# **Toyota Corolla**

1) Loading the data set

2) Removing the unwanted Sequence number

3) Removing the unwanted features from the data set

4) Used Box plot,Histogram, QQ plot to understand the features

5) Correlation Plot to understand multiple variable

6) Correlation Coefficient to find the strength and direction

7) library(corpcor) to Estimation of Co-variance

Plot is used to understand the relation of price and other features

Model 1 : All variable

Result : Rsq Value is 0.8638.

Found : CC and Doors variables are insignificant

Model 2 : Only with CC variable

Result : Rsq value is 0.01597

Found : CC data is significant . P value is less than .05

model 3 : Only with Doors Variable

Result : Rsq value is 0.03435

Found : Doors data is significant. P value is less than .05

model 4 : Only with cc and Doors Variable

Result : Rsq value is 0.04688

Found : Both data is significant. P value is less than .05

8) library(car) used to identify the influencing plot and VIF

Model 5 : Deleting influential records

Result : RSq Value is 0.8694

Found : Even after removing the influential value , Doors features is insignificant. P value is greater than .05

9) Added Variable plot to check correlation between variables and output variable

10) VIF shows that the model is not col-linearity in variables

Variance Inflation Factors ( VIF) greater than 10 means col-linearity in the variables. All the values are less than 10 in this dataset

Deleting influential records Model 5 has higher RSq Value.

Final Result

80% Train Data

20% Test Data

RSq Value is 0.8638

RMSE 1338.259