In [1]:
```python
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
```

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

In [5]:
```python
df.head()
```

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

In [6]:
```python
df.tail()
```

Out[6]:

	YearsExperience	Salary
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 [7]:
```python
df.info()
```

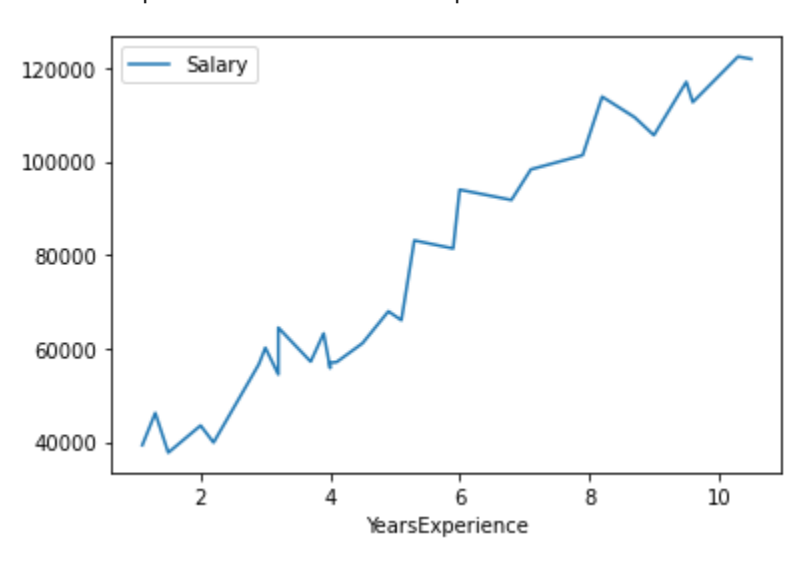
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 30 entries, 0 to 29
Data columns (total 2 columns):
 #   Column           Non-Null Count  Dtype  
---  ------           --------------  -----  
 0   YearsExperience  30 non-null     float64
 1   Salary           30 non-null     float64
