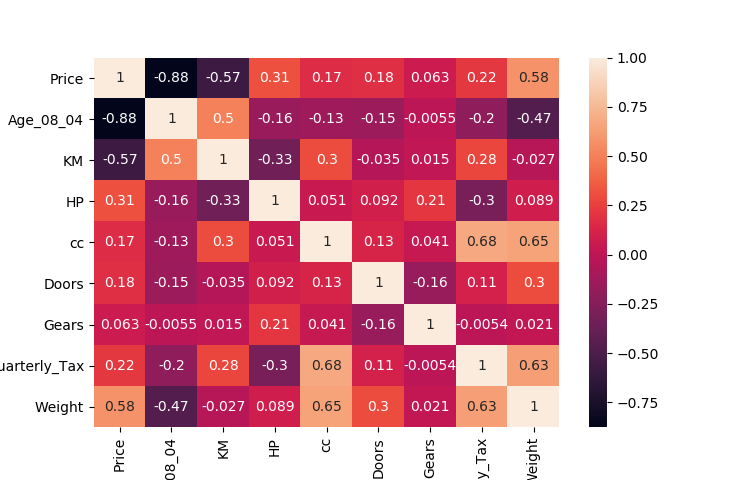
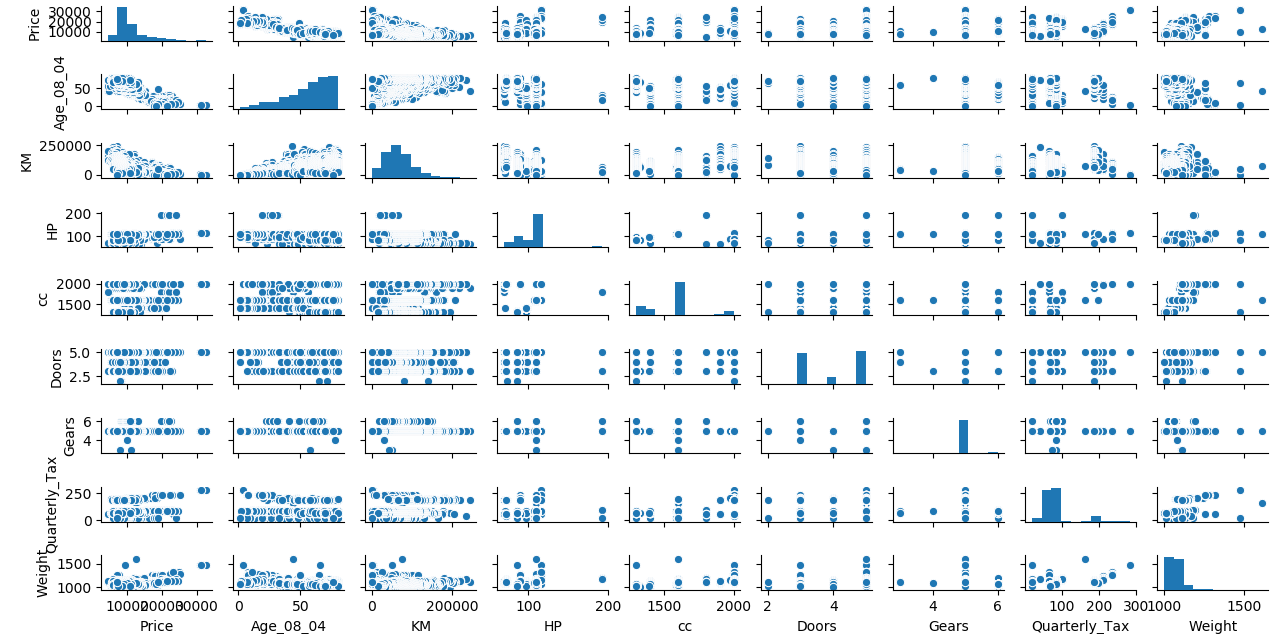
**MULTILINEAR RIGRESSION**

