

```
In [96]: import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
```

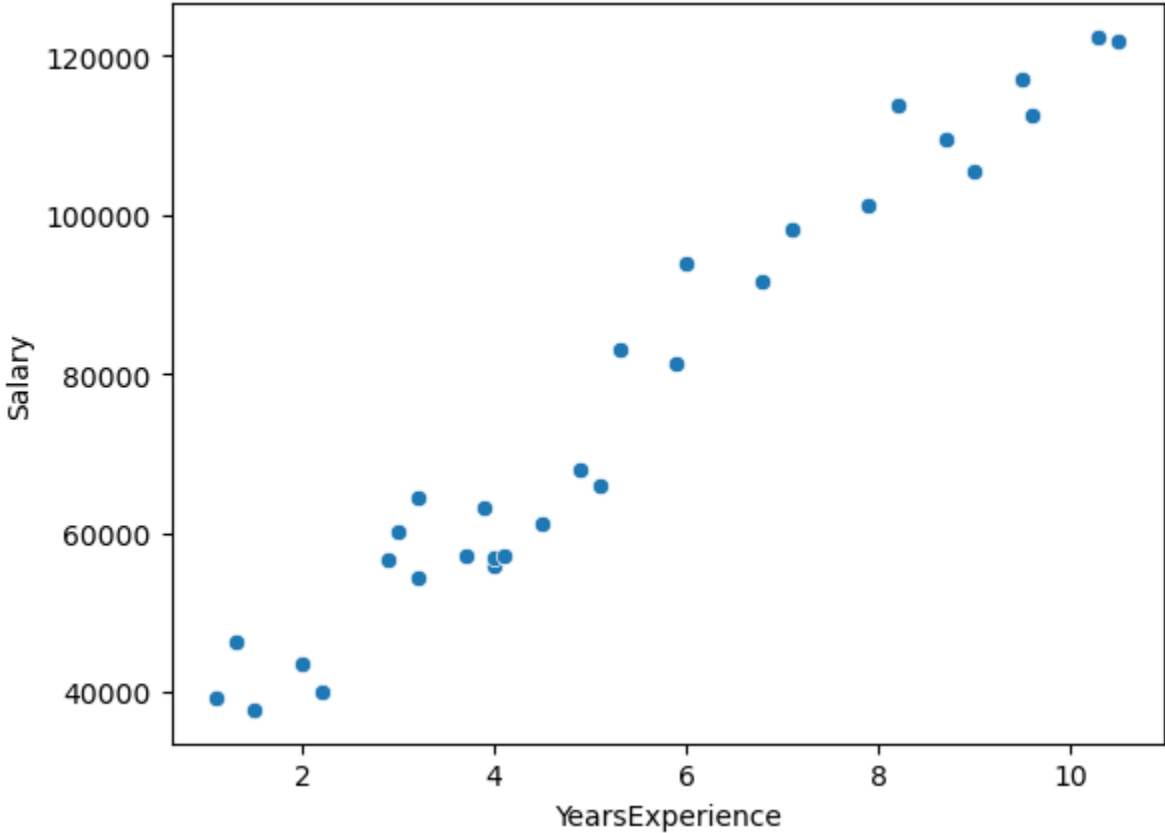
```
In [97]: df = pd.read_csv("Salary_Data.csv")
```

```
In [98]: df.head()
```

	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 [99]: sns.scatterplot(x="YearsExperience", y="Salary", data=df)
```

```
Out[99]: <AxesSubplot:xlabel='YearsExperience', ylabel='Salary'>
```



```
In [100... df.isnull().sum()
df.dropna(inplace=True)
```

```
In [101... df.isnull().sum()
```

```
Out[101]: YearsExperience    0
Salary                    0
dtype: int64
```

```
In [102... x=df.iloc[:,0].values
y=df.iloc[:,1].values
```

```
In [103... x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.2,random_state=3)
```