dtypes: float64(2)
memory usage: 608.0 bytes
```

In [9]:
```python
df.isnull().sum()
```

Out[9]:
```
YearsExperience    0
Salary             0
dtype: int64
```

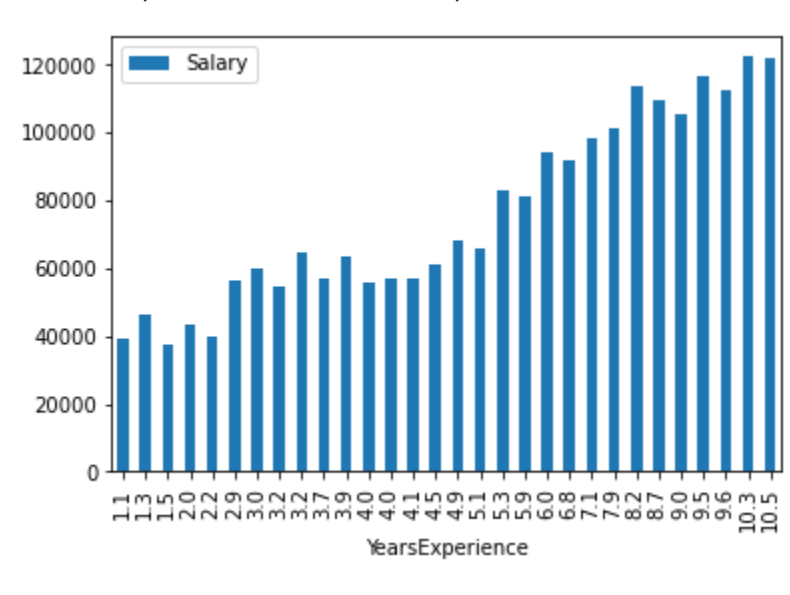
In [13]:
```python
df.plot(x='YearsExperience',y='Salary')
```

Out[13]: <AxesSubplot:xlabel='YearsExperience'>



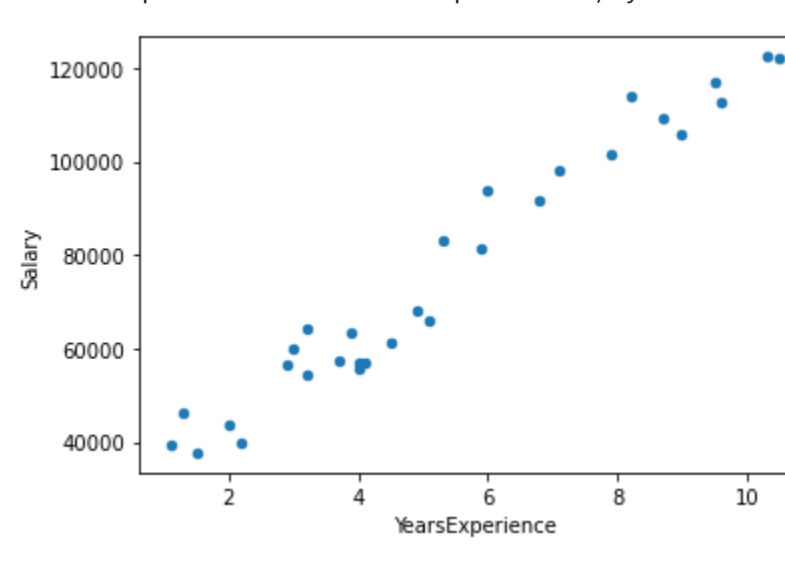
In [14]:
```python
df.plot(kind='bar',x='YearsExperience',y='Salary')
```

Out[14]: <AxesSubplot:xlabel='YearsExperience'>



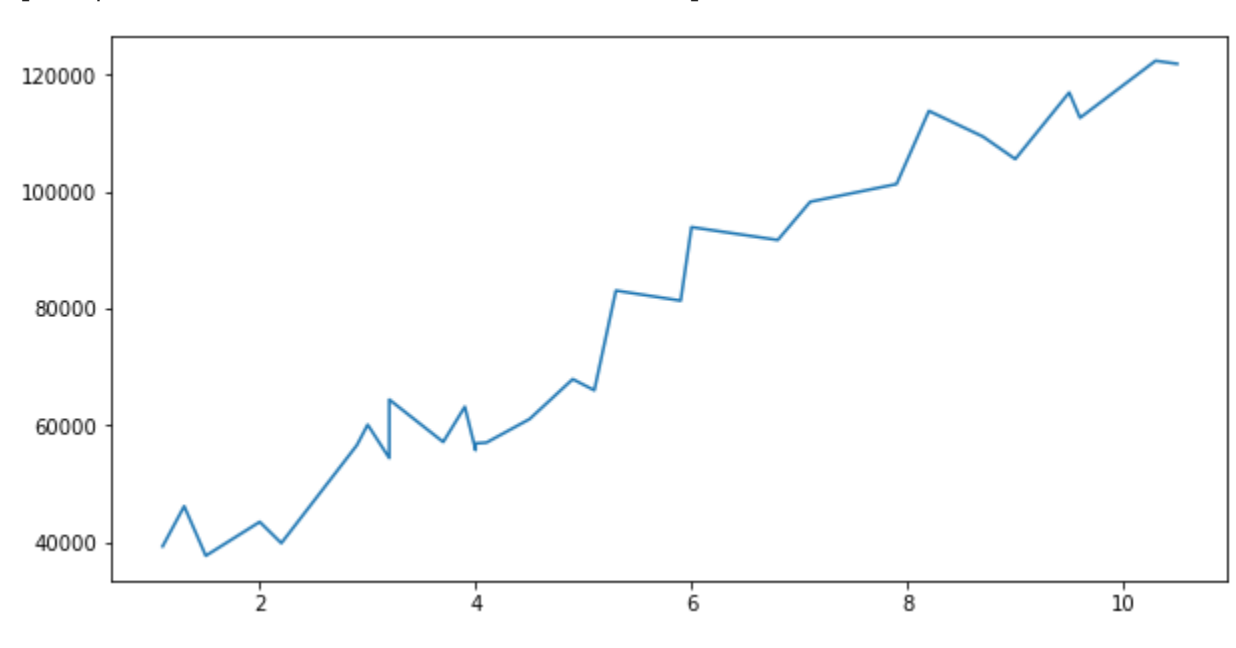
In [15]:
```python
df.plot(kind='scatter',x='YearsExperience',y='Salary')
```

Out[15]: <AxesSubplot:xlabel='YearsExperience', ylabel='Salary'>



