```
In [100]: import pandas as pd
     ...:
     ...: df=pd.read_csv('G:\Data Analysis\output.csv')
     ...: df=df.dropna()
     ...: X = df.iloc[:,[8,11,14,17,22,23,24,25,26,27,28,29,30,31,32,33,34,35,36,37]].values
     ...: y = df.iloc[:, 1].values

In [101]: from sklearn.model_selection import train_test_split
     ...: X_train, X_test, y_train, y_test = train_test_split(X, y, test_size = 0.4,
random_state = 0)
     ...:
     ...:
     ...: from sklearn.linear_model import LinearRegression
     ...: regressor = LinearRegression()
     ...: regressor.fit(X_train, y_train)
Out[101]:
LinearRegression(copy_X=True, fit_intercept=True, n_jobs=None,
         normalize=False)

In [102]: import statsmodels.formula.api as sm
     ...:
     ...: model1=sm.OLS(y_train,X_train)
     ...: result=model1.fit()
     ...: print(result.summary())
                            OLS Regression Results
==============================================================================
Dep. Variable:                      y   R-squared:                       0.765
Model:                            OLS   Adj. R-squared:                  0.763
Method:                 Least Squares   F-statistic:                     314.6
Date:                Sat, 06 Jul 2019   Prob (F-statistic):               0.00
Time:                        19:26:22   Log-Likelihood:                -3846.7
No. Observations:                1953   AIC:                             7733.
Df Residuals:                    1933   BIC:                             7845.
Df Model:                          20
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
x1             0.0101      0.005      1.859      0.063      -0.001       0.021
x2             0.0121      0.005      2.444      0.015       0.002       0.022
x3            -0.8680      0.122     -7.140      0.000      -1.106      -0.630
x4             0.0185      0.005      3.814      0.000       0.009       0.028
x5            -0.1383      0.047     -2.940      0.003      -0.231      -0.046
x6             0.0637      0.060      1.070      0.285      -0.053       0.181
x7             0.0837      0.049      1.696      0.090      -0.013       0.180
x8            -0.1006      0.043     -2.356      0.019      -0.184      -0.017
x9            -0.0791      0.052     -1.535      0.125      -0.180       0.022
x10            0.0012      0.000      2.669      0.008       0.000       0.002
x11           -0.0002      0.001     -0.290      0.771      -0.001       0.001
x12            0.0021      0.000      5.322      0.000       0.001       0.003
x13            0.2092      0.097      2.147      0.032       0.018       0.400
x14            0.0005      0.000      2.064      0.039    2.71e-05       0.001
x15           -0.0005      0.000     -1.568      0.117      -0.001       0.000
x16           -0.0015      0.000     -6.329      0.000      -0.002      -0.001
x17         8.215e-05      0.000      0.176      0.860      -0.001       0.001
x18           -0.0017      0.000     -3.627      0.000      -0.003      -0.001
x19           -0.0002      0.000     -0.493      0.622      -0.001       0.001
```

```
x20            0.0014      0.001      1.179      0.239      -0.001       0.004
==============================================================================
Omnibus:                     1949.242   Durbin-Watson:                   2.017
Prob(Omnibus):                  0.000   Jarque-Bera (JB):           150101.647
Skew:                           4.652   Prob(JB):                         0.00
Kurtosis:                      44.928   Cond. No.                     3.02e+03
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 3.02e+03. This might indicate that there are
strong multicollinearity or other numerical problems.

