In [1]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
```

In [2]:
```
df=pd.read_csv(r'file:///C:\Users\Hp\Downloads\Salary_Data.csv')
```

In [3]:
```
df
```

Out[3]:

	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

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

In [4]:
```
df.shape
```

Out[4]:
```
(30, 2)
```

In [6]:
```
df.describe()
```

Out[6]:

	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

In [7]:
```
df.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
```

In [8]:
```
x=df[['YearsExperience']]
```

In [9]:
```
y=df['Salary']
```

In [10]:
```
x
```

Out[10]:

	YearsExperience
0	1.1
1	1.3
2	1.5
3	2.0
4	2.2
5	2.9
6	3.0
7	3.2
8	3.2
9	3.7
10	3.9
11	4.0
12	4.0
13	4.1
14	4.5
15	4.9
16	5.1
17	5.3
18	5.9
19	6.0
20	6.8
21	7.1
22	7.9
23	8.2
24	8.7
25	9.0
26	9.5
27	9.6
28	10.3
29	10.5

In [11]:
```
y
```

Out[11]:
```
0     39343
1     46205
2     37731
3     43525
4     39891
5     56642
6     60150
7     54445
8     64445
9     57189
10    63218
11    55794
12    56957
13    57081
14    61111
15    67938
16    66029
17    83088
18    81363
19    93940
20    91738
21    98273
22   101302
23   113812
24   109431
25   105582
26   116969
27   112635
28   122391
29   121872
Name: Salary, dtype: int64
```

In [12]:
```
from sklearn.model_selection import train_test_split
```

In [14]:
```
x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.25,random_state=42)
```

In [15]:
```
from sklearn.linear_model import LinearRegression
```

In [16]:
```
lr=LinearRegression()
```

In [17]:
```
reg=lr.fit(x_train,y_train)
```

In [18]:
```
reg
```

Out[18]:
```
LinearRegression()
```

In [21]:
```
y_pred_train=reg.predict(x_train)
```

In [23]:
```
y_pred_test=reg.predict(x_test)
```

In [24]:
```
#metrics
from sklearn.metrics import mean_squared_error,mean_absolute_error,r2_score
```

In [28]:
```
print("mse=",mean_squared_error(y_pred_test,y_test))
print("mae=",mean_absolute_error(y_pred_test,y_test))
print("r2_score=",r2_score(y_pred_test,y_test))
```

```
mse= 38802588.99247065
mae= 5056.995466663594
r2_score= 0.9346100018376171
```

In [ ]:
```

```