```python
In [1]:   #Import libraries
          import pandas as pd
          import numpy as np
          from sklearn.linear_model import LinearRegression
          import matplotlib.pyplot as plt
```

```python
In [3]:   #load data
          saldf=pd.read_csv('/content/Salary Data.csv')
```

```python
In [4]:   #view data
          saldf.head()
```

Out[4]:

	Age	Gender	Education Level	Job Title	Years of Experience	Salary
0	32.0	Male	Bachelor's	Software Engineer	5.0	90000.0
1	28.0	Female	Master's	Data Analyst	3.0	65000.0
2	45.0	Male	PhD	Senior Manager	15.0	150000.0
3	36.0	Female	Bachelor's	Sales Associate	7.0	60000.0
4	52.0	Male	Master's	Director	20.0	200000.0

```python
In [5]:   saldf.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 375 entries, 0 to 374
Data columns (total 6 columns):
 #   Column               Non-Null Count  Dtype
---  ------               --------------  -----
 0   Age                  373 non-null    float64
 1   Gender               373 non-null    object
 2   Education Level      373 non-null    object
 3   Job Title            373 non-null    object
 4   Years of Experience  373 non-null    float64
 5   Salary               373 non-null    float64
dtypes: float64(3), object(3)
memory usage: 17.7+ KB
```

```python
In [6]:   saldf.isnull().sum()
```

Out[6]:

	0
Age	2
Gender	2
Education Level	2
Job Title	2
Years of Experience	2
Salary	2

dtype: int64

In [7]:
```python
#prepare independent and dependent variables
inp=saldf[['Years of Experience']]
out=saldf['Salary']
```

In [8]:
```python
#create an instance of LR
LR=LinearRegression()
```

In [9]:
```python
#train the model
df_cleaned = saldf.dropna(subset=['Years of Experience', 'Salary'])

inp_cleaned = df_cleaned[['Years of Experience']]
out_cleaned = df_cleaned['Salary']

LR.fit(inp_cleaned, out_cleaned)
```

Out[9]:
```
▼ LinearRegression ⓘ ⓘ

LinearRegression()
```

In [10]:
```python
#predict when experience is 5years
LR.predict([[5]])
```
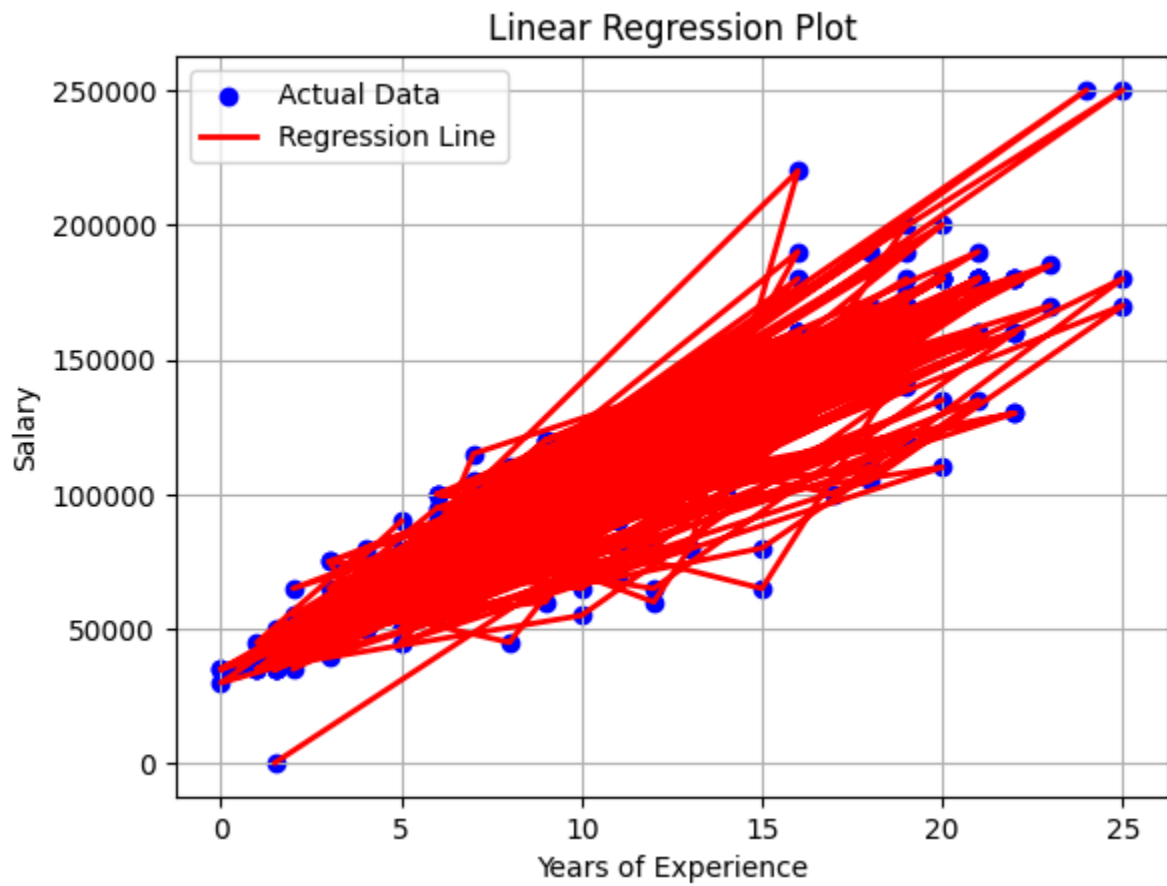
```
/usr/local/lib/python3.12/dist-packages/sklearn/utils/validation.py:2739: UserW
arning: X does not have valid feature names, but LinearRegression was fitted wi
th feature names
  warnings.warn(
```

Out[10]: array([66143.76948947])

In [11]:
```python
plt.scatter(inp,out, color='blue', label='Actual Data')
plt.plot(inp,out, color='red', linewidth=2, label='Regression Line')

plt.xlabel('Years of Experience')
plt.ylabel('Salary')
plt.title('Linear Regression Plot')
plt.legend()
```

```
plt.grid(True)
plt.show()
```

Linear Regression Plot

In [12]:
```python
import pickle
with open("Salaryprediction.pkl","wb") as f:
  pickle.dump(LR,f)
from google.colab import files
files.download('Salaryprediction.pkl')
```