In [103]: c=0
     ...: y_pred = regressor.predict(X_test)
     ...: for i in range(len(y_pred)):
     ...:     y_pred[i]=(y_pred[i]<=y_test[i]+1.5  and y_pred[i]>=y_test[i]-1.5)
     ...:     if(y_pred[i]):
     ...:         c+=1
     ...:
     ...: acc=float(c/len(y_test))

In [104]: def backwardElimination(x, sl):
     ...:     numVars = len(x[0])
     ...:     for i in range(0, numVars):
     ...:         regressor_OLS = sm.OLS(y, x).fit()
     ...:         maxVar = max(regressor_OLS.pvalues).astype(float)
     ...:         if maxVar > sl:
     ...:             for j in range(0, numVars - i):
     ...:                 if (regressor_OLS.pvalues[j].astype(float) == maxVar):
     ...:                     x = np.delete(x, j, 1)
     ...:     regressor_OLS.summary()
     ...:     return x
     ...:
     ...:
     ...: SL = 0.05
     ...: import numpy as np
     ...: X=np.append(arr=np.ones((3256,1)).astype(int),values =X ,axis=1)
     ...: X_opt = X[:,[0,1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20]]
     ...: X_Modeled = backwardElimination(X_opt, SL)

In [105]: from sklearn.model_selection import train_test_split
     ...: X_train, X_test, y_train, y_test = train_test_split(X_Modeled, y, test_size = 0.4,
random_state = 0)
     ...:
     ...: from sklearn.linear_model import LinearRegression
     ...: regressor = LinearRegression()
     ...: regressor.fit(X_train, y_train)
Out[105]:
LinearRegression(copy_X=True, fit_intercept=True, n_jobs=None,
         normalize=False)

In [106]: model1=sm.OLS(y_train,X_train)
     ...: result=model1.fit()
     ...: print(result.summary())
                            OLS Regression Results
==============================================================================
Dep. Variable:                      y   R-squared:                       0.764
Model:                            OLS   Adj. R-squared:                  0.762
```

```
Method:                 Least Squares   F-statistic:                     482.7
Date:                Sat, 06 Jul 2019   Prob (F-statistic):               0.00
Time:                        19:26:41   Log-Likelihood:                -3851.6
No. Observations:                1953   AIC:                             7729.
Df Residuals:                    1940   BIC:                             7802.
Df Model:                          13
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
x1             0.0176      0.004      4.361      0.000       0.010       0.026
x2            -0.8252      0.118     -6.983      0.000      -1.057      -0.593
x3             0.0205      0.005      4.431      0.000       0.011       0.030
x4            -0.1378      0.046     -3.004      0.003      -0.228      -0.048
x5             0.0507      0.050      1.013      0.311      -0.047       0.149
x6            -0.1090      0.042     -2.580      0.010      -0.192      -0.026
x7             0.0013      0.000      3.336      0.001       0.001       0.002
x8             0.0021      0.000      5.796      0.000       0.001       0.003
x9             0.2053      0.095      2.169      0.030       0.020       0.391
x10            0.0005      0.000      2.181      0.029     5.2e-05       0.001
x11           -0.0006      0.000     -1.949      0.051      -0.001    3.44e-06
x12           -0.0015      0.000     -6.386      0.000      -0.002      -0.001
x13           -0.0017      0.000     -3.624      0.000      -0.003      -0.001
==============================================================================
Omnibus:                     1948.016   Durbin-Watson:                   2.022
Prob(Omnibus):                  0.000   Jarque-Bera (JB):           149193.173
Skew:                           4.649   Prob(JB):                         0.00
Kurtosis:                      44.796   Cond. No.                     2.50e+03
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 2.5e+03. This might indicate that there are
strong multicollinearity or other numerical problems.
```

```
In [107]: c=0
     ...: y_pred = regressor.predict(X_test)
     ...: for i in range(len(y_pred)):
     ...:     y_pred[i]=(y_pred[i]<=y_test[i]+2 and y_pred[i]>=y_test[i]-2)
     ...:     if(y_pred[i]):
     ...:         c+=1
     ...:
     ...: accAfterBackElimination=float(c/len(y_test))

In [108]: c=0
     ...: y_pred = regressor.predict(X_test)
     ...: for i in range(len(y_pred)):
     ...:     y_pred[i]=(y_pred[i]<=y_test[i]+2 and y_pred[i]>=y_test[i]-2)
     ...:     if(y_pred[i]):
     ...:         c+=1
     ...:
     ...: accAfterBackElimination=float(c/len(y_test))

In [109]: accAfterBackElimination
Out[109]: 0.8948580199539524

In [110]:
```