

▼ Rollno: 1233

Name: Kirti Vichare

```
import matplotlib.pyplot as plt
import pandas as pd
import seaborn as sb
```

```
df= pd.read_csv("Salary_Data.csv")
df
```

|   | 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 |

```
df.columns

Index(['YearsExperience', 'Salary'], dtype='object')
```

```
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 |

```
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

28          10.3  122391.0
```

```
df.isnull()
```

|    | YearsExperience | Salary |
|----|-----------------|--------|
| 0  | False           | False  |
| 1  | False           | False  |
| 2  | False           | False  |
| 3  | False           | False  |
| 4  | False           | False  |
| 5  | False           | False  |
| 6  | False           | False  |
| 7  | False           | False  |
| 8  | False           | False  |
| 9  | False           | False  |
| 10 | False           | False  |
| 11 | False           | False  |
| 12 | False           | False  |
| 13 | False           | False  |
| 14 | False           | False  |
| 15 | False           | False  |
| 16 | False           | False  |
| 17 | False           | False  |
| 18 | False           | False  |
| 19 | False           | False  |
| 20 | False           | False  |
| 21 | False           | False  |
| 22 | False           | False  |
| 23 | False           | False  |
| 24 | False           | False  |
| 25 | False           | False  |
| 26 | False           | False  |
| 27 | False           | False  |
| 28 | False           | False  |

```
#check for null values
df.isnull().sum()

YearsExperience    0
Salary            0
dtype: int64
```

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

```
plt.plot(x,y)  
plt.title("Years of Experience vs Salary")  
plt.xlabel("Years of Experience")  
plt.ylabel("Salary")  
plt.show()
```



```
plt.scatter(x,y)  
plt.title("Years of Experience vs Salary")  
plt.xlabel("Years of Experience")  
plt.ylabel("Salary")  
plt.show()
```



```
from sklearn.model_selection import train_test_split  
x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.25,random_state=1)
```

```
x_train
```

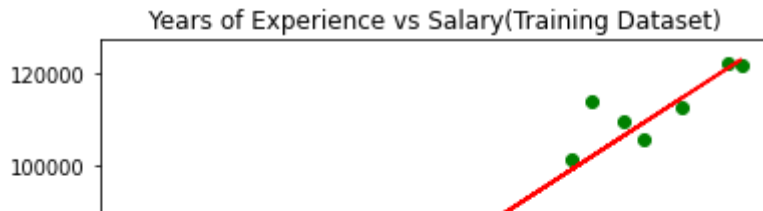
```
array([[ 8.7],  
       [ 7.9],  
       [ 8.2],  
       [ 2.2],  
       [ 1.5],  
       [ 9. ],  
       [ 3. ],  
       [ 5.9],  
       [ 4.1],  
       [ 3.2],  
       [ 9.6],  
       [ 1.3],  
       [ 5.1],  
       [ 1.1],  
       [ 4.9],  
      [10.5],  
      [10.3],  
       [ 3.7],  
       [ 3.2],  
       [ 4. ],  
       [ 4. ],  
       [ 2.9]])
```

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

```
LinearRegression()
```

```
y_pred = regressor.predict(x_test)  
y_train_pred = regressor.predict(x_train)
```

```
plt.scatter(x_train,y_train,color="green")  
plt.plot(x_train,y_train_pred,color="red")  
plt.title("Years of Experience vs Salary(Training Dataset)")  
plt.xlabel("Years of Experience")  
plt.ylabel("Salary")  
plt.show()
```



```
plt.scatter(x_test,y_test,color="blue")
plt.plot(x_test,y_pred,color="red")
plt.title("Years of Experience vs Salary(Training Dataset)")
plt.xlabel("Years of Experience")
plt.ylabel("Salary")
plt.show()
```



## Evaluation

```
from sklearn.metrics import r2_score
import numpy as np
```

```
print("Mean absolute error: %.2f" % np.mean(np.absolute(y_pred - y_test)))
print("Mean Squared error: %.2f" % np.mean((y_pred - y_test)**2))
print("Root Mean Squared error: %.2f" % np.sqrt(np.mean(y_pred - y_test)**2))
print("r2_score: %.2f" % r2_score(y_pred,y_test))
```

```
Mean absolute error: 5215.55
Mean Squared error: 40884620.50
Root Mean Squared error: 3378.35
r2_score: 0.89
```

```
print("Training Data")
print("Mean absolute error: %.2f" % np.mean(np.absolute(y_train_pred - y_train)))
print("Mean Squared error: %.2f" % np.mean((y_train_pred - y_train)**2))
print("Root Mean Squared error: %.2f" % np.sqrt(np.mean(y_train_pred - y_train)**2))
print("r2_score: %.2f" % r2_score(y_train_pred,y_train))
```

Training Data

Mean absolute error: 4538.96

Mean Squared error: 29258602.85

Root Mean Squared error: 0.00

r2\_score: 0.96

```
plt.scatter(x_test,y_test,color="blue")
plt.plot(x_test,y_pred,color="orange")
plt.title("Years of Experience vs Salary(Training Dataset)")
plt.xlabel("Years of Experience")
plt.ylabel("Salary")
plt.show()
```



---

✓ 0s completed at 3:18 PM ● ✕