```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import statsmodels.api as sm



df = pd.read_csv('MLR-Feature-Elimination.csv')
```

First we import all the necessary libraries and import the data on which we are going to work on.

```
y=df["c52"]
X = df.drop(columns=["c1", "c2", "c241","c52"])
X = sm.add_constant(X)
```

We define y as c52 and x as all the other columns excluding c1, c2, c241 and c52.

```
drop = []
r2 = []
mse_value = []
```

Here we create three arrays which store dropped variables, R^2 after dropping variables and Mean square Errors. We use a while loop to drop these variables and obviosly exclude the constant term .

```
model = sm.OLS(y, X).fit()
print(model.summary())


while len(X.columns) > 1:
    model = sm.OLS(y, X).fit()
    max_p_value = model.pvalues[1:].max()
    if max_p_value > 0.05:
        drop = model.pvalues[1:].idxmax()
        X = X.drop(columns=[dropped_var])
        model = sm.OLS(y, X).fit()
        drop.append(dropped_var)
        r2.append(model.rsquared)
        mse_value.append(model.mse_resid)
    else:
        break



summary_table = pd.DataFrame({
    "Dropped Variable": dropped_vars,
    "R2": r2_values,
    "MSE": mse_values
})
print(summary_table)
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                    c52   R-squared:                       0.785
Model:                            OLS   Adj. R-squared:                  0.777
Method:                 Least Squares   F-statistic:                     97.27
Date:                Sat, 02 Sep 2023   Prob (F-statistic):          1.11e-299
Time:                        19:37:45   Log-Likelihood:                -1479.4
No. Observations:                1025   AIC:                             3035.
Df Residuals:                     987   BIC:                             3222.
Df Model:                          37
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const       -136.4882    104.748     -1.303      0.193    -342.042      69.065
c26            0.3634      0.049      7.442      0.000       0.268       0.459
c27           -0.1911      0.895     -0.214      0.831      -1.948       1.565
c28            0.2266      0.044      5.150      0.000       0.140       0.313
c29           -0.4454      0.049     -9.115      0.000      -0.541      -0.349
```

```
c30            3.4632      0.458      7.568      0.000       2.565       4.361
c31            0.2667      0.035      7.699      0.000       0.199       0.335
c32            0.1781      0.199      0.895      0.371      -0.212       0.569
c33           -0.6545      0.464     -1.412      0.158      -1.564       0.255
c39           12.9984      1.470      8.845      0.000      10.114      15.882
c139          -0.8439      0.225     -3.745      0.000      -1.286      -0.402
c142           0.0454      0.067      0.682      0.495      -0.085       0.176
c143          -0.1537      0.039     -3.956      0.000      -0.230      -0.077
c155          -0.0342      0.013     -2.684      0.007      -0.059      -0.009
c157           0.2501      0.041      6.100      0.000       0.170       0.331
c158           0.2836      0.023     12.121      0.000       0.238       0.329
c160           0.0040      0.002      2.206      0.028       0.000       0.008
c161           0.0105      0.001      9.632      0.000       0.008       0.013
c162           0.0027      0.002      1.649      0.099      -0.001       0.006
c163           0.0081      0.002      3.724      0.000       0.004       0.012
c7             0.3236      0.292      1.108      0.268      -0.250       0.897
c8            -0.4465      0.137     -3.257      0.001      -0.716      -0.177
c9            -0.6863      0.075     -9.097      0.000      -0.834      -0.538
c10            8.8046      1.541      5.715      0.000       5.781      11.828
c11           -0.1706      0.042     -4.044      0.000      -0.253      -0.088
c12           -0.3028      0.109     -2.765      0.006      -0.518      -0.088
c13            0.0757      0.052      1.455      0.146      -0.026       0.178
c15           -0.4255      0.058     -7.336      0.000      -0.539      -0.312
c16           -0.5046      0.103     -4.919      0.000      -0.706      -0.303
c17           -0.0792      0.021     -3.707      0.000      -0.121      -0.037
c19            0.3822      0.218      1.756      0.079      -0.045       0.809
c20            0.2279      0.042      5.449      0.000       0.146       0.310
c21           -0.1647      0.049     -3.329      0.001      -0.262      -0.068
c22           -0.1258      0.036     -3.455      0.001      -0.197      -0.054
c23           -0.3301      0.048     -6.848      0.000      -0.425      -0.235
c34           -0.5123      1.765     -0.290      0.772      -3.976       2.951
