```python
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_absolute_error,
mean_squared_error,r2_score
%matplotlib inline

df=pd.read_csv("salary_prediction_dataset.csv")

df.head()
```

```
    YearsExperience          Salary
0         8.116262   41469.816295
1        19.063572   88641.137315
2        14.907885   76440.413727
3        12.374511   64925.484725
4         3.964354   22620.723082
```

```python
df.isnull().sum()
```

```
YearsExperience     0
Salary              0
dtype: int64
```

```python
plt.figure(figsize=(8,5))
sns.scatterplot(x=df['YearsExperience'],y=df['Salary'])
plt.xlabel("Years of Experience")
plt.ylabel("Salary")
plt.title("Years of Experience vc Salary")
plt.show()
```

Years of Experience vc Salary


```python
x=df[['YearsExperience']]
y=df['Salary']
x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.2,random_state=42)
x_train.shape,x_test.shape
```

```
((800, 1), (200, 1))
```

```python
model=LinearRegression()
model.fit(x_train,y_train)
print(f"Intercept:{model.intercept_}")
print(f"Coefficient:{model.coef_[0]}")
```

```
Intercept:995.5770464802335
Coefficient:4953.709417260947
```

```python
y_pred=model.predict(x_test)
results=pd.DataFrame({'Actual Salary':y_test.values,'Predicted Salary':y_pred})
results.head()
```

```
   Actual Salary  Predicted Salary
0   46144.843128      41798.916309
1   85731.005173      82808.016799
2   55059.748116      49496.929481
```

```
3    42143.595304        39465.280850
4    88585.759395        95369.573814

plt.figure(figsize=(8,5))
plt.scatter(x_train,y_train,color='blue',label="Actual Salary(Training
DAta)")
plt.plot(x_train,model.predict(x_train),color='red',label="Regression"
)
plt.xlabel("Years of Experience")
plt.ylabel("Salary")
plt.title("Salary Prediction(Training Data)")
plt.legend()
plt.show()

plt.figure(figsize=(8,5))
plt.scatter(x_test,y_test,color='green',label="Actual Salary(Test
Data)")
plt.scatter(x_test,y_pred,color='red',label="Predicted Salary(Test
Data")
plt.plot(x_train,model.predict(x_train),color='red',label="Regression"
)
plt.xlabel("Years of Experience")
plt.ylabel("Salary")
plt.title("Salary Prediction(Test Data)")
plt.legend()
plt.show()
```

Salary Prediction(Training Data)

