

```
In [1]: import warnings
warnings.simplefilter("ignore")
```

```
In [2]: import pandas as pd
import numpy as np
```

```
In [3]: import matplotlib.pyplot as plt
%matplotlib inline
```

```
In [4]: dataset = pd.read_csv('Salary_Data.csv')
```

```
In [5]: dataset
```

```
Out[5]:
```

	YearsExperience	Salary
0	1.1	39343.0
1	1.3	46205.0
2	1.5	37731.0
3	2.0	43525.0
4	2.2	39891.0
5	2.9	56642.0
6	3.0	60150.0
7	3.2	54445.0
8	3.2	64445.0
9	3.7	57189.0
10	3.9	63218.0
11	4.0	55794.0
12	4.0	56957.0
13	4.1	57081.0
14	4.5	61111.0
15	4.9	67938.0
16	5.1	66029.0
17	5.3	83088.0
18	5.9	81363.0
19	6.0	93940.0
20	6.8	91738.0
21	7.1	98273.0
22	7.9	101302.0
23	8.2	113812.0

	YearsExperience	Salary
24	8.7	109431.0
25	9.0	105582.0
26	9.5	116969.0
27	9.6	112635.0
28	10.3	122391.0
29	10.5	121872.0

In [6]: `dataset.shape`

Out[6]: (30, 2)

In [7]: `dataset.head()`

Out[7]:

	YearsExperience	Salary
0	1.1	39343.0
1	1.3	46205.0
2	1.5	37731.0
3	2.0	43525.0
4	2.2	39891.0

In [8]: `x = dataset.iloc[:,0]`

In [9]: `x`

Out[9]:

0	1.1
1	1.3
2	1.5
3	2.0
4	2.2
5	2.9
6	3.0
7	3.2
8	3.2
9	3.7
10	3.9
11	4.0
12	4.0
13	4.1
14	4.5
15	4.9
16	5.1
17	5.3
18	5.9
19	6.0
20	6.8
21	7.1
22	7.9
23	8.2
24	8.7

```
25     9.0
26     9.5
27     9.6
28    10.3
29    10.5
Name: YearsExperience, dtype: float64
```

```
In [10]: x.shape
```

```
Out[10]: (30,)
```

```
In [11]: x = dataset.iloc[:,0].values.reshape(-1,1)
```

```
In [12]: x.shape
```

```
Out[12]: (30, 1)
```

```
In [13]: y = dataset.iloc[:, -1].values.reshape(-1,1)
```

```
In [14]: y.shape
```

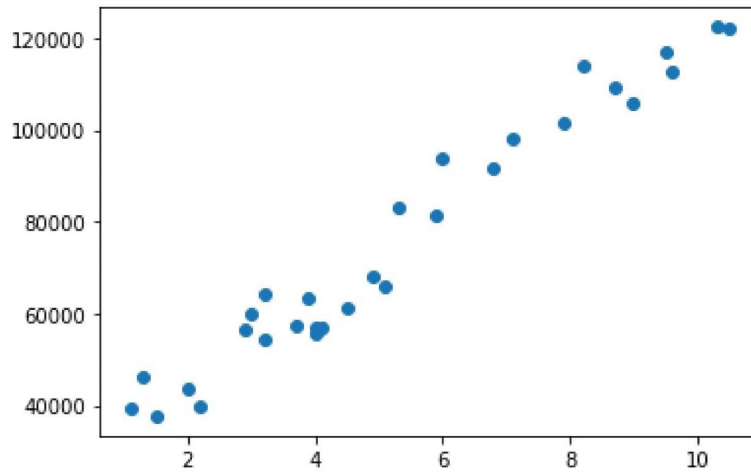
```
Out[14]: (30, 1)
```

```
In [15]: y
```

```
Out[15]: array([[ 39343.],
 [ 46205.],
 [ 37731.],
 [ 43525.],
 [ 39891.],
 [ 56642.],
 [ 60150.],
 [ 54445.],
 [ 64445.],
 [ 57189.],
 [ 63218.],
 [ 55794.],
 [ 56957.],
 [ 57081.],
 [ 61111.],
 [ 67938.],
 [ 66029.],
 [ 83088.],
 [ 81363.],
 [ 93940.],
 [ 91738.],
 [ 98273.],
 [101302.],
 [113812.],
 [109431.],
 [105582.],
 [116969.],
 [112635.],
 [122391.],
 [121872.]])
```

```
In [16]: plt.scatter(x,y)
plt.show
```

```
Out[16]: <function matplotlib.pyplot.show(close=None, block=None)>
```



```
In [17]: from sklearn.model_selection import train_test_split
```

```
In [18]: x_train,x_test,y_train,y_test = train_test_split(x,y,test_size = 0.2,random_state = 0)
```

```
In [19]: x_train.shape
```

```
Out[19]: (24, 1)
```

```
In [20]: x_test.shape
```

```
Out[20]: (6, 1)
```

```
In [21]: y_train.shape
```

```
Out[21]: (24, 1)
```

```
In [22]: y_test.shape
```

```
Out[22]: (6, 1)
```

```
In [23]: from sklearn.linear_model import LinearRegression
```

```
In [24]: lm = LinearRegression()
```

```
In [25]: lm.fit(x_train,y_train)
```

```
Out[25]: LinearRegression()
```

```
In [26]: y_pred = lm.predict(x_test)
```

```
In [27]: y_pred
```

```
Out[27]: array([[ 40748.96184072],
 [122699.62295594],
 [ 64961.65717022],
 [ 63099.14214487],
 [115249.56285456],
 [107799.50275317]])
```

```
In [28]: plt.scatter(x,y,color='blue')  
plt.plot(x_test,y_pred,color='red')
```

Out[28]: [

