04-Groupby

October 19, 2022

---

1 Groupby

The groupby method allows you to group rows of data together and call aggregate functions

[31]:
```
import pandas as pd
# Create dataframe
data = {'Company':['GOOG','GOOG','MSFT','MSFT','FB','FB'],
       'Person':['Sam','Charlie','Amy','Vanessa','Carl','Sarah'],
       'Sales':[200,120,340,124,243,350]}
```

[32]:
```
df = pd.DataFrame(data)
```

[33]:
```
df
```

[33]:

	Company	Person	Sales
0	GOOG	Sam	200
1	GOOG	Charlie	120
2	MSFT	Amy	340
3	MSFT	Vanessa	124
4	FB	Carl	243
5	FB	Sarah	350

Now you can use the .groupby() method to group rows together based off of a column name.For instance let's group based off of Company. This will create a DataFrameGroupBy object:

[34]:
```
df.groupby('Company')
```

[34]:
```
<pandas.core.groupby.DataFrameGroupBy object at 0x113014128>
```

You can save this object as a new variable:

[35]:
```
by_comp = df.groupby("Company")
```

And then call aggregate methods off the object:

[36]:
```
by_comp.mean()
```

[36]:
```
         Sales
Company
FB       296.5
GOOG     160.0
MSFT     232.0
```

[37]:
```
df.groupby('Company').mean()
```

[37]:
```
         Sales
Company
FB       296.5
GOOG     160.0
MSFT     232.0
```

More examples of aggregate methods:

[38]:
```
by_comp.std()
```

[38]:
```
              Sales
Company
FB        75.660426
GOOG      56.568542
MSFT     152.735065
```

[39]:
```
by_comp.min()
```

[39]:
```
          Person  Sales
Company
FB          Carl    243
GOOG     Charlie    120
MSFT         Amy    124
```

[40]:
```
by_comp.max()
```

[40]:
```
          Person  Sales
Company
FB         Sarah    350
GOOG         Sam    200
MSFT     Vanessa    340
```

[41]:
```
by_comp.count()
```

[41]:
```
         Person  Sales
Company
FB            2      2
GOOG          2      2
MSFT          2      2
```

[42]:
```
by_comp.describe()
```

[42]:
```
                  Sales
Company
FB      count    2.000000
        mean   296.500000
        std     75.660426
        min    243.000000
        25%    269.750000
        50%    296.500000
        75%    323.250000
        max    350.000000
GOOG    count    2.000000
        mean   160.000000
        std     56.568542
        min    120.000000
        25%    140.000000
        50%    160.000000
        75%    180.000000
        max    200.000000
MSFT    count    2.000000
        mean   232.000000
        std    152.735065
        min    124.000000
        25%    178.000000
        50%    232.000000
        75%    286.000000
        max    340.000000
```

[43]:
```
by_comp.describe().transpose()
```

[43]:
```
Company    FB                                                        GOOG  \
          count   mean        std    min     25%    50%     75%    max count
Sales       2.0  296.5  75.660426  243.0  269.75  296.5  323.25  350.0   2.0

Company         …                MSFT                                \
           mean  …      75%    max count   mean         std    min    25%
Sales     160.0  …    180.0  200.0   2.0  232.0  152.735065  124.0  178.0

Company
            50%    75%    max
```

```
Sales    232.0  286.0  340.0

[1 rows x 24 columns]
```

[44]:
```
by_comp.describe().transpose()['GOOG']
```

[44]:
```
        count   mean        std    min    25%    50%    75%    max
Sales     2.0  160.0  56.568542  120.0  140.0  160.0  180.0  200.0
```

2 Great Job!