
(http://www.pieriandata.com)

Groupby

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

In [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]}
```

In [32]:

```
df = pd.DataFrame(data)
```

In [33]:

```
df
```

Out[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:

In [34]:

```
df.groupby('Company')
```

Out[34]:

```
<pandas.core.groupby.DataFrameGroupBy object at 0x113014128>
```

You can save this object as a new variable:

In [35]:

```
by_comp = df.groupby("Company")
```

And then call aggregate methods off the object:

In [36]:

```
by_comp.mean()
```

Out[36]:

Company	Sales
FB	296.5
GOOG	160.0
MSFT	232.0

In [37]:

```
df.groupby('Company').mean()
```

Out[37]:

Company	Sales
FB	296.5
GOOG	160.0
MSFT	232.0

More examples of aggregate methods:

In [38]:

```
by_comp.std()
```

Out[38]:

Company	Sales
FB	75.660426
GOOG	56.568542
MSFT	152.735065

In [39]:

```
by_comp.min()
```

Out[39]:

Company	Person	Sales
FB	Carl	243
GOOG	Charlie	120
MSFT	Amy	124

In [40]:

```
by_comp.max()
```

Out[40]:

Company	Person	Sales
FB	Sarah	350
GOOG	Sam	200
MSFT	Vanessa	340

In [41]:

```
by_comp.count()
```

Out[41]:

Company	Person	Sales
FB	2	2
GOOG	2	2
MSFT	2	2

In [42]:

```
by_comp.describe()
```

Out[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

In [43]:

```
by_comp.describe().transpose()
```

Out[43]:

Company	FB										
	count	mean	std	min	25%	50%	75%	max	count	mean	...	7
Sales	2.0	296.5	75.660426	243.0	269.75	296.5	323.25	350.0	2.0	160.0	...	18

1 rows × 24 columns

In [44]:

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

Out[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

Great Job!