**MULTILINEAR RIGRESSION**

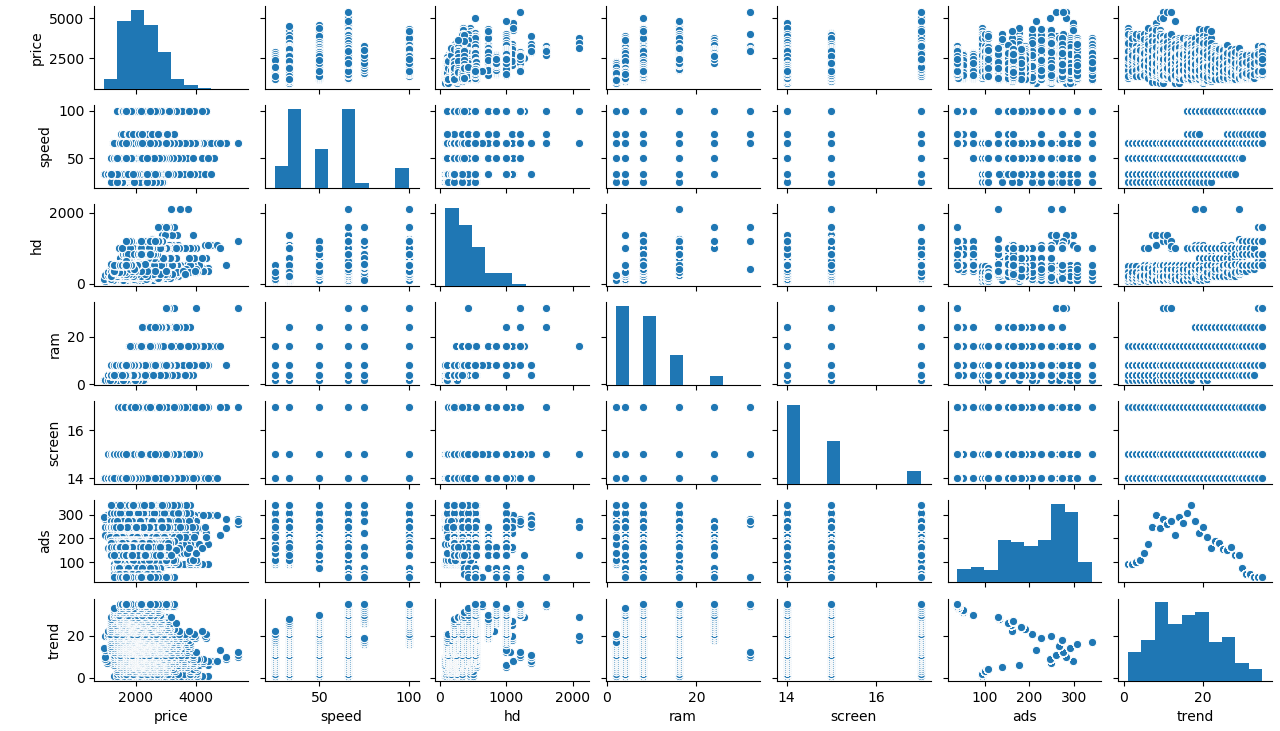
**Business Problem** = ﻿﻿Prepare a prediction model sales Price of the computer.

* **Name of the File: -** Computer\_Data.csv
* **Size of the File: -** 253 KB
* **Data: -** 6259 Observation, 10 Variable
* **Column Name: -** price, speed, hd, ram, screen, cd, multi, premium, ads ,trend

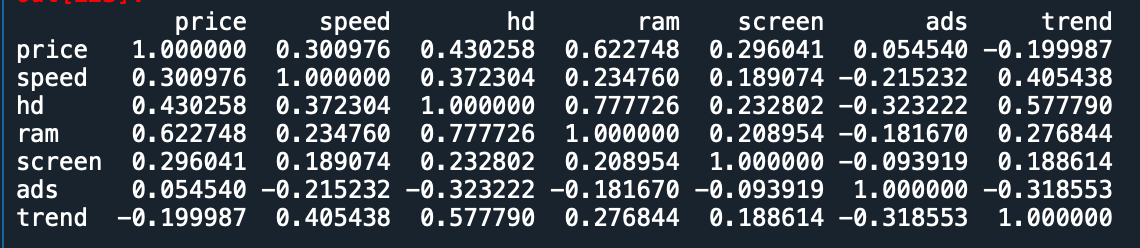
**Exploratory data Analysis** =

* **Outliers: -**  yes outliers are presents.
* **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 there are not have any strong correlation