**Business Problem** = ﻿Prepare a prediction model for Predicting Price.

* **Name of the File: -** ToyotaCorolla.csv
* **Size of the File: -** 244 KB
* **Data: -** 1435 Observation, 9 Variable
* **Column Name: -** Price,Age\_08\_04,KM,HP,cc,Doors,Gears,Quarterly\_Tax,Weight

**Exploratory data Analysis** =

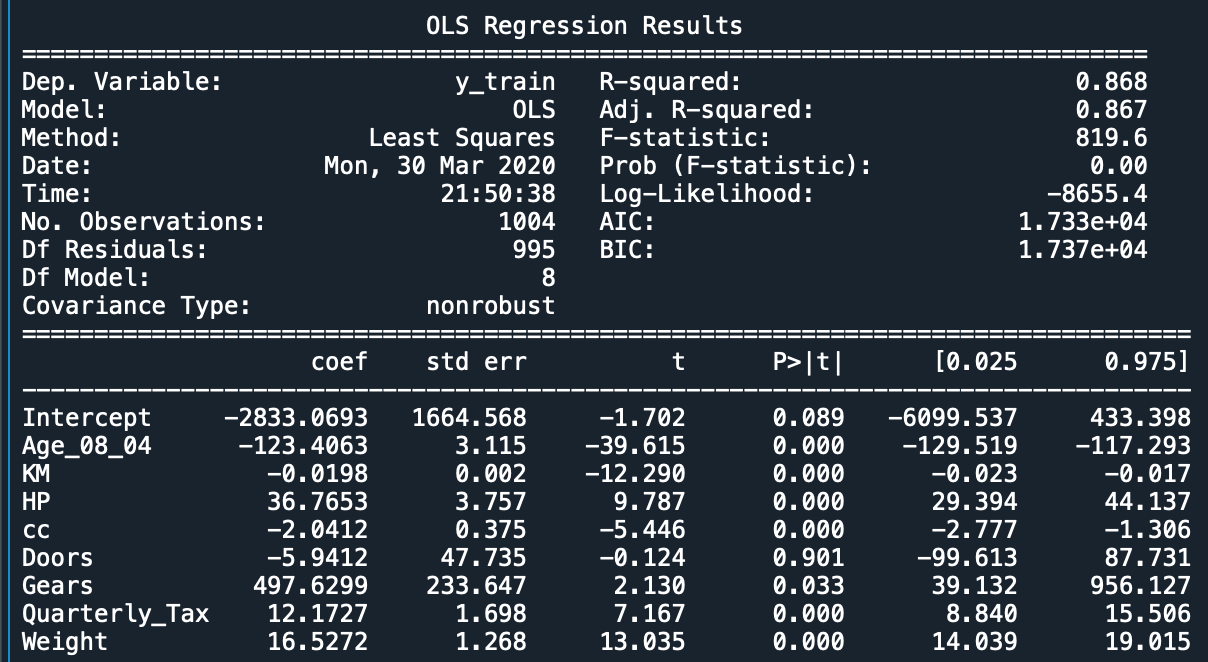
* **Outliers: -**  yes outliers are presents and deleting one outlier .
* **Missing Value: -** Data don’t have Missing Values
* **Normality: -** Data are not normal
* **Transformation: -**  May be Required to improve accuracy

******Scatter plot =** From below scatter plot we can say that Age and Price, Km and Price is having good negative correlation and rest are having weak or moderate correlation.