```
In [104... x_train=x_train.reshape(-1,1)
y_train=y_train.reshape(-1,1)
x_test=x_test.reshape(-1,1)
y_test=y_test.reshape(-1,1)
```

```
In [105... model=LinearRegression()
model
```

```
Out[105]: LinearRegression()
```

```
In [106... model.fit(x_train,y_train)
```

```
Out[106]: LinearRegression()
```

```
In [107... y_pred=model.predict(x_test)
```

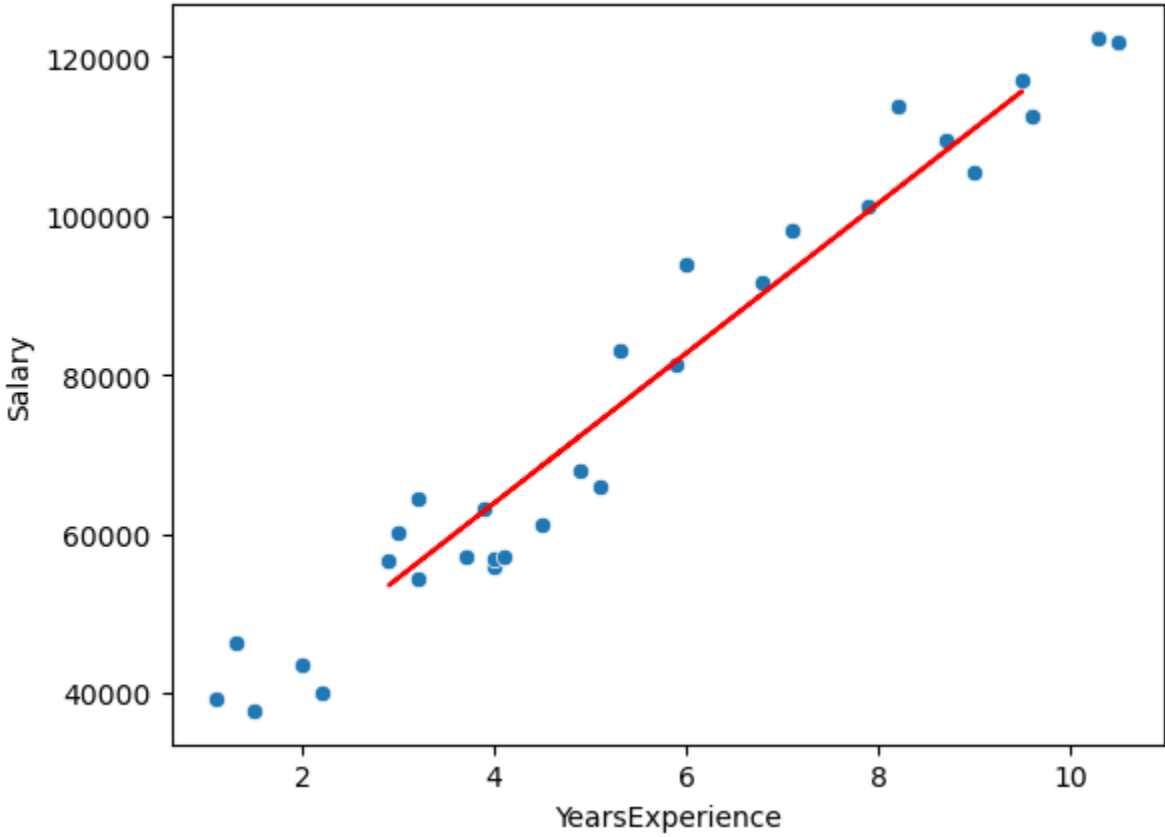
```
In [108... from sklearn.metrics import r2_score
score=r2_score(y_test,y_pred)
score
```

```
Out[108]: 0.9695039421049821
```

```
In [110... # Create a scatterplot
sns.scatterplot(x="YearsExperience", y="Salary", data=df)

# Plot the red line on top of the scatterplot
plt.plot(x_test, y_pred, color="red")

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



```
In [ ]:
```