c35            6.3277      1.616      3.917      0.000       3.157       9.498
c36           -1.8280     90.385     -0.020      0.984    -179.197     175.541
==============================================================================
Omnibus:                       40.476   Durbin-Watson:                   0.546
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              113.185
Skew:                          -0.049   Prob(JB):                     2.64e-25
Kurtosis:                       4.625   Cond. No.                     3.48e+06
```

```
final_model = sm.OLS(y, X).fit()
print(final_model.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                    c52   R-squared:                       0.782
Model:                            OLS   Adj. R-squared:                  0.776
Method:                 Least Squares   F-statistic:                     127.9
Date:                Wed, 30 Aug 2023   Prob (F-statistic):          1.30e-306
Time:                        00:01:18   Log-Likelihood:                -1485.0
No. Observations:                1025   AIC:                             3028.
Df Residuals:                     996   BIC:                             3171.
Df Model:                          28
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const       -125.8111     17.336     -7.257      0.000    -159.831     -91.791
c26            0.3663      0.047      7.734      0.000       0.273       0.459
c28            0.2515      0.034      7.454      0.000       0.185       0.318
c29           -0.4401      0.047     -9.342      0.000      -0.532      -0.348
c30            3.4022      0.447      7.610      0.000       2.525       4.280
c31            0.2953      0.022     13.236      0.000       0.252       0.339
c33           -0.2699      0.082     -3.293      0.001      -0.431      -0.109
c39           14.6285      1.314     11.134      0.000      12.050      17.207
c139          -0.4288      0.045     -9.546      0.000      -0.517      -0.341
c143          -0.1502      0.037     -4.063      0.000      -0.223      -0.078
c155          -0.0418      0.012     -3.576      0.000      -0.065      -0.019
c157           0.2573      0.039      6.587      0.000       0.181       0.334
c158           0.2832      0.023     12.429      0.000       0.239       0.328
c160           0.0039      0.002      2.152      0.032       0.000       0.007
c161           0.0110      0.001     10.690      0.000       0.009       0.013
c163           0.0087      0.002      4.141      0.000       0.005       0.013
c8            -0.4440      0.131     -3.381      0.001      -0.702      -0.186
c9            -0.6920      0.066    -10.507      0.000      -0.821      -0.563
c10            8.9184      1.518      5.876      0.000       5.940      11.897
c11           -0.1655      0.041     -4.004      0.000      -0.247      -0.084
c12           -0.3099      0.105     -2.939      0.003      -0.517      -0.103
c15           -0.4338      0.052     -8.278      0.000      -0.537      -0.331
c16           -0.4095      0.087     -4.699      0.000      -0.580      -0.238
c17           -0.0800      0.021     -3.797      0.000      -0.121      -0.039
c20            0.2353      0.040      5.827      0.000       0.156       0.315
c21           -0.1557      0.048     -3.213      0.001      -0.251      -0.061
```

```
c22           -0.1296      0.035     -3.717      0.000      -0.198      -0.061
c23           -0.3024      0.045     -6.746      0.000      -0.390      -0.214
c35            6.4165      1.516      4.232      0.000       3.441       9.392
==============================================================================
Omnibus:                       53.080   Durbin-Watson:                   0.546
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              178.982
Skew:                          -0.077   Prob(JB):                     1.36e-39
Kurtosis:                       5.041   Cond. No.                     5.63e+05
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 5.63e+05. This might indicate that there are
strong multicollinearity or other numerical problems.
```

The dropped variable list is given by

	Dropped Variable	R2	MSE
0	c36	0.784773	1.089361
1	c27	0.784763	1.088310
2	c34	0.784744	1.087306
3	c142	0.784643	1.086720
4	c32	0.784505	1.086317
5	c7	0.784152	1.087002
6	c13	0.783681	1.088280
7	c162	0.783137	1.089919
8	c19	0.782408	1.092487

Here we can see that p values of most of the variables are 0.00 which are equally significant on the basis of the p-values. But on the basis of coefficients, variables with larger magnitude coefficients are more significant than other and we can see that c39 has the highest coefficient and hence the greatest significance.