**Toyota Data Report:**

* The Toyota data set has Many feature ‘s but we are considering only the Important and Needed for the problem features
* Preparing the Prediction Model for Predicting the “Price”
* The Dimensions of the Data set is (1436,9)
* The main Features we consider in the data set are
* *"Price","Age\_08\_04","KM","HP","cc","Doors","Gears","Quarterly\_Tax","Weight".*
* After that I was calculated the 4 Business moments

**Checking for the Outliers and Missing Values**

* They are no Outliers and Missing values in the Data set.
* Plotted using Boxplot to identify the outliers

**Checking the Correlation and Best Fit Model using OLS**

* All the features are moderately correlated
* By considering all the features I created a model called “all\_vars” this contains two in-Significant variables they are

(CC & Doors) and R2 🡺0.864, AIC🡺2.477

* Checking the In-Significant variable one by one by using SLR technique then they became Significant variables
* By combining both In-Significant variables with dependent variable “Price” they became significant variables

VIF Values:

| **Variables** | **VIF** |
| --- | --- |
| **0** | CC | 1.163894 |
| **1** | Age\_04\_mnths | 1.884620 |
| **2** | KM | 1.756905 |
| **3** | HP | 1.419422 |
| **4** | Doors | 1.156575 |
| **5** | Gears | 1.098723 |
| **6** | Quarterly\_Tax | 2.311431 |
| **7** | Weight | 2.516420 |

**Calculated VIF values (Variance Inflation Factor):**

By Observing the VIF values in the above can say that there is NO Multicollinearity Problem exist there is no VIF values more than 5 or 10 if the VIF values are greater than 5 or 10 then it considered as a problem of Multicollinearity

**Residual plot for Hoscedasticity:**

by observing the above graph, we are not able to see any pattern in the data set

**# Model Deletion Diagnostics**

**Finding Influential variables and Outliers**

**Cooks Distance value is 80**

Here the data point 80 is high influential observations when compared to normal data

**Influence Plot**

By observing the above influential observation with data usually "CC " has Higher value when compared to others

So, I Dropped the value of 80

Final Model: created a final\_model1 using the ols

R^2 🡺0.869

AIC🡺24687

By Observing the "final\_model1" model has low Aic value and Same R2 and Same Adjusted R2 values in my case "doors" are not that much important feature to predict the price while compared to other features