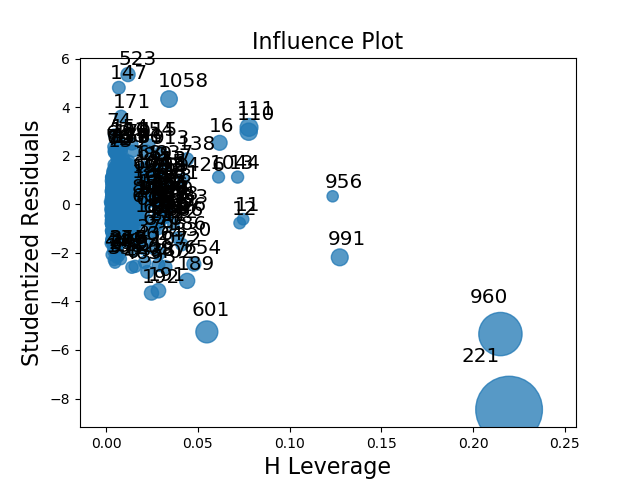
**Correlation Coefficient (r) =**  From heat map we can see that Age and price is have highest correlation and after that Km and Price is have moderate correlation and also Weight is having moderate positive correlation with price and we cannot observe any high correlation between any two input variable so it means data don’t have any collinearity problem.

**Model 1 Building =**

* **Summary: -**

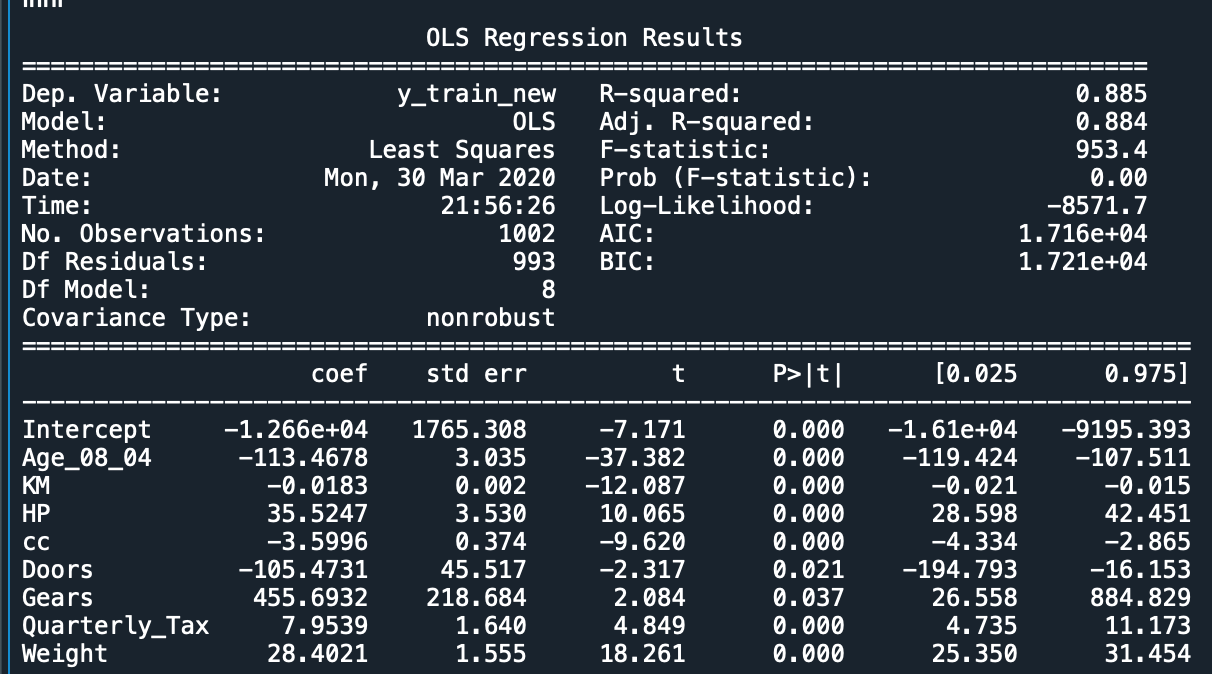
****

We can see from above summary the variable Door is insignificant we have to check there is any influencer observation present in doors data. So we use influencer plot to check this.

* **Influencer plot: -** There are two influencer observation 960 and 221 which may create problem in our model.

**Model 2 Building =** Building model after removing influencer.

* **Summary: -**

****

In above model R2 value is strong, Adj. R2 is also strong and all variable are significant we can use this model for prediction.

