Consider only the below columns and prepare a prediction model for predicting Price.

Corolla=Corolla[c("Price","Age\_08\_04","KM","HP","cc","Doors","Gears","Quarterly\_Tax","Weight")]

Exploratory data analysis:

First moment business decision:

Price Age\_08\_04 KM HP cc Doors Gears

Min. : 4350 Min. : 1.00 Min. : 1 Min. : 69.0 Min. : 1300 Min. :2.000 Min. :3.000

1st Qu.: 8450 1st Qu.:44.00 1st Qu.: 43000 1st Qu.: 90.0 1st Qu.: 1400 1st Qu.:3.000 1st Qu.:5.000

Median : 9900 Median :61.00 Median : 63390 Median :110.0 Median : 1600 Median :4.000 Median :5.000

Mean :10731 Mean :55.95 Mean : 68533 Mean :101.5 Mean : 1577 Mean :4.033 Mean :5.026

3rd Qu.:11950 3rd Qu.:70.00 3rd Qu.: 87021 3rd Qu.:110.0 3rd Qu.: 1600 3rd Qu.:5.000 3rd Qu.:5.000

Max. :32500 Max. :80.00 Max. :243000 Max. :192.0 Max. :16000 Max. :5.000 Max. :6.000

Quarterly\_Tax Weight

Min. : 19.00 Min. :1000

1st Qu.: 69.00 1st Qu.:1040

Median : 85.00 Median :1070

Mean : 87.12 Mean :1072

3rd Qu.: 85.00 3rd Qu.:1085

Max. :283.00 Max. :1615

#2nd moment business decision

sd(Price)

#3626.965

sd(Age\_08\_04)

#18.59999

sd(KM)

#37506.45

sd(HP)

#14.98108

sd(cc)

# 424.3868

sd(Doors)

#0.9526766

sd(Gears)

#0.1885104

sd(Quarterly\_Tax)

#41.12861

sd(Weight)

#52.64112

##########variance########

var(Price)#13154872

var(Age\_08\_04)#345.9596

var(KM)#1406733707

var(HP)# 224.4327

var(cc)#180104.1

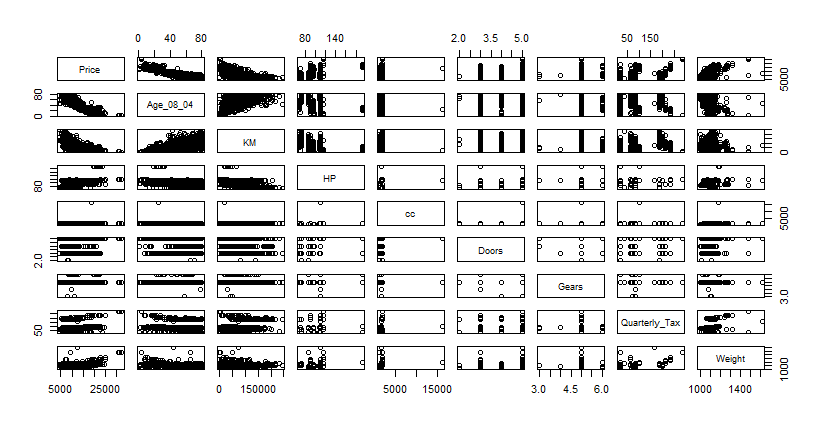
var(Doors)#0.9075927

var(Gears)#0.03553619

var(Quarterly\_Tax)#1691.563

var(Weight)# 2771.088

**pair plot:**

****

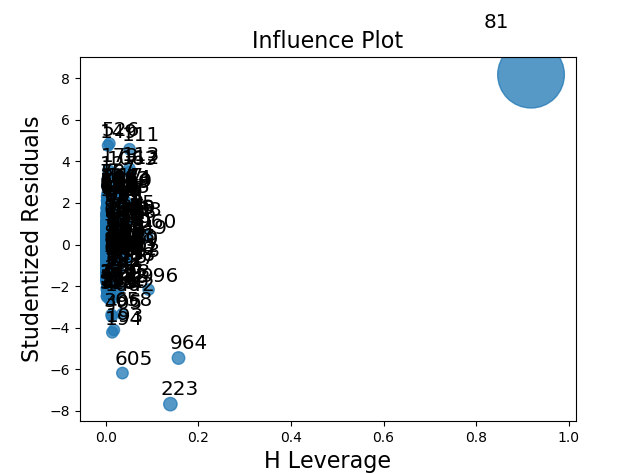
coefficient of determination:

model with all the columns:

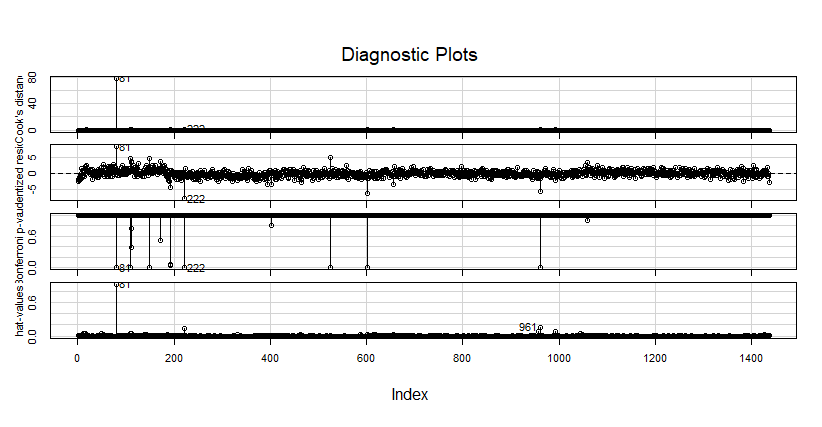
Multiple R-squared: 0.8638, Adjusted R-squared: 0.863

p-value: < 2.2e-16 for overall model

values for cc and Doors have higher values than the significance values.

So using the influence index plot try to figure out the values which are influencing the model.

Try removing the value and build the model again and see if there is any change in the p value.



Variance inflation factor:

Age\_08\_04 KM HP cc Doors Gears Quarterly\_Tax Weight

1.884620 1.756905 1.419422 1.163894 1.156575 1.098723 2.311431 2.516420

Vif value should be less than 10 :

As all the vif values are less than 10 and the p value remains same after removing the 81th index, so try removing the column with the highest p value.

final model:

created after removing the columns with the higher p values.

Multiple R-squared: 0.8636>0.8=>strong correlation

Adjusted R-squared: 0.863

P value : < 2.2e-16<0.05=>overall model is good

Added variable plots: 