In [21]:

```
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np
```

In [7]:

```
df = pd.read_csv('C:/Users/User/Downloads/Salary_Data.csv')
df
```

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
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
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 [9]:

```
df.shape
```

Out[9]:

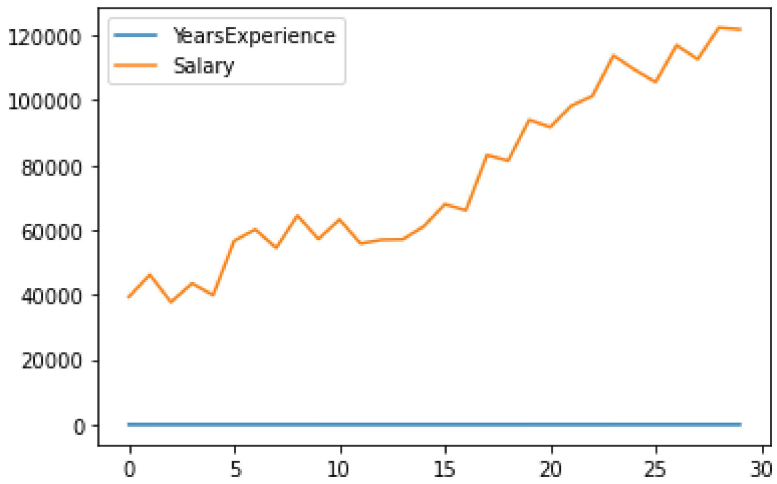
```
(30, 2)
```

In [10]:

```
df.plot()
```

Out[10]:

```
<AxesSubplot:>
```



In [11]:

```
df.corr()
```

Out[11]:

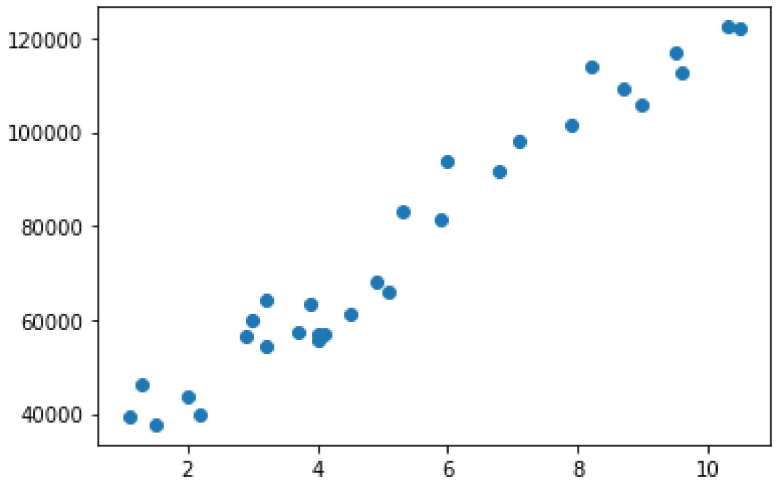
	YearsExperience	Salary
YearsExperience	1.000000	0.978242
Salary	0.978242	1.000000

In [14]:

```
plt.scatter(x=df.YearsExperience,y=df.Salary)
```

Out[14]:

```
<matplotlib.collections.PathCollection at 0x217e9770940>
```



In [16]:

```
from sklearn.linear_model import LinearRegression
lr = LinearRegression()
```

In [22]:

```
x = np.array(df['YearsExperience']).reshape(-1,1)
y = np.array(df['Salary']).reshape(-1,1)
y
```

Out[22]:

```
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 [27]:

```
from sklearn.model_selection import train_test_split
train_x,test_x,train_y,test_y = train_test_split(x,y,random_state = 0)
```

In [24]:

```
lr.fit(train_x,train_y)
```

Out[24]:

```
LinearRegression()
```

In [28]:

```
y_predict = lr.predict(test_x)
y_predict
```

Out[28]:

```
array([[ 41056.25705466],
       [123597.70938378],
       [ 65443.50433372],
       [ 63567.56223533],
       [116093.94099022],
       [108590.17259667],
       [117031.91203942],
       [ 64505.53328452]])
```

In [30]:

```
from sklearn.metrics import r2_score
lracc = r2_score(test_y,y_predict)
lracc*100
```

Out[30]:

```
97.79208335417601
```