-31
37	57.506904	5	Barbara	2020-12-31
900	47.943711	4	Michael	2020-12-31

2000 rows × 4 columns

You want the amount spent per person…

That's easy enough:

```
(test
 .groupby('name')['spending']
 .sum()
 .reset_index()
 )
```

	name	spending
0	Barbara	18763.542549
1	Michael	20542.536837
2	Richard	20293.571894
3	Robert	19657.780899
4	Susan	21083.771545

Of course, you can also group by name and date:

```
(test
 .groupby(['date', 'name'])['spending']
 .sum()
 .reset_index()
 .head(6)
 )
```

	date	name	spending
0	2020-01-01	Barbara	61.504729
1	2020-01-01	Michael	8.552533
2	2020-01-01	Richard	88.218989
3	2020-01-01	Robert	91.914874
4	2020-01-01	Susan	23.026872
5	2020-01-02	Barbara	96.325127

That gives you the amount spent per person, per day…

But what if you want the amount spent per person, per month ?

One solution is to extract the month from the data column:

```
test['month'] = test['date'].dt.month
(test
 .groupby(['month', 'name'])['spending']
 .sum()
 .reset_index()
 .head(6)
 )
```

	month	name	spending
0	1	Barbara	1547.929915
1	1	Michael	1933.735591
2	1	Richard	1380.436786
3	1	Robert	1471.838783
4	1	Susan	1677.703572
5	2	Barbara	1992.982753

But if you want to group by more complex periods, (let's say a couple of days or two weeks), this is not gonna work

There is a Pandas method that helps you “group by period of time”
Here comes df.resample():

```
(test
 .resample("W", on="date")['spending']
 .sum()
 .reset_index().head(6)
 )
```

	date	spending
0	2020-01-05	1243.257215
1	2020-01-12	1932.799026
2	2020-01-19	1755.861310
3	2020-01-26	1461.857024
4	2020-02-02	2325.407612
5	2020-02-09	2116.374003

As you can see, you’ve got one result per week.
You can use “2W”, or “6M”, and it will get you what you expect… How amazing is that ? :)

But how do we use resample and groupby together ?

```
grouper = pd.Grouper(key='date', freq='3M')
(test
 .groupby([grouper, 'name'])['spending']
 .sum()
 .reset_index().head(6)
 )
```

	date	name	spending
0	2020-01-31	Barbara	1547.929915
1	2020-01-31	Michael	1933.735591
2	2020-01-31	Richard	1380.436786
3	2020-01-31	Robert	1471.838783
4	2020-01-31	Susan	1677.703572
5	2020-04-30	Barbara	5603.028158

If you liked this content, please like and share it :)

(if you didn't like it, leave a comment :˚)