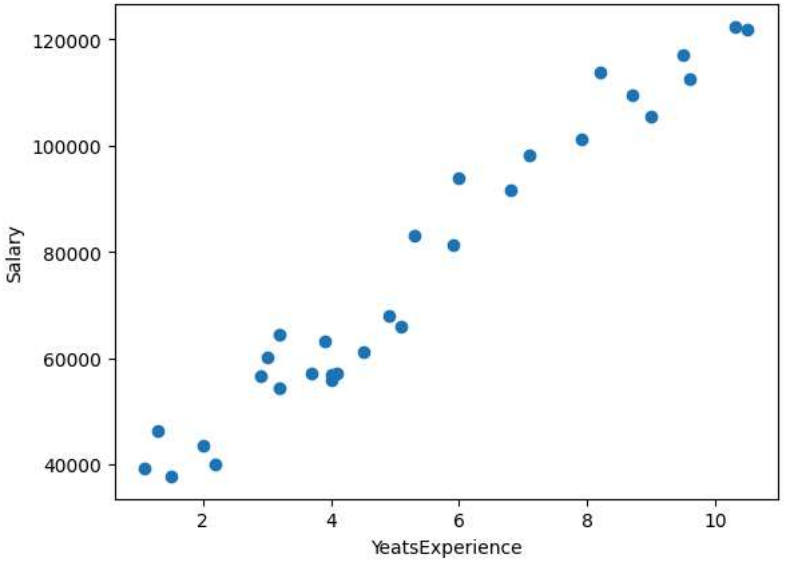
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
```

```
df=pd.read_csv("/content/salary_data.csv")
df.describe()
```

	YearsExperience	Salary
count	30.000000	30.000000
mean	5.313333	76003.000000
std	2.837888	27414.429785
min	1.100000	37731.000000
25%	3.200000	56720.750000
50%	4.700000	65237.000000
75%	7.700000	100544.750000
max	10.500000	122391.000000

```
plt.scatter(df['YearsExperience'],df['Salary'])
plt.xlabel('YeatsExperience')
plt.ylabel('Salary')
```

```
Text(0, 0.5, 'Salary')
```




```
x=df.iloc[:,0:1]
y=df.iloc[:,-1]
y
```

	Salary
0	39343.0
1	46205.0
2	37731.0
3	43525.0
4	39891.0
5	56642.0
6	60150.0
7	54445.0
8	64445.0
9	57189.0
10	63218.0
11	55794.0
12	56957.0
13	57081.0
14	61111.0
15	67938.0
16	66029.0
17	83088.0
18	81363.0
19	93940.0
20	91738.0
21	98273.0
22	101302.0
23	113812.0
24	109431.0
25	105582.0
26	116969.0
27	112635.0
28	122391.0
29	121872.0

dtype: float64

```
from sklearn.model_selection import train_test_split
x_train,x_test,y_train,y_test = train_test_split(x,y, test_size=0.2, random_state=2)


from sklearn.linear_model import LinearRegression
lr=LinearRegression()
lr.fit(x_train,y_train)
```

x_test

	YearsExperience
1	1.3
0	1.1
14	4.5
9	3.7
21	7.1
19	6.0

y_test

	Salary
1	46205.0
0	39343.0
14	61111.0
9	57189.0
21	98273.0
19	93940.0

dtype: float64

x_train

	YearsExperience
23	8.2
6	3.0
3	2.0
20	6.8
5	2.9
27	9.6
12	4.0
4	2.2
10	3.9
16	5.1
28	10.3
25	9.0
17	5.3
2	1.5
7	3.2
26	9.5
24	8.7
18	5.9
11	4.0
22	7.9
29	10.5
13	4.1
15	4.9
8	3.2

```
y_train
```

	Salary
23	113812.0
6	60150.0
3	43525.0
20	91738.0
5	56642.0
27	112635.0
12	56957.0
4	39891.0
10	63218.0
16	66029.0
28	122391.0
25	105582.0
17	83088.0
2	37731.0
7	54445.0
26	116969.0
24	109431.0
18	81363.0
11	55794.0
22	101302.0
29	121872.0
13	57081.0
15	67938.0
8	64445.0

dtype: float64

```
plt.scatter(df['YearsExperience'],df['Salary'])
plt.plot(x_train,lr.predict(x_train),color='red')
plt.xlabel('YearsExperience')
plt.ylabel('Salary')
```

Text(0, 0.5, 'Salary')