

```
In [1]: #Name :janhavi Nitin warghade  
#Roll No.:69  
#Sec:3C  
#Sub:E.T.1
```

```
In [2]: #Aim :To perform simple linear regression
```

```
In [3]: import pandas as pd
```

```
In [4]: import os
```

```
In [5]: os.getcwd
```

```
Out[5]: <function nt.getcwd()>
```

```
In [6]: df=pd.read_csv("C:\\Users\\DELL\\OneDrive\\Documents\\Salary.csv")
```

In [7]: df

Out[7]:

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

```
In [8]: df.head()
```

```
Out[8]:
```

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

```
In [9]: df.tail()
```

```
Out[9]:
```

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

```
In [10]: df.shape
```

```
Out[10]: (35, 2)
```

```
In [11]: df.size
```

```
Out[11]: 70
```

```
In [12]: df.ndim
```

```
Out[12]: 2
```

```
In [14]: 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
```

```
df.
```

```
In [15]: df.describe
```

```
Out[15]: <bound method NDFrame.describe of      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
30        11.2     127345
31        11.5     126756
32        12.3     128765
33        12.9     135675
34        13.5     139465>
```

```
In [16]: df.isnull().sum()
```

```
Out[16]: YearsExperience    0
Salary                    0
dtype: int64
```

```
In [17]: x=df.iloc[:, -1].values
y=df.iloc[:, -1].values
```

```
In [18]: 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 [19]: 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 [20]: import matplotlib.pyplot as plt
import seaborn as sns
import numpy as np
```

```
In [21]: 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=
```

```
In [22]: print(x_train)
```

```
[[ 7.1]
 [ 4.9]
 [10.5]
 [ 6. ]
 [ 4. ]
 [ 3.2]
 [ 5.1]
 [ 3.7]
 [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 [23]: print(y_train)
```

```
[ 98273  67938 121872  93940  56957  64445  66029  57189 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 [24]: print(y_test)
```

```
[116969  57081 109431]
```

```
In [25]: from sklearn.linear_model import LinearRegression
lr = LinearRegression()
lr.fit(x_train,y_train)
```

```
Out[25]: LinearRegression()
```

```
In [26]: m=lr.coef_
```

```
In [27]: print("coefficient :",m)
coefficient : [8627.16113876]
```

```
In [28]: #Assigning Y-intercept to a
C=lr.intercept_
```

```
In [29]: print("Intercept :",C)
Intercept : 29441.35981114542
```

```
In [30]: lr.score(x_test,y_test)*100
```

```
Out[30]: 94.589314712631
```

```
In [ ]:
```