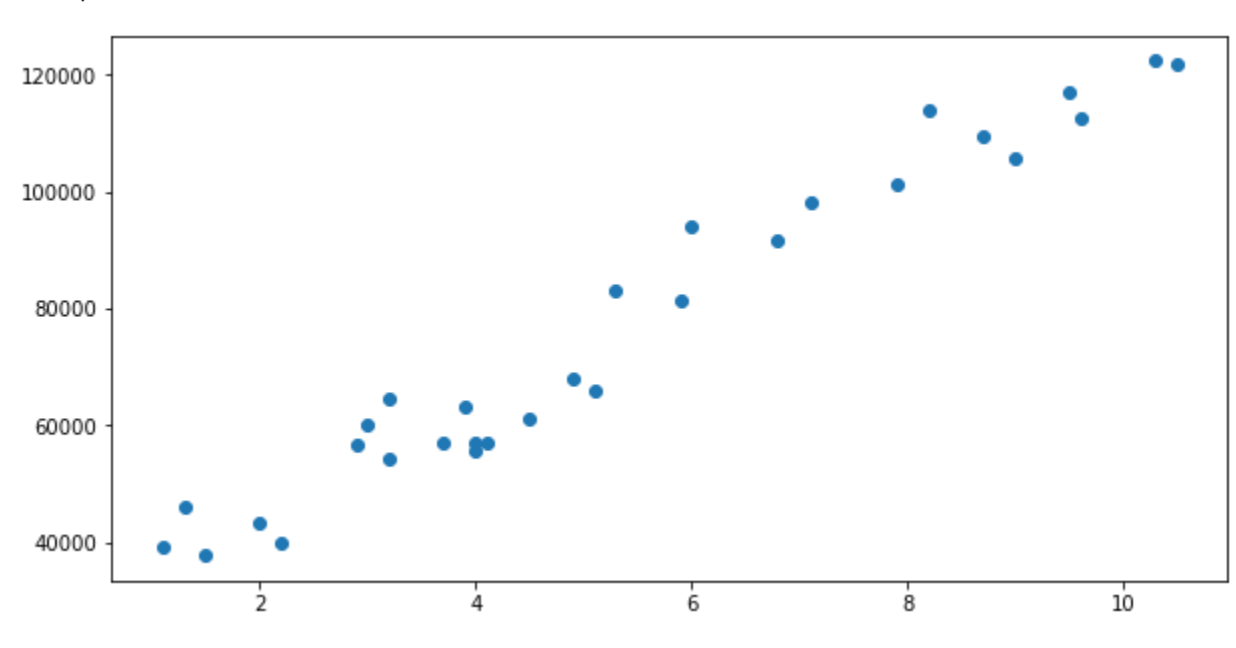
In [16]:
```python
plt.figure(figsize=(10,5))
plt.plot(df.YearsExperience,df.Salary)
```

Out[16]: [<matplotlib.lines.Line2D at 0x28e2fb35100>]



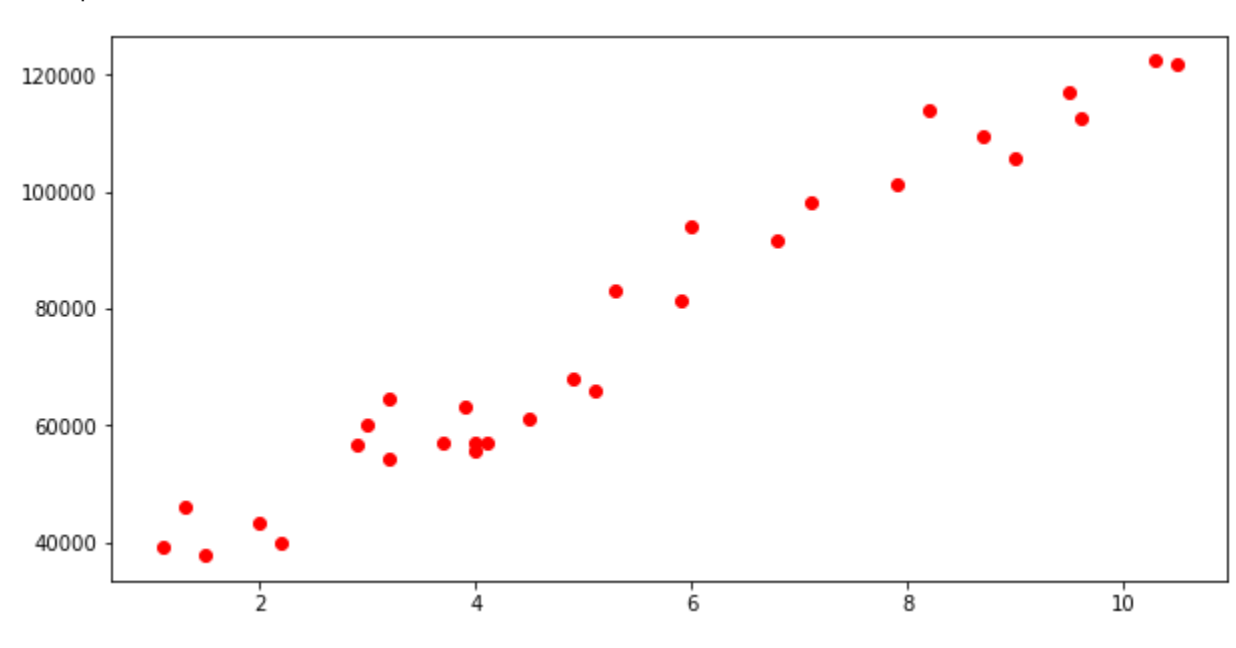
In [17]:
```python
plt.figure(figsize=(10,5))
plt.scatter(df.YearsExperience,df.Salary)
```

Out[17]: <matplotlib.collections.PathCollection at 0x28e2fb80c70>



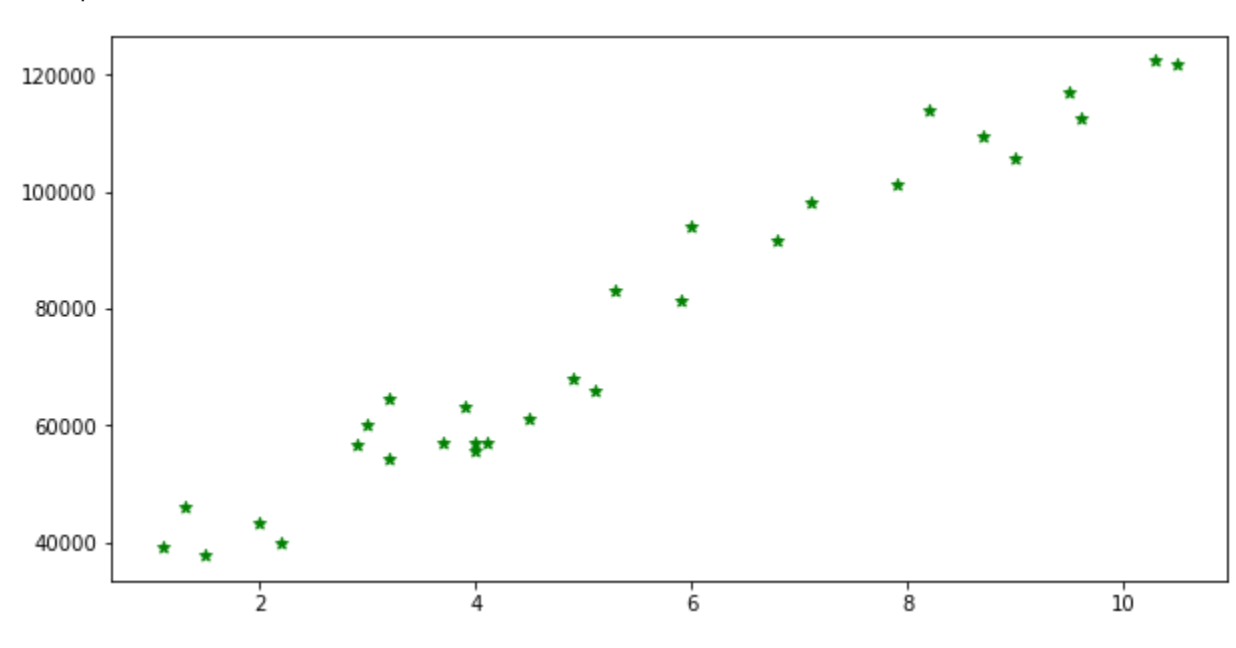
In [18]:
```python
plt.figure(figsize=(10,5))
plt.scatter(df.YearsExperience,df.Salary,color='red')
```

Out[18]: <matplotlib.collections.PathCollection at 0x28e2fbd28e0>



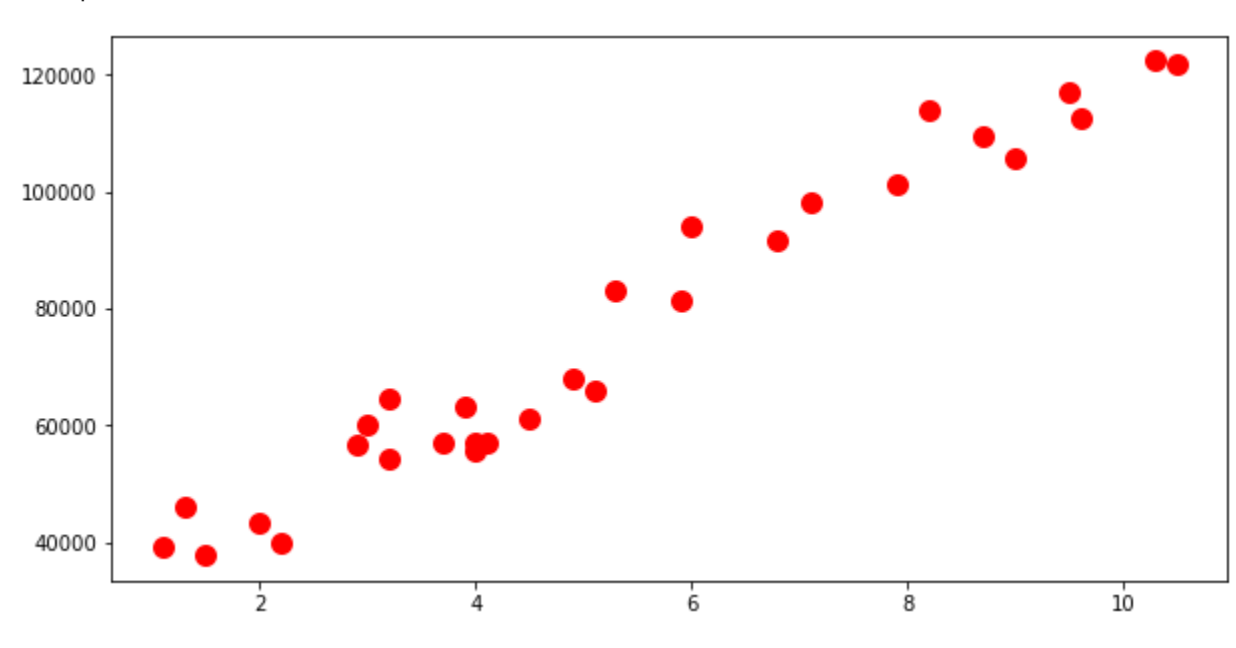
In [19]:
```python
