In [ ]:
```python
#7

from google.colab import files
uploaded = files.upload()

import numpy as np
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
import pickle

df = pd.read_csv('Salary_data.csv')
print("=== Dataset Preview ===")
display(df)

print("\n=== Dataset Info ===")
df.info()

df.dropna(inplace=True)

print("\n=== After Dropping Missing Values ===")
df.info()

print("\n=== Descriptive Statistics ===")
display(df.describe())

features = df.iloc[:, [0]].values
label = df.iloc[:, [1]].values

x_train, x_test, y_train, y_test = train_test_split(features, label, test_size

model = LinearRegression()
model.fit(x_train, y_train)

print("\n=== Model Performance ===")
print("Training Score:", model.score(x_train, y_train))
print("Testing Score:", model.score(x_test, y_test))
print("Coefficient:", model.coef_)
print("Intercept:", model.intercept_)

pickle.dump(model, open('SalaryPred.model', 'wb'))

model = pickle.load(open('SalaryPred.model', 'rb'))

yr_of_exp = float(input("Enter Years of Experience: "))
yr_of_exp_NP = np.array([[yr_of_exp]])
Salary = model.predict(yr_of_exp_NP)

print(f"\nEstimated Salary for {yr_of_exp} years of experience is {Salary[0][0
```

Browse...

Upload widget is only available when the cell has been executed in the current browser session. Please rerun this cell to enable.

```
Saving Salary_data.csv to Salary_data (1).csv
=== Dataset Preview ===
```

	YearsExperience	Salary
0	1.1	39343
1	1.3	46205
2	1.5	37731
3	2.0	43525
4	2.2	39891
5	2.9	56642
6	3.0	60150
7	3.2	54445
8	3.2	64445
9	3.7	57189
10	3.9	63218
11	4.0	55794
12	4.0	56957
13	4.1	57081
14	4.5	61111
15	4.9	67938
16	5.1	66029
17	5.3	83088
18	5.9	81363
19	6.0	93940
20	6.8	91738
21	7.1	98273
22	7.9	101302
23	8.2	113812
24	8.7	109431
25	9.0	105582
26	9.5	116969
27	9.6	112635
28	10.3	122391
29	10.5	121872

```
=== Dataset 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     int64  
dtypes: float64(1), int64(1)
memory usage: 612.0 bytes

=== After Dropping Missing Values ===
<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     int64  
dtypes: float64(1), int64(1)
memory usage: 612.0 bytes

=== Descriptive Statistics ===
```

	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

```
=== Model Performance ===
Training Score: 0.9603182547438908
Testing Score: 0.9184170849214232
Coefficient: [[9281.30847068]]
Intercept: [27166.73682891]
Enter Years of Experience: 4

Estimated Salary for 4.0 years of experience is 64291.97
```