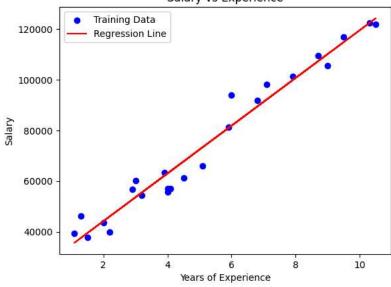
```
from google.colab import files
uploaded = files.upload()
Choose Files | Salary_dataset.csv
     • Salary dataset.csv(text/csv) - 454 bytes, last modified: 6/26/2025 - 100% done
import pandas as pd
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
data = pd.read_csv('Salary_dataset.csv')
print(data.head())
print(data.shape)
print(data.isnull().sum())
x = data.iloc[:, :1].values
y = data.iloc[:, 1:2].values
x_train, x_test, y_train, y_test = train_test_split(x, y, test_size=0.2, random_state=42)
model = LinearRegression()
model.fit(x_train, y_train)
y_pred = model.predict(x_test)
print("Predicted:", y_pred)
print("Actual:", y_test)
plt.scatter(x_train, y_train, color='blue', label='Training Data')
plt.plot(x_train, model.predict(x_train), color='red', label='Regression Line')
plt.title('Salary vs Experience')
plt.xlabel('Years of Experience')
plt.ylabel('Salary')
plt.legend()
plt.show()
plt.scatter(x_test, y_test, color='red')
plt.plot(x_train, model.predict(x_train), color='green')
plt.title('SALARY VS EXPERIENCE (testing set)')
plt.xlabel('Years of Experience')
plt.ylabel('Salary')
plt.show()
```

```
YearsExperience
                     Salary
               1.1 39343.0
                    46205.0
               1.3
2
                   37731.0
3
               2.0 43525.0
4
               2.2 39891.0
(30, 2)
YearsExperience
Salary
dtype: int64
Predicted: [[115790.21011287]
 [ 71498.27809463]
[102596.86866063]
 [ 75267.80422384]
   55477.79204548]
[ 60189.69970699]]
Actual: [[112635.]
 [ 67938.]
 [113812.]
 [ 83088.]
 [ 64445.]
[ 57189.]]
```





SALARY VS EXPERIENCE (testing set)