plt.figure(figsize=(10,5))
plt.scatter(df.YearsExperience,df.Salary,color='green',marker='*')
```

Out[19]: <matplotlib.collections.PathCollection at 0x28e30bfb640>

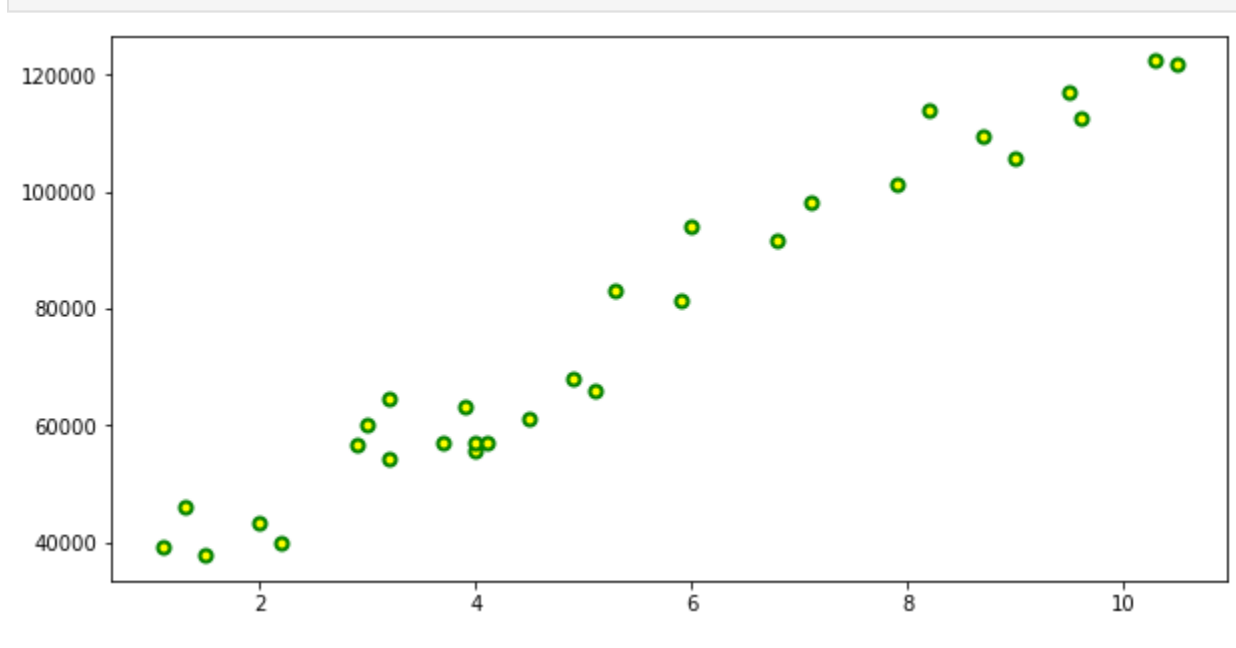


In [20]:
```python
plt.figure(figsize=(10,5))
plt.scatter(df.YearsExperience,df.Salary,color='red',linewidth=5)
```

Out[20]: <matplotlib.collections.PathCollection at 0x28e30c521c0>



In [21]:
```python
plt.figure(figsize=(10,5))
plt.scatter(df.YearsExperience,df.Salary,color='yellow',linewidth=2,edgecolor='green')
plt.show()
```



FROM SCRATCH

In [24]:
```python
x=df.iloc[:,:-1].values
y=df.iloc[:,-1].values
```

In [25]:
```python
x
```

Out[25]:
```
array([[ 1.1],
       [ 1.3],
       [ 1.5],
       [ 2. ],
       [ 2.2],
       [ 2.9],
       [ 3. ],
       [ 3.2],
       [ 3.2],
       [ 3.7],
       [ 3.9],
       [ 4. ],
       [ 4. ],
       [ 4.1],
       [ 4.5],
       [ 4.9],
       [ 5.1],
       [ 5.3],
       [ 5.9],
       [ 6. ],
       [ 6.8],
       [ 7.1],
       [ 7.9],
       [ 8.2],
       [ 8.7],
       [ 9. ],
       [ 9.5],
       [ 9.6],
       [10.3],
       [10.5]])
```

In [26]:
```python
y
```

Out[26]:
```
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.])
```

w0=y_mean-w1*x_mean

w1= sum(x-x_mean)*(y-y_mean)/sum(x-x_mean)^2

In [27]:
```python
# Mean_X and Mean_Y
mean_x=np.mean(x)
mean_y=np.mean(y)