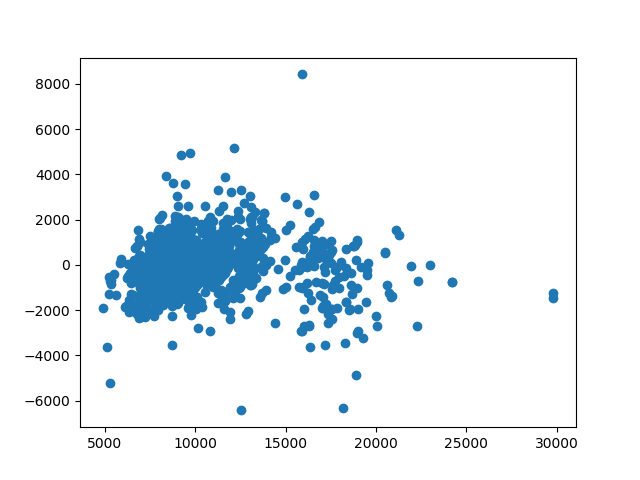
* **Intersect(B0): - -**12660
* **Slope(B1): -** -113.47, -0.018, 35.5, -3.6, -105.47, 455.7, 7.95, 28.4
* **Coefficient of**

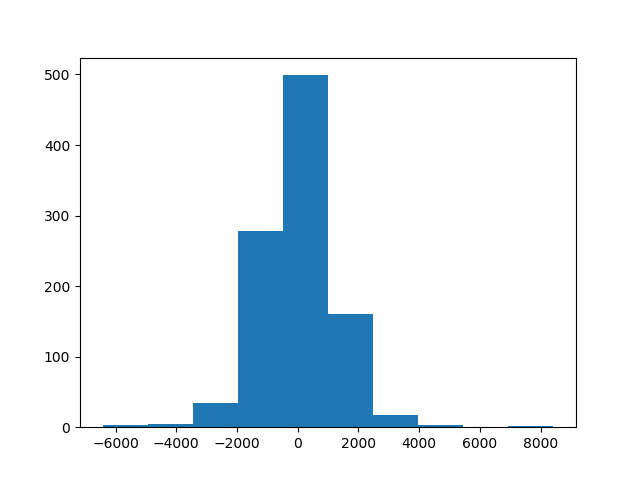
**Determination (R2): -** 0.885 - strong Correlation

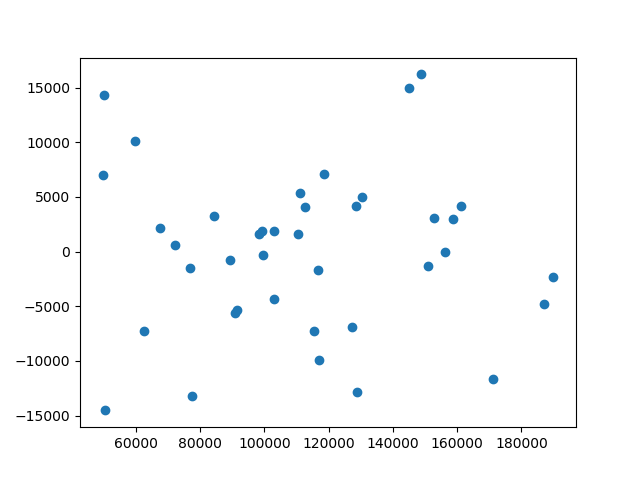
* **Pvalue: -** Pvalue less than 0.05, we can use intercept & slop for model building.
* **Model: - ﻿﻿Price =** -12660 - 113.46(Age\_08\_04) - 0.0183(KM) + 35.52(HP) –

3.60(cc) - 105.47(Doors) + 455.7(Gears) + 7.95(Quarterly Tax) + 28.40(Weight)

