In [2]:

```
import pandas as pd
import matplotlib.pyplot as plt
from numpy.polynomial.polynomial import polyfit
import statsmodels.api as sm
import numpy as np
data=pd.read_csv('Salary_Data.csv')
data.head()
```

executed in 3.29s, finished 09:12:57 2021-11-26

Out[2]:

	YearsExperience	Salary
0	1.1	39343.0
1	1.3	46205.0
2	1.5	37731.0
3	2.0	43525.0
4	2.2	39891.0

In [3]:

```
data.info()
```

executed in 48ms, finished 08:53:10 2021-11-13

```
<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     float64
dtypes: float64(2)
memory usage: 608.0 bytes
```

In [4]:

```
data.corr()
```

executed in 24ms, finished 08:55:18 2021-11-13

Out[4]:

	YearsExperience	Salary
YearsExperience	1.000000	0.978242
Salary	0.978242	1.000000

In [5]:

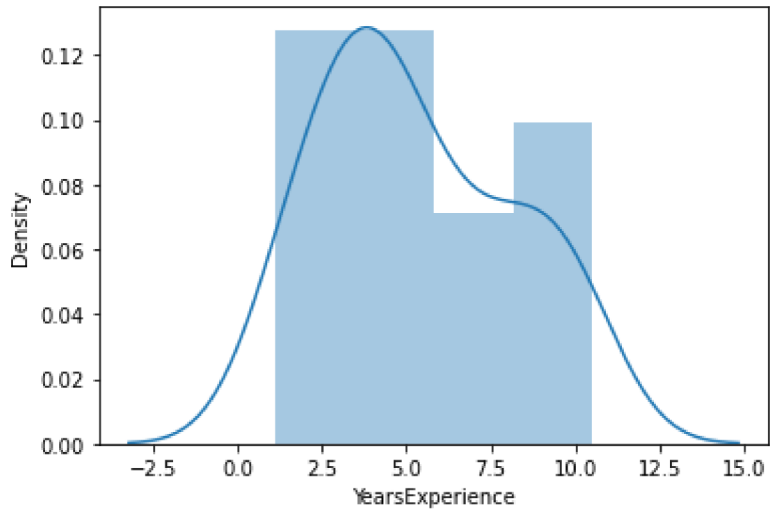
```
import seaborn as sns
sns.distplot(data['YearsExperience'])
```

executed in 2.48s, finished 09:16:52 2021-11-13

```
C:\Users\win\anaconda3\lib\site-packages\seaborn\distributions.py:2557: Futu
reWarning: `distplot` is a deprecated function and will be removed in a futu
re version. Please adapt your code to use either `displot` (a figure-level f
unction with similar flexibility) or `histplot` (an axes-level function for
histograms).
  warnings.warn(msg, FutureWarning)
```

Out[5]:

```
<AxesSubplot:xlabel='YearsExperience', ylabel='Density'>
```



In [8]:

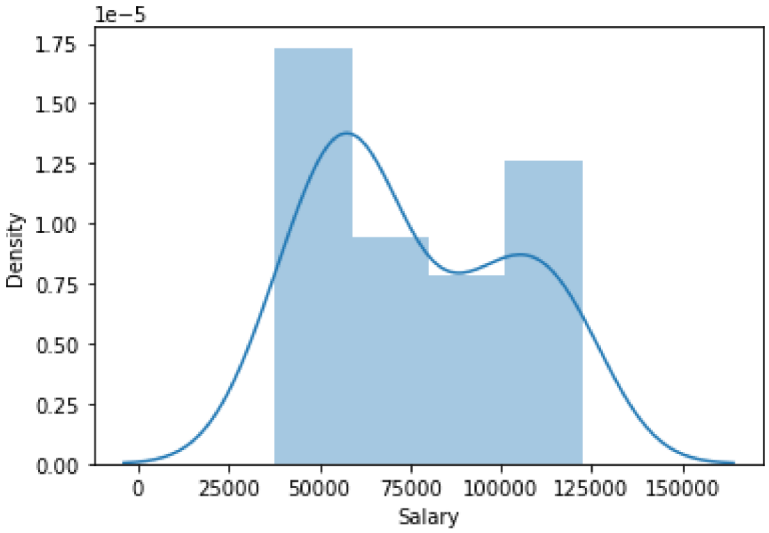
```
import seaborn as sns
sns.distplot(data['Salary'])
```

