```
# Experiment No: 9
```

```
# Aim: Simple Linear Regression
```

```
# Name: Sakshi Padmakar Yeole
```

```
# Class: 3rd year(B)
```

```
# Roll No: 69
```

```
# Date: 5th October 2024
```

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

```python
import pandas as pd
```

```python
import os
```

```python
os.getcwd()
```

```
'C:\\Users\\hp'
```

```python
os.chdir("C:\\Users\\hp\\OneDrive\\Desktop")
```

```python
df=pd.read_csv("Salary.csv")
```

```python
df.head()
```

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

```python
df.tail()
```

	YearsExperience	Salary
30	11.2	127345
31	11.5	126756
32	12.3	128765
33	12.9	135675
34	13.5	139465

```python
df.head(20)
```

	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

In [13]:
```python
df[5:10]
```

Out[13]:

	YearsExperience	Salary
5	2.9	56642
6	3.0	60150
7	3.2	54445
8	3.2	64445
9	3.7	57189

In [14]:
```python
df.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: 688.0 bytes
```

In [15]:
```python
df.describe()
```

Out[15]:

	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 [16]:
```python
df.shape
```

Out[16]:
```
(35, 2)
```

```
In [17]:   df.size

Out[17]:   70


In [18]:   df.ndim

Out[18]:   2


In [19]:   df.columns

Out[19]:   Index(['YearsExperience', 'Salary'], dtype='object')


In [20]:   df.isnull().sum()

Out[20]:   YearsExperience    0
           Salary             0
           dtype: int64


In [21]:   # Assiging values in X & Y
           x=df.iloc[:,:-1].values
           y=df.iloc[:,-1].values


           #x=df['year of experience']
           #y=df['salary']

In [22]:   print(x)

           [[ 1.1]
            [ 1.3]
            [ 1.5]
            [ 2. ]
            [ 2.2]
            [ 2.9]
            [ 3. ]
            [ 3.2]
            [ 3.2]
            [ 3.7]
            [ 3.9]
            [ 4. ]
            [ 4. ]
            [ 4.1]
            [ 4.5]
            [ 4.9]
            [ 5.1]
            [ 5.3]
            [ 5.9]
            [ 6. ]
            [ 6.8]
            [ 7.1]
            [ 7.9]
            [ 8.2]
            [ 8.7]
            [ 9. ]
            [ 9.5]
            [ 9.6]
            [10.3]
            [10.5]
            [11.2]
            [11.5]
            [12.3]
            [12.9]
            [13.5]]


In [23]:   print(y)

           [ 39343  46205  37731  43525  39891  56642  60150  54445  64445  57189
```

```
 63218  55794  56957  57081  61111  67938  66029  83088  81363  93940
 91738  98273 101302 113812 109431 105582 116969 112635 122391 121872
127345 126756 128765 135675 139465]
```

In [24]:
```python
# Splitting testdata into x_train,x_test,y_train,y_test
from sklearn.model_selection import train_test_split
x_train,x_test,y_train,y_test = train_test_split(x,y,test_size=.3,random_state=42)
```

In [25]:
```python
print(x_train)
```

```
[[12.9]
 [ 1.1]
 [ 2.2]
 [ 5.3]
 [ 9.6]
 [ 2.9]
 [ 4. ]
 [ 1.3]
 [ 1.5]
 [12.3]
 [ 2. ]
 [11.2]
 [ 8.2]
 [11.5]
 [ 3.9]
 [ 7.9]
 [ 5.9]
 [ 9. ]
 [ 3. ]
 [ 6.8]
 [13.5]
 [ 3.2]
 [ 4.5]
 [10.3]]
```

In [26]:
```python
print(x_test)
```

```
[[ 9.5]
 [ 4.1]
 [ 8.7]
 [ 7.1]
 [ 4.9]
 [10.5]
 [ 6. ]
 [ 4. ]
 [ 3.2]
 [ 5.1]
 [ 3.7]]
```

In [27]:
```python
print(y_train)
```

```
[135675  39343  39891  83088 112635  56642  55794  46205  37731 128765
  43525 127345 113812 126756  63218 101302  81363 105582  60150  91738
 139465  54445  61111 122391]
```

In [28]:
```python
print(y_test)
```

```
[116969  57081 109431  98273  67938 121872  93940  56957  64445  66029
  57189]
```

In [29]:
```python
from sklearn.linear_model import LinearRegression
lr=LinearRegression()
lr.fit(x_train,y_train)
```

Out[29]:
```
LinearRegression()
```

In [30]:
```python
#Assigning Coefficient (slope) to m
m = lr.coef_
```

```
In [31]:    print("Coefficient  :" , m)

Coefficient  : [8555.33918938]
```

```
In [32]:    #Assigning Y-intercept to a
            c = lr.intercept_
```

```
In [33]:    print("Intercept : ", c)

Intercept :  29602.07353482097
```

```
In [34]:    lr.score(x_test,y_test) * 100
```

Out[34]:   91.71426108885095

```
In [35]:    df.isnull().any()
```

Out[35]:   YearsExperience   False
            Salary   False
            dtype: bool

```
In [ ]:
```