

In [17]:

#Jose Neto

'''Linear regression script, we have two columns of data: years of experience and salary. I that can predict the salary of a person based in the years of experience. For that, we plot and, after observe a linear behavior, we use a linear regression algorithm, that is trained algorithm'''

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from sklearn.linear_model import LinearRegression
```

In [13]:

```
data = pd.read_csv('linear-regression-dataset.csv')
```

In [14]:

data

Out[14]:

	experience	salary
0	0.5	2500
1	0.0	2250
2	1.0	2750
3	5.0	8000
4	8.0	9000
5	4.0	6900
6	15.0	20000
7	7.0	8500
8	3.0	6000
9	2.0	3500
10	12.0	15000
11	10.0	13000
12	14.0	18000
13	6.0	7500

In [15]:

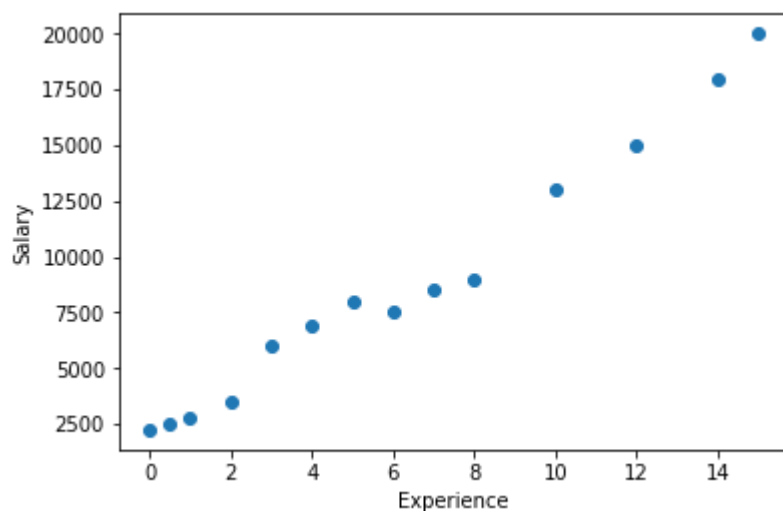
```
data.describe()
```

Out[15]:

	experience	salary
count	14.000000	14.000000
mean	6.250000	8778.571429
std	4.979767	5733.495185
min	0.000000	2250.000000
25%	2.250000	4125.000000
50%	5.500000	7750.000000
75%	9.500000	12000.000000
max	15.000000	20000.000000

In [16]:

```
#Exploratory analisys, is i  
plt.scatter(data.experience, data.salary)  
plt.xlabel("Experience")  
plt.ylabel("Salary")  
plt.show()
```



In [18]:

```
#Seems linear, applying linear regression  
linear_reg = LinearRegression()  
x = data.experience.values.reshape(-1,1)  
y = data.salary.values.reshape(-1,1)  
linear_reg.fit(x,y)
```

Out[18]:

```
LinearRegression(copy_X=True, fit_intercept=True, n_jobs=None,  
                 normalize=False)
```

In [20]:

```
#Ok, the model is trained, applying a prediction  
#Predicting 23 years of experience
```

```
twentythree_salary = linear_reg.predict([[23]])  
twentythree_salary
```

Out[20]:

```
array([[27845.90372791]])
```

In []:

```
#The salary for a 23 years old person is 27845.9
```