# Total length of the Dataset
n=len(x)

# Using The formula caluculate the w0 and w1
numer=0
denom=0

for i in range(n):
    numer+=(x[i]-mean_x)*(y[i]-mean_y)
    denom+=(x[i]-mean_x)**2

w1=numer/denom
w0=mean_y-(w1*mean_x)
print('The Coefficients are',w0,w1)
```

```
The Coefficients are [25792.20019867] [9449.96232146]
```

In [28]:
```python
plt.scatter(x,y,color='m')

# claculate the y_pred
y_pred=w0+w1*x

#plot the regressor line
plt.plot(x,y_pred,color='g')

#putting the lables
plt.xlabel('Years of Experience in a company')
plt.ylabel('Salary of the Employee')
plt.title('Salary vs Years of Experience')

#show the plot
plt.show()
```

In [29]:
```python
y
```

Out[29]:
```
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 [30]:
```python
y_pred
```

Out[30]:
```
array([[ 36187.15875227],
       [ 38077.15121656],
       [ 39967.14368085],
       [ 44692.12484158],
       [ 46582.11730587],
       [ 53197.09093089],
       [ 54142.08716303],
       [ 56032.07962732],
       [ 56032.07962732],
       [ 60757.06078805],
       [ 62647.05325234],
       [ 63592.04948449],
       [ 63592.04948449],
       [ 64537.04571663],
       [ 68317.03064522],
       [ 72097.0155738 ],
       [ 73987.00803809],
       [ 75877.00050238],
       [ 81546.97789525],
       [ 82491.9741274 ],
       [ 90051.94398456],
       [ 92886.932681  ],
       [100446.90253816],
       [103281.8912346 ],
       [108006.87239533],
       [110841.86109176],
       [115566.84225249],
       [116511.83848464],
       [123126.81210966],
       [125016.80457395]])
```