Salary Prediction(Test Data)
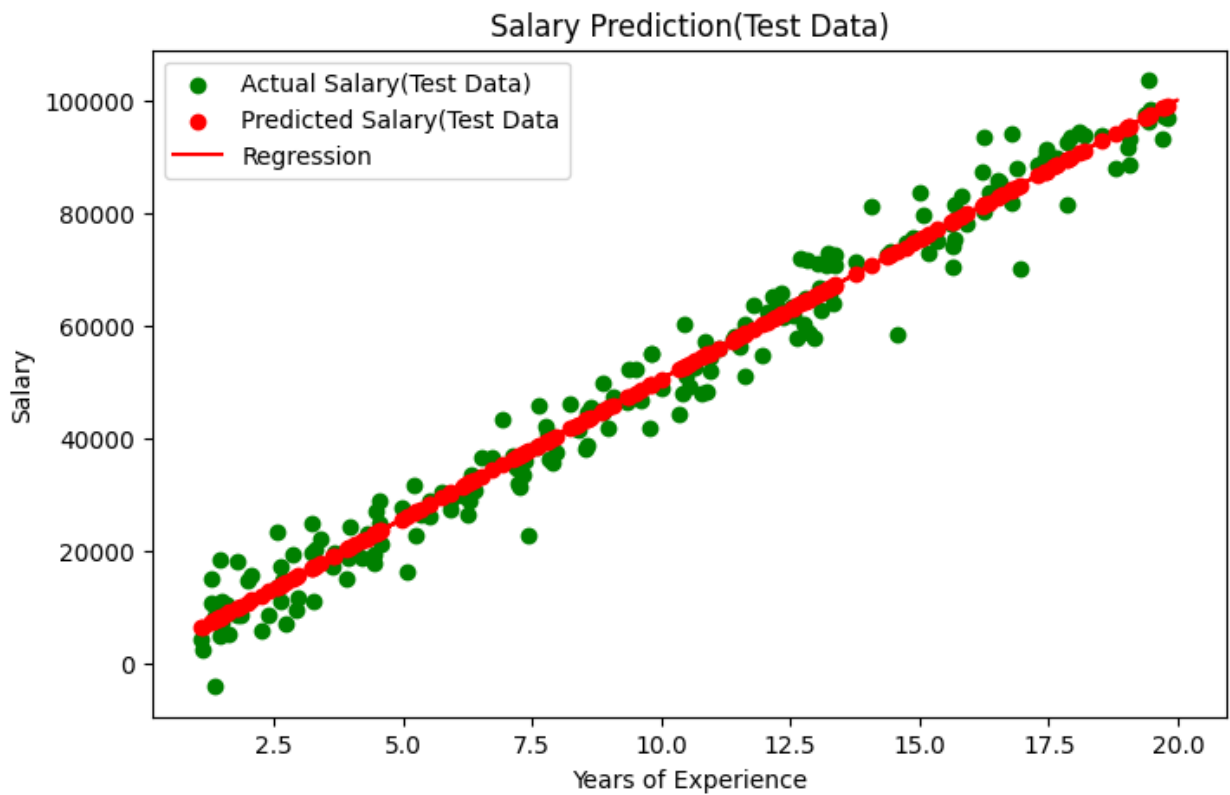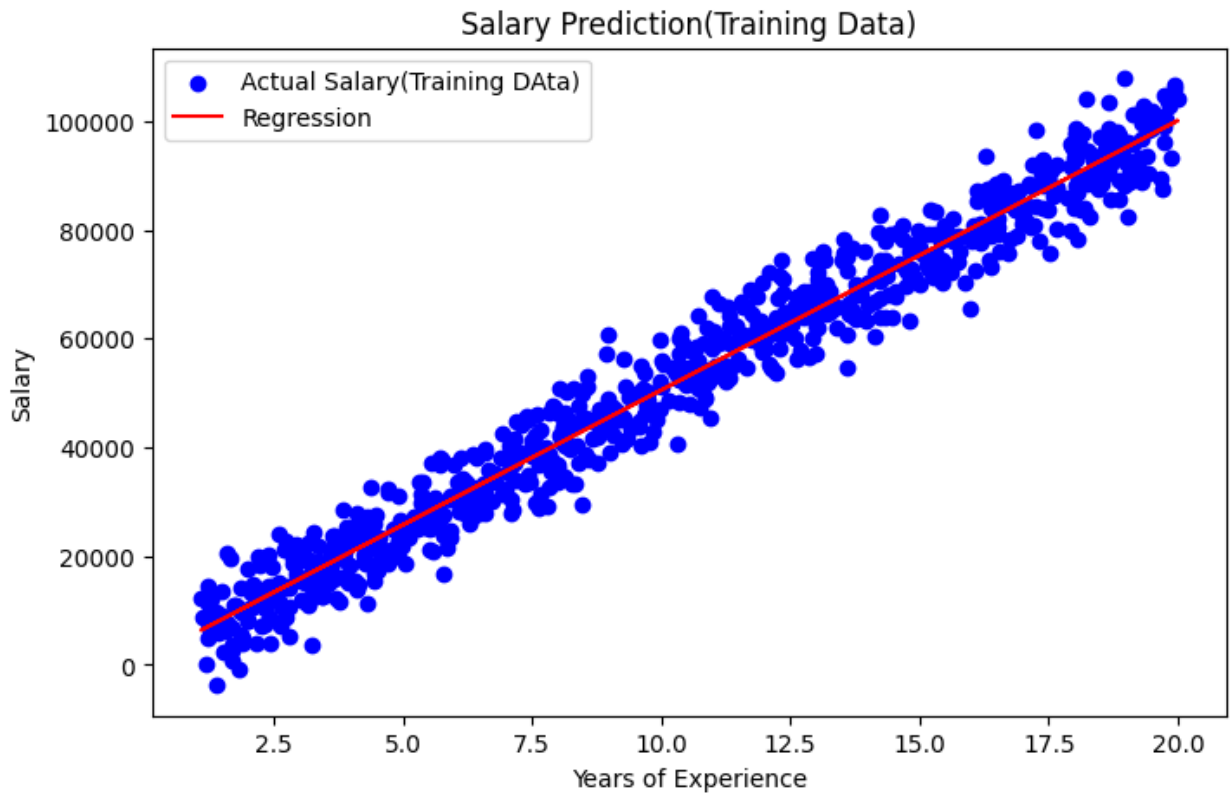
```
mae=mean_absolute_error(y_test,y_pred)
mse=mean_squared_error(y_test,y_pred)
r2=r2_score(y_test,y_pred)
print(f"Mean Absolute Error(MAE):{mae}")
print(f"Mean Squared Error(MSE):{mse}")
print(f"R2 Score:{r2}")
```

```
Mean Absolute Error(MAE):3602.3955438952053
Mean Squared Error(MSE):21434311.66454385
R2 Score:0.9720753738957405
```

```
experience = float(input("enter years of experience"))
predicted_salary=model.predict([[experience]])[0]
print(f"Predicted Salary for {experience} years of experience is:
{predicted_salary:.2f}")
```

```
enter years of experience 50
```

```
Predicted Salary for 50.0 years of experience is:248681.05
```

```
C:\Users\hp\AppData\Roaming\Python\Python312\site-packages\sklearn\
base.py:493: UserWarning: X does not have valid feature names, but
LinearRegression was fitted with feature names
  warnings.warn(
```