* **RMSE:** - 1256

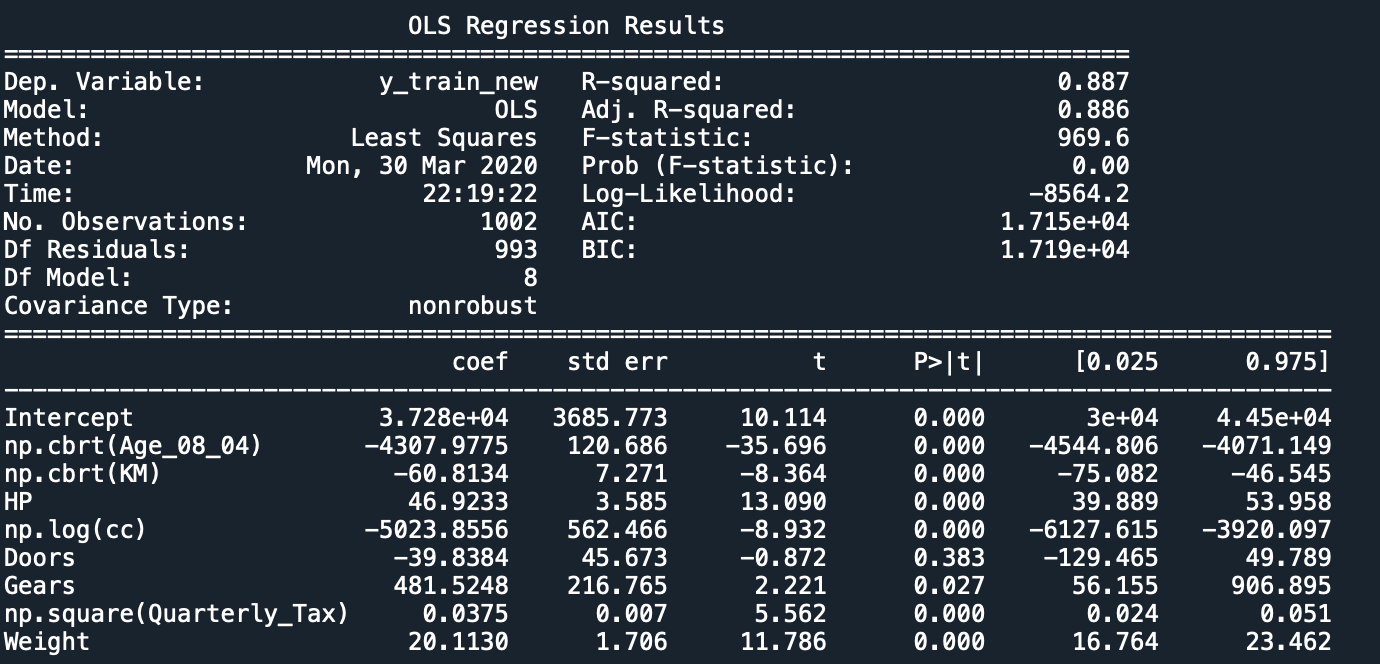
**Residual Scatter Plot**: -

****Error do not have any pattern. It means errors are independent and from below histogram we can say that errors are nearly normal distributed but having some outliers.

****Our R value is strong but we also try to improve it by applying varies transformation.

**Model 3 Building =** Model after selecting best transformations of each variable.

* **Summary: -**

****

After trying several transformation the model is not able to significant therefore we cannot use transformations.

After considering all aspect we can say that model 2 is give us more accurate result as compare to other model.so we used model 2 as final model.

**Final Models with 95% confidence interval: -**

**﻿Price = -**16100 - 119.42(Age\_08\_04) - 0.021(KM) + 28.6(HP) - 4.334(cc) - 194.8(Doors) +

26.558(Gears) + 4.735(Quarterly Tax) + 25.35(Weight)

**Price** = -9195 - 107.51(Age\_08\_04) - 0.015(KM) + 42.45(HP) - 2.865(cc) - 16.153(Doors) +

884.82(Gears) + 11.173(Quarterly Tax) + 31.45(Weight)

**Python code file**: - [Toyota Corolla Analysis.py](https://github.com/nilaydeshmukh0/Multiple-Linear-Regression-With-EDA/blob/master/Toyota%20Corolla%20Analysis/Toyota%20Corolla%20Analysis.py)