Practical No. 08

Aim: To perform and analysis of Linear Regression Algorithm

In [5]:

```
#Name : Taufiq Rafik Nagori
#Roll no. : 77 (BDA-B77)
#Section : B
#Subject : PE-II
```

In [7]:

```
import os
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

In [9]:

```
os.getcwd()
```

Out[9]:

```
'C:\\Users\\USER'
```

In [11]:

```
os.chdir("C:\\Users\\USER\\Desktop")
```

In [13]:

```
data = pd.read_csv("Salary.csv")
```

In [15]:

```
data.head()
```

Out[15]:

	YearsExperience	Salary
0	1.1	39343
1	1.3	46205
2	1.5	37731
3	2.0	43525
4	2.2	39891

In [17]:

```
data.tail()
```

Out[17]:

	YearsExperience	Salary
30	11.2	127345
31	11.5	126756
32	12.3	128765
33	12.9	135675

	YearsExperience	Salary
34	13.5	139465

In [19]:

```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 35 entries, 0 to 34
Data columns (total 2 columns):
 #   Column           Non-Null Count  Dtype  
---  ------           --------------  -----  
 0   YearsExperience  35 non-null     float64
 1   Salary           35 non-null     int64  
dtypes: float64(1), int64(1)
memory usage: 692.0 bytes
```

In [21]:

```
data.describe()
```

Out[21]:

	YearsExperience	Salary
count	35.000000	35.000000
mean	6.308571	83945.600000
std	3.618610	32162.673003
min	1.100000	37731.000000
25%	3.450000	57019.000000
50%	5.300000	81363.000000
75%	9.250000	113223.500000
max	13.500000	139465.000000

In [23]:

```
data.shape
```

Out[23]:

```
(35, 2)
```

In [25]:

```
data.size
```

Out[25]:

```
70
```

In [27]:

```
data.columns
```

Out[27]:

```
Index(['YearsExperience', 'Salary'], dtype='object')
```

In [29]:

```
data.ndim
```

Out[29]:

```
2
```

In [31]:

```
data.isnull()
```

Out[31]:

	YearsExperience	Salary
0	False	False
1	False	False
2	False	False
3	False	False
4	False	False
5	False	False
6	False	False
7	False	False
8	False	False
9	False	False
10	False	False
11	False	False
12	False	False
13	False	False
14	False	False
15	False	False
16	False	False
17	False	False
18	False	False
19	False	False
20	False	False
21	False	False
22	False	False
23	False	False
24	False	False
25	False	False
26	False	False
27	False	False
28	False	False
29	False	False
30	False	False
31	False	False

	YearsExperience	Salary
32	False	False
33	False	False
34	False	False

In [33]:

```
data.isnull().sum()
```

Out[33]:

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

In [35]:

```
sns.scatterplot(x='YearsExperience', y='Salary',data=data)
plt.title('Experience Vs Salary')
plt.show()
```

In [37]:

```
x = data[['YearsExperience']]
y = data['Salary']
```

In [39]:

```
from sklearn.model_selection import train_test_split
```

In [41]:

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

In [43]:

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

In [45]:

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

Out[45]:

LinearRegression

LinearRegression()

In [47]:

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

In [49]:

```
from sklearn.metrics import mean_squared_error, r2_score
print("Mean Squared Error:", mean_squared_error(y_test, y_predict))
print("R² Score:", r2_score(y_test, y_predict))
```

```
Mean Squared Error: 55761791.30626011
R² Score: 0.891423414004278
```

In [51]:

```
from sklearn.metrics import r2_score,mean_absolute_error
```

In [53]:

```
mean_absolute_error(y_test,y_predict)
```

Out[53]:

```
6692.364094497281
```

In [55]:

```
plt.scatter(x_test, y_test, color='blue', label='Actual')
plt.plot(x_test, y_predict, color='red', linewidth=2, label='Predicted')
plt.title("Linear Regression - Experience vs Salary")
plt.xlabel("Years of Experience")
plt.ylabel("Salary")
plt.show()
```