```
Python 3.6.1 |Anaconda 4.4.0 (64-bit)| (default, May 11 2017, 13:25:24) [MSC v.1900 64 bit
(AMD64)]
Type "copyright", "credits" or "license" for more information.

IPython 5.3.0 -- An enhanced Interactive Python.
?         -> Introduction and overview of IPython's features.
%quickref -> Quick reference.
help      -> Python's own help system.
object?   -> Details about 'object', use 'object??' for extra details.

In [1]: """
   ...: Created on Sat Aug 26 14:33:16 2017
   ...:
   ...: @author: Harshil
   ...: """
   ...:
   ...: import numpy as np
   ...: import matplotlib.pyplot as plt
   ...: import pandas as pd
   ...:
   ...: dataset=pd.read_csv('50_Startups.csv')
   ...: x=dataset.iloc[:, :-1].values
   ...: y=dataset.iloc[:, 4].values
   ...:
   ...: from sklearn.preprocessing import LabelEncoder, OneHotEncoder
   ...: labelencoder_X = LabelEncoder()
   ...: x[:, 3] = labelencoder_X.fit_transform(x[:, 3])
   ...: onehotencoder = OneHotEncoder(categorical_features = [3])
   ...: x= onehotencoder.fit_transform(x).toarray()
   ...:
   ...: #avoiding dummy variable trap
   ...: x=x[:, 1:]
   ...:
   ...: from sklearn.cross_validation import train_test_split
   ...: x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.2,random_state=0)
   ...:
   ...: from sklearn.linear_model import LinearRegression
   ...: regressor=LinearRegression()
   ...: regressor.fit(x_train,y_train)
   ...:
   ...: y_pred=regressor.predict(x_test)
   ...:
   ...: import statsmodels.formula.api as sm
   ...: x=np.append(arr=np.ones((50,1)).astype(int),values=x,axis=1 )
   ...: x_opt=x[:,[0,1,2,3,4,5]]
   ...: regressor_OLS=sm.OLS(endog=y,exog=x_opt).fit()
   ...: regressor_OLS.summary()
   ...: x_opt=x[:,[0,1,3,4,5]]
   ...: regressor_OLS=sm.OLS(endog=y,exog=x_opt).fit()
   ...: regressor_OLS.summary()
   ...: x_opt=x[:,[0,3,4,5]]
   ...: regressor_OLS=sm.OLS(endog=y,exog=x_opt).fit()
   ...: regressor_OLS.summary()
   ...: x_opt=x[:,[0,3,5]]
   ...: regressor_OLS=sm.OLS(endog=y,exog=x_opt).fit()
   ...: regressor_OLS.summary()
   ...: x_opt=x[:,[0,3]]
   ...: regressor_OLS=sm.OLS(endog=y,exog=x_opt).fit()
   ...: regressor_OLS.summary()
C:\Users\Harshil\Anaconda3\lib\site-packages\sklearn\cross_validation.py:44:
DeprecationWarning: This module was deprecated in version 0.18 in favor of the model_selection
module into which all the refactored classes and functions are moved. Also note that the
interface of the new CV iterators are different from that of this module. This module will be
removed in 0.20.
  "This module will be removed in 0.20.", DeprecationWarning)
Out[1]:
<class 'statsmodels.iolib.summary.Summary'>
"""
                            OLS Regression Results
```

```
==============================================================================
Dep. Variable:                      y   R-squared:                       0.947
Model:                            OLS   Adj. R-squared:                  0.945
Method:                 Least Squares   F-statistic:                     849.8
Date:                Sat, 26 Aug 2017   Prob (F-statistic):           3.50e-32
Time:                        16:20:36   Log-Likelihood:                -527.44
No. Observations:                  50   AIC:                             1059.
Df Residuals:                      48   BIC:                             1063.
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const       4.903e+04   2537.897     19.320      0.000    4.39e+04    5.41e+04
x1             0.8543      0.029     29.151      0.000       0.795       0.913
==============================================================================
Omnibus:                       13.727   Durbin-Watson:                   1.116
Prob(Omnibus):                  0.001   Jarque-Bera (JB):               18.536
Skew:                          -0.911   Prob(JB):                     9.44e-05
Kurtosis:                       5.361   Cond. No.                     1.65e+05
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 1.65e+05. This might indicate that there are
strong multicollinearity or other numerical problems.
"""
```

In [2]: