

```
In [28]: import pandas as pd
import numpy as np
import seaborn as sns
```

```
In [29]: df=pd.read_csv(r"D:\NIT\DECEMBER\11 DEC (SLR(SIMPLE))\11th - Regression model\S
```

```
In [30]: df.head()
```

```
Out[30]:
```

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

```
In [31]: df.shape
```

```
Out[31]: (30, 2)
```

mean

```
In [33]: df.mean()
```

```
Out[33]: YearsExperience    5.313333
Salary                76003.000000
dtype: float64
```

```
In [34]: df["Salary"].mean()
```

```
Out[34]: 76003.0
```

median

```
In [36]: df.median() # this will give median of entire dataframe
```

```
Out[36]: YearsExperience    4.7
Salary                65237.0
dtype: float64
```

```
In [37]: df['Salary'].median() # this will give us median of that particular column
```

```
Out[37]: 65237.0
```

Mode

```
In [39]: df['Salary'].mode() # this will give us mode of that particular column
```

```
Out[39]: 0      37731
1      39343
2      39891
3      43525
4      46205
5      54445
6      55794
7      56642
8      56957
9      57081
10     57189
11     60150
12     61111
13     63218
14     64445
15     66029
16     67938
17     81363
18     83088
19     91738
20     93940
21     98273
22    101302
23    105582
24    109431
25    112635
26    113812
27    116969
28    121872
29    122391
Name: Salary, dtype: int64
```

Variance

```
In [41]: df.var() # this will give variance of entire dataframe
```

```
Out[41]: YearsExperience    8.053609e+00
Salary                    7.515510e+08
dtype: float64
```

```
In [42]: df['Salary'].var() # this will give us variance of that particular column
```

```
Out[42]: 751550960.4137931
```

Standard deviation

```
In [44]: df.std() # this will give standard deviation of entire dataframe
```

```
Out[44]: YearsExperience    2.837888
Salary                    27414.429785
dtype: float64
```

```
In [45]: df['Salary'].std() # this will give us standard deviation of that particular col
```

Out[45]: 27414.4297845823

Correlation

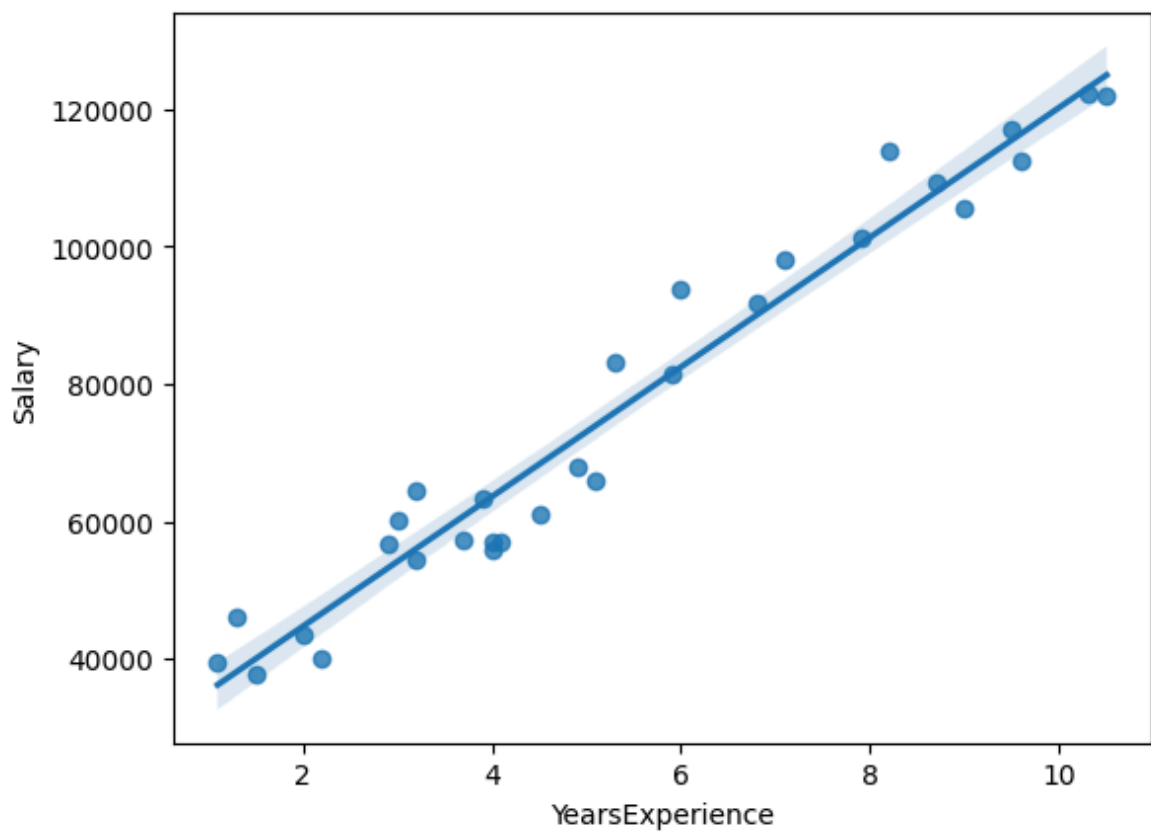
```
In [47]: df.corr() # this will give correlation of entire dataframe
```

```
Out[47]:
```

	YearsExperience	Salary
YearsExperience	1.000000	0.978242
Salary	0.978242	1.000000

```
In [57]: sns.regplot(data=df,x='YearsExperience',y='Salary')
```

```
Out[57]: <Axes: xlabel='YearsExperience', ylabel='Salary'>
```



```
In [49]: df
```

Out[49]:

	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

Skewness

```
In [51]: df.skew() # this will give skewness of entire dataframe
```

```
Out[51]: YearsExperience    0.37956  
Salary          0.35412  
dtype: float64
```

```
In [52]: df['Salary'].skew() # this will give us skewness of that particular column
```

```
Out[52]: 0.35411967922959153
```

```
In [53]: df.describe()
```

```
Out[53]:
```

	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 [ ]:
```