executed in 424ms, finished 09:19:16 2021-11-13

```
C:\Users\win\anaconda3\lib\site-packages\seaborn\distributions.py:2557: Futu
reWarning: `distplot` is a deprecated function and will be removed in a futu
re version. Please adapt your code to use either `displot` (a figure-level f
unction with similar flexibility) or `histplot` (an axes-level function for
histograms).
  warnings.warn(msg, FutureWarning)
```

Out[8]:

```
<AxesSubplot:xlabel='Salary', ylabel='Density'>
```



In [4]:

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

executed in 17ms, finished 09:14:13 2021-11-26

In [5]:

```
b, m = polyfit(x, y, 1)
plt.scatter(x, y)
plt.plot(x, y, '.')
plt.plot(x, b + m * x, '-')
plt.title('Scatter plot Salary Hike')
plt.xlabel('Years of Experience')
plt.ylabel('Salary Hike')
plt.show()
```

executed in 369ms, finished 09:14:30 2021-11-26

In [13]:

```
model.params
```

executed in 16ms, finished 09:22:26 2021-11-13

Out[13]:

```
Intercept   -2.383161
Salary       0.000101
dtype: float64
```

In [17]:

```
print(model.tvalues,'\n',model.pvalues)
```

executed in 16ms, finished 09:24:22 2021-11-13

```
Intercept    -7.281283
Salary       24.950094
dtype: float64
 Intercept    6.300123e-08
Salary       1.143068e-20
dtype: float64
```

In [18]:

```
(model.rsquared,model.rsquared_adj)
```

executed in 32ms, finished 09:25:29 2021-11-13

Out[18]:

```
(0.9569566641435086, 0.9554194021486339)
```

In [6]:

```
model = sm.OLS(y, x).fit()
predictions = model.predict(x)
```

executed in 19ms, finished 09:14:58 2021-11-26

In [7]:

```
model.summary()
```

executed in 53ms, finished 09:15:10 2021-11-26

Out[7]:

OLS Regression Results

Dep. Variable:	Salary	**R-squared (uncentered):**	0.973
Model:	OLS	**Adj. R-squared (uncentered):**	0.972
Method:	Least Squares	**F-statistic:**	1048.
Date:	Fri, 26 Nov 2021	**Prob (F-statistic):**	2.56e-24
Time:	09:15:10	**Log-Likelihood:**	-327.28
No. Observations:	30	**AIC:**	656.6
Df Residuals:	29	**BIC:**	658.0
Df Model:	1		
Covariance Type:	nonrobust		

	coef	**std err**	**t**	**P>|t|**	**[0.025**	**0.975]**
YearsExperience	1.325e+04	409.401	32.376	0.000	1.24e+04	1.41e+04

Omnibus:	0.610	**Durbin-Watson:**	0.323
Prob(Omnibus):	0.737	**Jarque-Bera (JB):**	0.671
Skew:	-0.121	**Prob(JB):**	0.715
Kurtosis:	2.308	**Cond. No.**	1.00

Notes:
[1] R^2 is computed without centering (uncentered) since the model does not contain a constant.
[2] Standard Errors assume that the covariance matrix of the errors is correctly specified.

In [9]:

```
x_log = np.log(data['YearsExperience'])
```

executed in 16ms, finished 09:15:39 2021-11-26

In [10]:

```
model = sm.OLS(y, x_log).fit()
predictions = model.predict(x_log)
```

executed in 15ms, finished 09:15:51 2021-11-26

In [11]:

```
model.summary()
```

executed in 46ms, finished 09:16:02 2021-11-26

Out[11]:

