Module 02

Groupby

Data Science Developer

Purwadhika
Startup and Coding School

Create DataFrame

In [1]:
```
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 [2]:
```
df = pd.DataFrame(data)
```

In [3]:
```
df
```

Out[3]:

	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

Purwadhika
Startup and Coding School

Groupby() method

mean()

```
In [4]: df.groupby('Company')
Out[4]: <pandas.core.groupby.groupby.DataFrameGroupBy object at 0x000002654B5F12E8>
```

You can save this object as a new variable:

```
In [5]: by_comp = df.groupby("Company")
```

And then call aggregate methods off the object:

```
In [6]: by_comp.mean()
```

Out[6]:

Company	Sales
FB	296.5
GOOG	160.0
MSFT	232.0

```
In [7]: df.groupby('Company').mean()
```

Out[7]:

Company	Sales
FB	296.5
GOOG	160.0
MSFT	232.0

Purwadhika
Startup and Coding School

More Aggregate Methods

```
In [8]: by_comp.std()
```

Out[8]:

Company	Sales
FB	75.660426
GOOG	56.568542
MSFT	152.735065

```
In [9]: by_comp.min()
```

Out[9]:

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

```
In [10]: by_comp.max()
```

Out[10]:

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

```
In [11]: by_comp.count()
```

Out[11]:

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

Purwadhika
Startup and Coding School

More Aggregate Methods

In [12]:
```
by_comp.describe()
```

Out[12]:

	Sales							
	count	mean	std	min	25%	50%	75%	max
Company								
FB	2.0	296.5	75.660426	243.0	269.75	296.5	323.25	350.0
GOOG	2.0	160.0	56.568542	120.0	140.00	160.0	180.00	200.0
MSFT	2.0	232.0	152.735065	124.0	178.00	232.0	286.00	340.0

In [13]:
```
by_comp.describe().transpose()
```

Out[13]:

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

More Aggregate Methods

```
In [15]: by_comp.describe().transpose()['GOOG']
```

```
Out[15]: Sales  count      2.000000
         mean     160.000000
         std       56.568542
         min      120.000000
         25%      140.000000
         50%      160.000000
         75%      180.000000
         max      200.000000
         Name: GOOG, dtype: float64
```

```
In [17]: by_comp.describe().transpose()['GOOG'].loc['Sales'].loc['25%']
```

```
Out[17]: 140.0
```

Purwadhika
Startup and Coding School