In [31]:
```python
from sklearn.metrics import r2_score
r2_score(y,y_pred)*100
```

Out[31]: 95.69566641435085

From Using Libraries

In [32]:
```python
from sklearn.linear_model import LinearRegression
```

In [33]:
```python
model=LinearRegression()
```

In [34]:
```python
model.fit(x,y)
```

Out[34]: LinearRegression()

In [35]:
```python
y_pred_lib=model.predict(x)
```

In [36]:
```python
y
```

Out[36]:
```
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 [37]:
```python
y_pred_lib
```

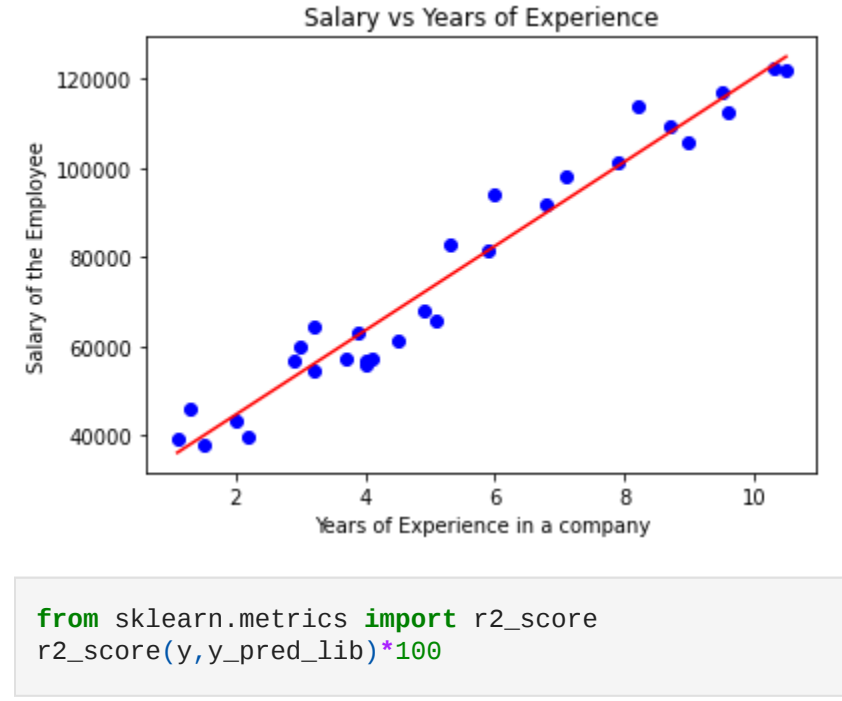
Out[37]:
```
array([ 36187.15875227,  38077.15121656,  39967.14368085,  44692.12484158,
        46582.11730587,  53197.09093089,  54142.08716303,  56032.07962732,
        56032.07962732,  60757.06078805,  62647.05325234,  63592.04948449,
        63592.04948449,  64537.04571663,  68317.03064522,  72097.0155738 ,
        73987.00803809,  75877.00050238,  81546.97789525,  82491.9741274 ,
        90051.94398456,  92886.932681  , 100446.90253816, 103281.8912346 ,
       108006.87239533, 110841.86109176, 115566.84225249, 116511.83848464,
       123126.81210966, 125016.80457395])
```

In [38]:
```python
plt.scatter(x,y,color='b')

#plot the regressor line
plt.plot(x,y_pred_lib,color='r')

#putting the lables
plt.xlabel('Years of Experience in a company')
plt.ylabel('Salary of the Employee')
plt.title('Salary vs Years of Experience')

#show the plot
plt.show()
```



In [39]:
```python
from sklearn.metrics import r2_score
r2_score(y,y_pred_lib)*100
```

Out[39]: 95.69566641435085

In [40]:
```python
x=[[12]]
model.predict(x)
```

Out[40]: array([139191.74805613])

In [41]:
```python
w0+w1*x
25792.20019867+ (9449.96232146*12)
```

Out[41]: 139191.74805619

End