OLS Regression Results

Dep. Variable:	Salary	**R-squared (uncentered):**	0.979
Model:	OLS	**Adj. R-squared (uncentered):**	0.978
Method:	Least Squares	**F-statistic:**	1338.
Date:	Fri, 26 Nov 2021	**Prob (F-statistic):**	8.06e-26
Time:	09:16:02	**Log-Likelihood:**	-323.70
No. Observations:	30	**AIC:**	649.4
Df Residuals:	29	**BIC:**	650.8
Df Model:	1		
Covariance Type:	nonrobust		

	coef	**std err**	**t**	**P>|t|**	**[0.025**	**0.975]**
YearsExperience	4.909e+04	1341.796	36.583	0.000	4.63e+04	5.18e+04

Omnibus:	10.249	**Durbin-Watson:**	0.421
Prob(Omnibus):	0.006	**Jarque-Bera (JB):**	8.950
Skew:	1.106	**Prob(JB):**	0.0114
Kurtosis:	4.507	**Cond. No.**	1.00

Notes:
[1] R^2 is computed without centering (uncentered) since the model does not contain a constant.
[2] Standard Errors assume that the covariance matrix of the errors is correctly specified.

In [12]:

```
y_log = np.log(data['Salary'])
```

executed in 7ms, finished 09:16:25 2021-11-26

In [13]:

```
model = sm.OLS(y_log, x).fit()
predictions = model.predict(x)
```

executed in 15ms, finished 09:16:36 2021-11-26

In [14]:

```
model.summary()
```

executed in 37ms, finished 09:16:48 2021-11-26

Out[14]:

OLS Regression Results

Dep. Variable:	Salary	**R-squared (uncentered):**	0.809
Model:	OLS	**Adj. R-squared (uncentered):**	0.802
Method:	Least Squares	**F-statistic:**	122.8
Date:	Fri, 26 Nov 2021	**Prob (F-statistic):**	6.09e-12
Time:	09:16:48	**Log-Likelihood:**	-90.160
No. Observations:	30	**AIC:**	182.3
Df Residuals:	29	**BIC:**	183.7
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
YearsExperience	1.6755	0.151	11.083	0.000	1.366	1.985

Omnibus:	3.609	**Durbin-Watson:**	0.016
Prob(Omnibus):	0.165	**Jarque-Bera (JB):**	2.045
Skew:	-0.389	**Prob(JB):**	0.360
Kurtosis:	1.985	**Cond. No.**	1.00

Notes:
[1] R^2 is computed without centering (uncentered) since the model does not contain a constant.
[2] Standard Errors assume that the covariance matrix of the errors is correctly specified.

In [15]:

```
model = sm.OLS(y_log, x_log).fit()
predictions = model.predict(x_log)
```

executed in 11ms, finished 09:17:07 2021-11-26

In [16]:

```
model.summary()
```

executed in 43ms, finished 09:17:21 2021-11-26

Out[16]:

OLS Regression Results

Dep. Variable:	Salary	**R-squared (uncentered):**	0.878
Model:	OLS	**Adj. R-squared (uncentered):**	0.874
Method:	Least Squares	**F-statistic:**	209.1
Date:	Fri, 26 Nov 2021	**Prob (F-statistic):**	8.60e-15
Time:	09:17:21	**Log-Likelihood:**	-83.410
No. Observations:	30	**AIC:**	168.8
Df Residuals:	29	**BIC:**	170.2
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
YearsExperience	6.4461	0.446	14.461	0.000	5.534	7.358

Omnibus:	2.550	**Durbin-Watson:**	0.026
Prob(Omnibus):	0.279	**Jarque-Bera (JB):**	2.163
Skew:	0.640	**Prob(JB):**	0.339
Kurtosis:	2.697	**Cond. No.**	1.00

Notes:
[1] R^2 is computed without centering (uncentered) since the model does not contain a constant.
[2] Standard Errors assume that the covariance matrix of the errors is correctly specified.

