```
Start coding or generate with AI.
```

```
from google.colab import drive
drive.mount('/content/drive')
```

```
import pandas as pd
```

```
df = pd.read_csv('/content/Salary_dataset.csv')
```

```
x=df.YearsExperience
y=df.Salary
```

```
import numpy as np
```

```
x=np.array(x).reshape(-1,1)
```

```
y=np.array(y).reshape(-1,1)
```

```
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=42
)
```

```
x_train.shape, y_train.shape, x_test.shape, y_test.shape
```

```
((24, 1), (24, 1), (6, 1), (6, 1))
```

```
x_train, x_test, y_train, y_test = train_test_split(
    x, y,
    test_size=0.2,
    random_state=0
)
```

```
x_train
```

```
array([[ 9.7],
       [ 4.1],
       [ 5.4],
       [ 8. ],
       [ 3. ],
       [ 5.2],
       [ 3.3],
       [ 4.6],
       [ 8.3],
       [ 6.9],
       [ 1.4],
       [10.6],
       [ 3.1],
       [ 2.3],
       [ 6. ],
       [ 6.1],
       [ 3.8],
       [ 3.3],
       [ 9.1],
       [ 2.1],
       [ 1.2],
       [ 7.2],
       [ 5. ],
       [ 4.1]])
```

```
from sklearn.linear_model import LinearRegression
```

```
model = LinearRegression()
```

```
model.fit(x_train, y_train)
```

```
▾ LinearRegression
LinearRegression()
```

```
y_pred = model.predict(x_test)
```

```
print("Actual salary:",y_test[:5])
print("Predicted salary:",y_pred[:5])
```

```
Actual salary: [[ 37732.]
 [122392.]
 [ 57082.]
 [ 63219.]
 [116970.]]
Predicted salary: [[ 40749.96184072]
 [122700.62295594]
 [ 64962.65717022]
 [ 63100.14214487]
 [115250.56285456]]
```

```
#Calculate Mean Squared Error
```

```
from sklearn.metrics import mean_squared_error
```

```
mse = mean_squared_error(y_test, y_pred)
print("Mean Squared Error:", mse)
```

```
Mean Squared Error: 12823412.298126565
```

```
#calculate RMSE
```

```
rmse = np.sqrt(mse)
print("Root Mean Squared Error:", rmse)
```

```
Root Mean Squared Error: 3580.9792373213454
```

```
#From R*2 Value
```

```
from sklearn.metrics import r2_score
```

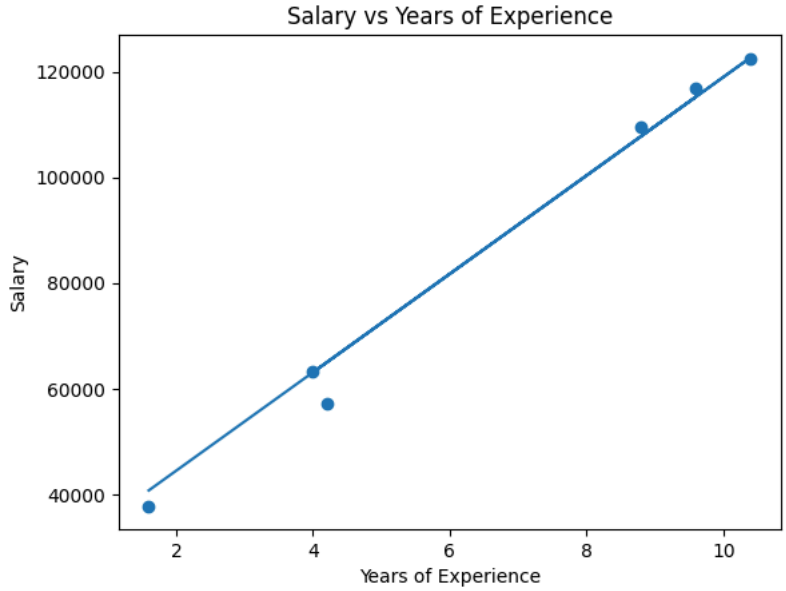
```
R2 = r2_score(y_test, y_pred)
print("R-Squared value:", R2)
```

```
R-Squared value: 0.988169515729126
```

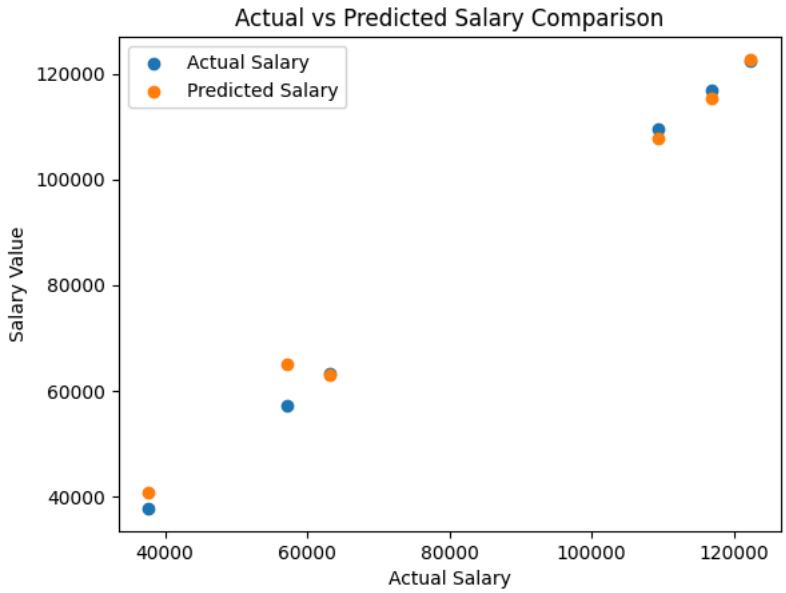
```
import matplotlib.pyplot as plt
```

```
plt.scatter(x_test, y_test)
plt.plot(x_test, y_pred)

plt.xlabel('Years of Experience')
plt.ylabel('Salary')
plt.title('Salary vs Years of Experience')
plt.show()
```




```
plt.scatter(y_test, y_test, label="Actual Salary")
plt.scatter(y_test, y_pred, label="Predicted Salary")
plt.xlabel("Actual Salary")
plt.ylabel("Salary Value")
plt.title("Actual vs Predicted Salary Comparison")
plt.legend()
plt.show()
```




```
slope = model.coef_[0]
intercept = model.intercept_

print("Slope (m):", slope)
print("Intercept (c):", intercept)
```

```
Slope (m): [9312.57512673]
Intercept (c): [25849.84163796]
```

```
experience = [[5]]
```

```
predicted_salary = model.predict(experience)
print("Predicted salary for 5 years experience:", predicted_salary[0])
```

```
Predicted salary for 5 years experience: [72412.7172716]
```

```
print("Actual salaries (sample):", y_test[:5])
print("Predicted salaries (sample):", y_pred[:5])
```

```
Actual salaries (sample): [[ 37732.]
 [122392.]
 [ 57082.]
 [ 63219.]
 [116970.]]
Predicted salaries (sample): [[ 40749.96184072]
 [122700.62295594]
 [ 64962.65717022]
 [ 63100.14214487]
```