**Multi-Linear Regression**

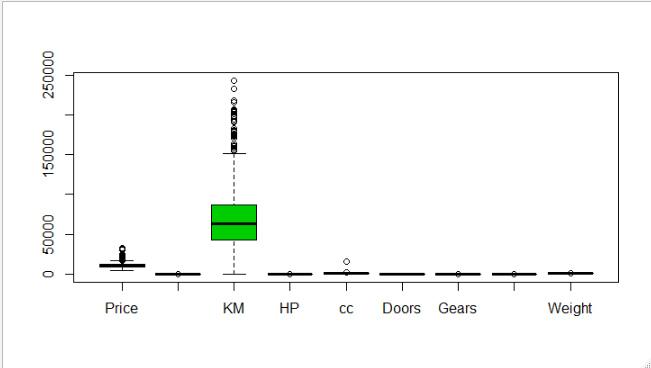
**Example- Toyota Corolla Dataset**

**Target variable is Price**

**Summary 🡺**

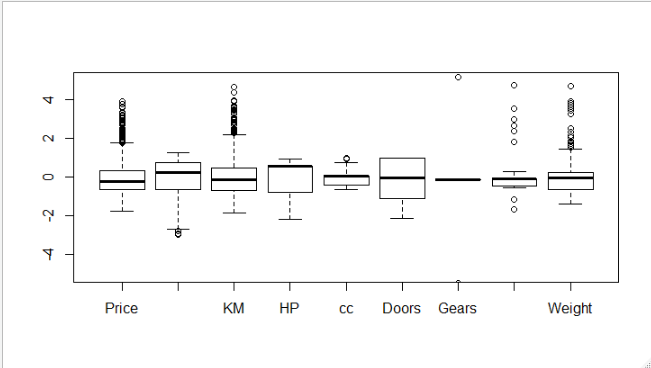
|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Price** | **Age\_08\_04** | **KM** | **HP** | **cc** | **Doors** | **Gears** | **Quarterly\_Tax** | **Weight** |
| **Min. : 4350** | **Min. : 1.00** | **Min. : 1** | **Min. : 69.0** | **Min. : 1300** | **Min. :2.000** | **Min. :3.000** | **Min. : 19.00** | **Min. :1000** |
| **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** | **1st Qu.: 69.00** | **1st Qu.:1040** |
| **Median:9900** | **Median:61.00** | **Median :63390** | **Median:110.0** | **Median:1600** | **Median:4.000** | **Median:5.000** | **Median :85.00** | **Median:1070** |
| **Mean :10731** | **Mean :55.95** | **Mean : 68533** | **Mean :101.5** | **Mean : 1577** | **Mean :4.033** | **Mean :5.026** | **Mean : 87.12** | **Mean :1072** |
| **3rdQu.:11950** | **3rd Qu.:70.00** | **3rd Qu.: 87021** | **3rd Qu.:110.0** | **3rd Qu.: 1600** | **3rd Qu.:5.000** | **3rd Qu.:5.000** | **3rd Qu.: 85.00** | **3rd Qu.:1085** |
| **Max. :32500** | **Max. :80.00** | **Max. :243000** | **Max. :192.0** | **Max. :16000** | **Max. :5.000** | **Max. :6.000** | **Max. :283.00** | **Max. :1615** |

**Box Plot 🡺**



**Based on above summary and box plot we can see that outliers are available in the dataset.**

**So will do scale free plot.**



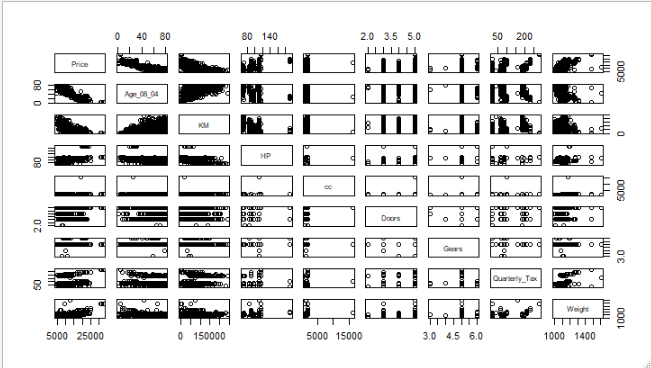
**We can see that large number of outliers are in the dataset.**

**Correlation 🡺**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Price** | **Age\_08\_04** | **KM** | **HP** | **cc** | **Doors** | **Gears** | **Quarterly\_Tax** | **Weight** |
| **Price** | **1** | **-0.876590497** | **-0.569960165** | **0.31498983** | **0.126389197** | **0.18532555** | **0.063103857** | **0.219196911** | **0.581197589** |
| **Age\_08\_04** | **-0.876590497** | **1** | **0.50567218** | **-0.15662202** | **-0.098083739** | **-0.148359215** | **-0.005363947** | **-0.198430508** | **-0.470253184** |
| **KM** | **-0.569960165** | **0.50567218** | **1** | **-0.333537948** | **0.102682891** | **-0.036196614** | **0.015023328** | **0.278164697** | **-0.028598457** |
| **HP** | **0.31498983** | **-0.15662202** | **-0.333537948** | **1** | **0.035855803** | **0.092424496** | **0.209477146** | **-0.298431717** | **0.089614059** |
| **cc** | **0.126389197** | **-0.098083739** | **0.102682891** | **0.035855803** | **1** | **0.079903296** | **0.014629352** | **0.306995798** | **0.335637399** |
| **Doors** | **0.18532555** | **-0.148359215** | **-0.036196614** | **0.092424496** | **0.079903296** | **1** | **-0.16014143** | **0.109363225** | **0.302617644** |
| **Gears** | **0.063103857** | **-0.005363947** | **0.015023328** | **0.209477146** | **0.014629352** | **-0.16014143** | **1** | **-0.005451955** | **0.020613284** |
| **Quarterly\_Tax** | **0.219196911** | **-0.198430508** | **0.278164697** | **-0.298431717** | **0.306995798** | **0.109363225** | **-0.005451955** | **1** | **0.626133733** |
| **Weight** | **0.581197589** | **-0.470253184** | **-0.028598457** | **0.089614059** | **0.335637399** | **0.302617644** | **0.020613284** | **0.626133733** | **1** |