In [18]:

```
x_sqrt = np.sqrt(data['YearsExperience'])
```

executed in 20ms, finished 09:17:44 2021-11-26

In [19]:

```
model = sm.OLS(y, x_sqrt).fit()
predictions = model.predict(x_sqrt)
```

executed in 14ms, finished 09:18:05 2021-11-26

In [20]:

```
model.summary()
```

executed in 49ms, finished 09:18:24 2021-11-26

Out[20]:

OLS Regression Results

Dep. Variable:	Salary	**R-squared (uncentered):**	0.989
Model:	OLS	**Adj. R-squared (uncentered):**	0.989
Method:	Least Squares	**F-statistic:**	2697.
Date:	Fri, 26 Nov 2021	**Prob (F-statistic):**	3.62e-30
Time:	09:18:24	**Log-Likelihood:**	-313.35
No. Observations:	30	**AIC:**	628.7
Df Residuals:	29	**BIC:**	630.1
Df Model:	1		
Covariance Type:	nonrobust		

	coef	**std err**	**t**	**P>|t|**	**[0.025**	**0.975]**
YearsExperience	3.48e+04	670.056	51.932	0.000	3.34e+04	3.62e+04

Omnibus:	5.654	**Durbin-Watson:**	0.734
Prob(Omnibus):	0.059	**Jarque-Bera (JB):**	1.849
Skew:	-0.040	**Prob(JB):**	0.397
Kurtosis:	1.786	**Cond. No.**	1.00

Notes:
[1] R^2 is computed without centering (uncentered) since the model does not contain a constant.
[2] Standard Errors assume that the covariance matrix of the errors is correctly specified.

In [21]:

```
y_sqrt = np.sqrt(data['Salary'])
```

executed in 17ms, finished 09:18:46 2021-11-26

In [22]:

```
model = sm.OLS(y_sqrt, x).fit()
predictions = model.predict(x)
```

executed in 10ms, finished 09:19:02 2021-11-26

In [23]:

```
model.summary()
```

executed in 62ms, finished 09:19:12 2021-11-26

Out[23]:

OLS Regression Results

Dep. Variable:	Salary	**R-squared (uncentered):**	0.906
Model:	OLS	**Adj. R-squared (uncentered):**	0.902
Method:	Least Squares	**F-statistic:**	278.1
Date:	Fri, 26 Nov 2021	**Prob (F-statistic):**	2.12e-16
Time:	09:19:12	**Log-Likelihood:**	-175.75
No. Observations:	30	**AIC:**	353.5
Df Residuals:	29	**BIC:**	354.9
Df Model:	1		
Covariance Type:	nonrobust		

	coef	**std err**	**t**	**P>|t|**	**[0.025**	**0.975]**
YearsExperience	43.7142	2.621	16.676	0.000	38.353	49.076

Omnibus:	2.437	**Durbin-Watson:**	0.046
Prob(Omnibus):	0.296	**Jarque-Bera (JB):**	1.859
Skew:	-0.445	**Prob(JB):**	0.395
Kurtosis:	2.167	**Cond. No.**	1.00

Notes:
[1] R^2 is computed without centering (uncentered) since the model does not contain a constant.
[2] Standard Errors assume that the covariance matrix of the errors is correctly specified.

In [24]:

```
model = sm.OLS(y_sqrt, x_sqrt).fit()
predictions = model.predict(x_sqrt)
```

executed in 7ms, finished 09:19:35 2021-11-26

In [25]:

```
model.summary()
```

executed in 40ms, finished 09:19:44 2021-11-26

Out[25]:

OLS Regression Results

Dep. Variable:	Salary	**R-squared (uncentered):**	0.988
Model:	OLS	**Adj. R-squared (uncentered):**	0.987
Method:	Least Squares	**F-statistic:**	2338.
Date:	Fri, 26 Nov 2021	**Prob (F-statistic):**	2.81e-29
Time:	09:19:44	**Log-Likelihood:**	-145.12
No. Observations:	30	**AIC:**	292.2
Df Residuals:	29	**BIC:**	293.6
Df Model:	1		
Covariance Type:	nonrobust		

	coef	**std err**	**t**	**P>|t|**	**[0.025**	**0.975]**
YearsExperience	118.8652	2.458	48.352	0.000	113.837	123.893

Omnibus:	2.762	**Durbin-Watson:**	0.231
Prob(Omnibus):	0.251	**Jarque-Bera (JB):**	2.152
Skew:	0.653	**Prob(JB):**	0.341
Kurtosis:	2.881	**Cond. No.**	1.00

Notes:
[1] R^2 is computed without centering (uncentered) since the model does not contain a constant.
[2] Standard Errors assume that the covariance matrix of the errors is correctly specified.