**From the above table it is clearly seen that Price and Age are highly negatively correlated.**

**Pairs Plot 🡺**



**Model-1 🡺**

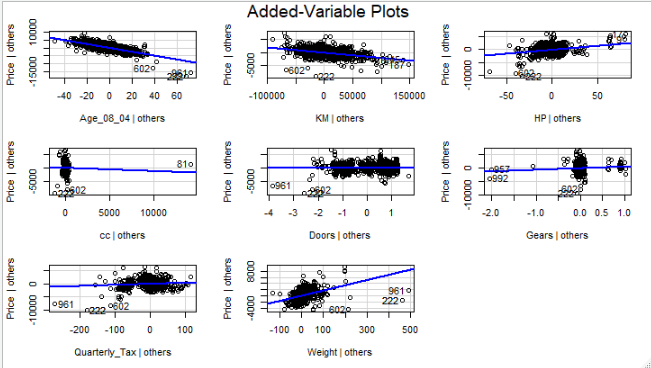
model\_T\_1 <- lm(Price~.,data = Corolla)

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

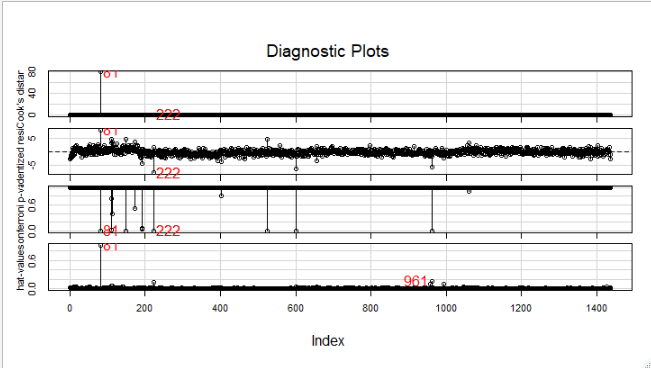
correlation is 0.9293884

RMSE value as 1338.258

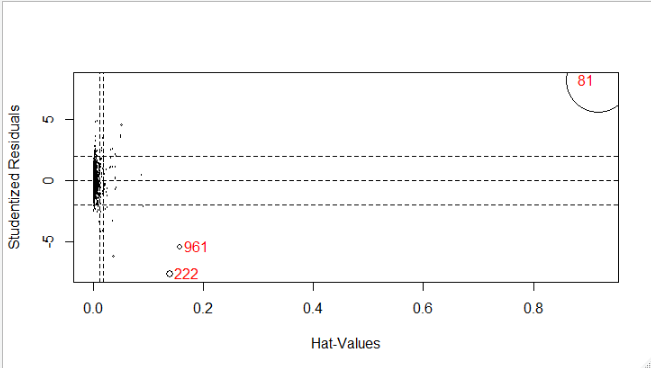
**AV Plot 🡺**



**From the above plot, cc, Doors and Gears are insignificant for our model.**



**Influence Plot**



**From the above plot, 81,222,961 are more influencing our model.**

**Model-2** 🡺

df\_Corola <- Corolla[-c(influence\_index),]

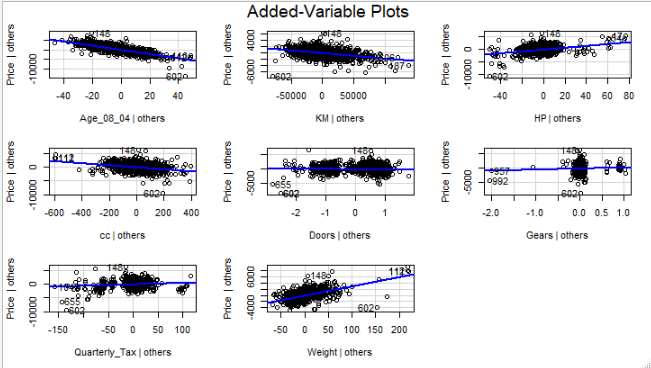
> model\_T\_2 <- lm(Price~.,data = df\_Corola)

Multiple R-squared: 0.8852, Adjusted R-squared: 0.8845

correlation as 0.9408425

RMSE value as 1227.474

**AV Plot**



**From the above plot now cc is showing somewhat significance but Doors and Gears are not.**

**Comparison 🡺**

|  |  |  |  |
| --- | --- | --- | --- |
| **Model No** | **R^2** | **RMSE** | **Cor** |
| **Model-1** | **0.8698** | **1338.25** | **0.929** |
| **Model-2** | **0.8852** | **1227.474** | **0.9408425** |

**From above table we can infer Model-2 is best model as it is having less RMSE and highly correlated between predicted and actual value with 80% variation in Price.**