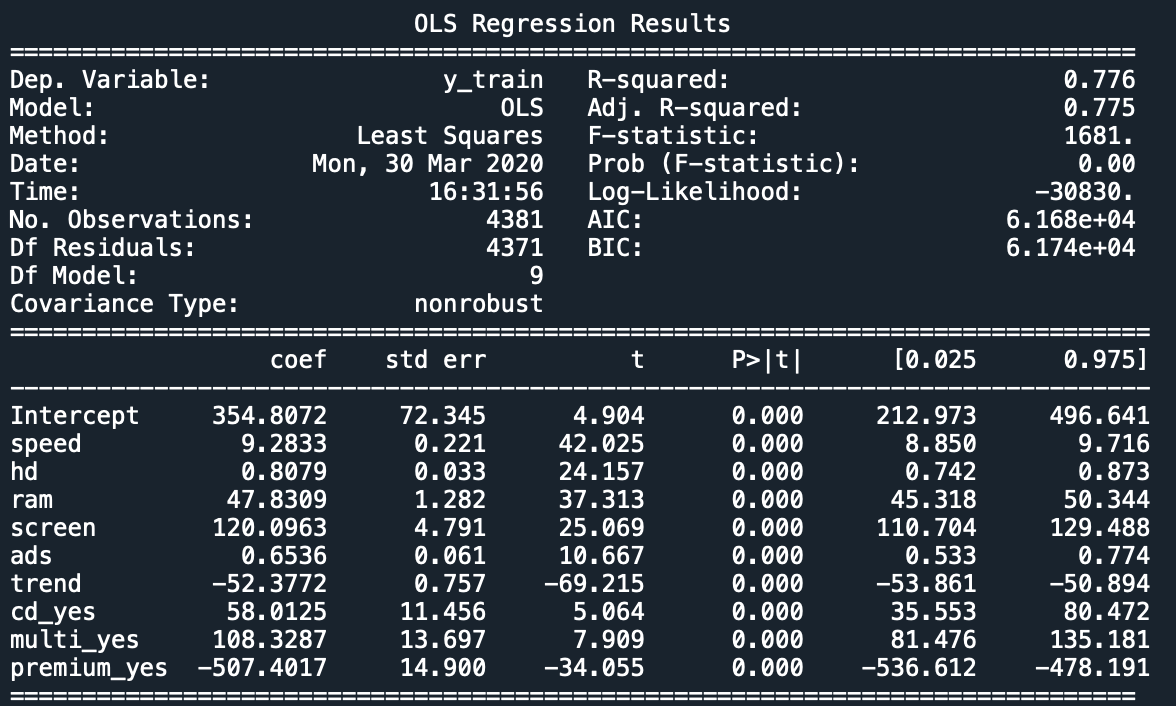


**Correlation Coefficient (r) =**  Correlation between price and ram is having moderate correlation and all other having weak correlation

****

**Model 1 Building =**

* **Summary: -**

****

In above model R2 value is moderate, Adj. R2 is also moderate and all variable are significant

* **Intersect(B0): -** 354.80
* **Slope(B1): -** 9.28, 0.80, 47.83, 120.09, 0.65, -52.37, 58.01, 108.32, -507.40
* **Coefficient of**

**Determination (R2): -** 0.76 - Moderate Correlation

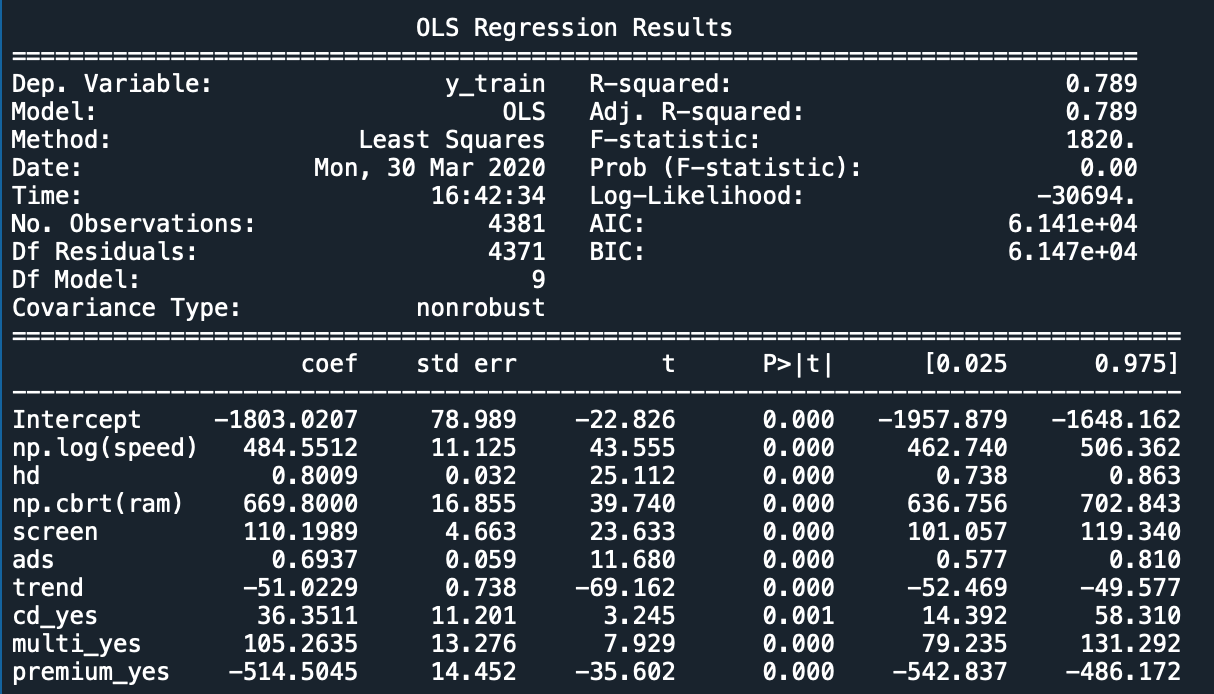
* **Pvalue: -** Pvalue less than 0.05, we can use intercept & slop for model building.
* **Model: - ﻿Price =**354.8+9.28(speed)+0.8(hd)+47.8(ram)+120(screen)+0.65(ads)-

52.4(trend)+58(cd)+108.32(multi)-507.4(premium)

* **RMSE:** - 275

**Model 2 Building =** Building model after considering transformation on each variable to increase R2

* **Summary: -**

****

In above model R2 value is moderate, Adj. R2 is also moderate and all variable are significant

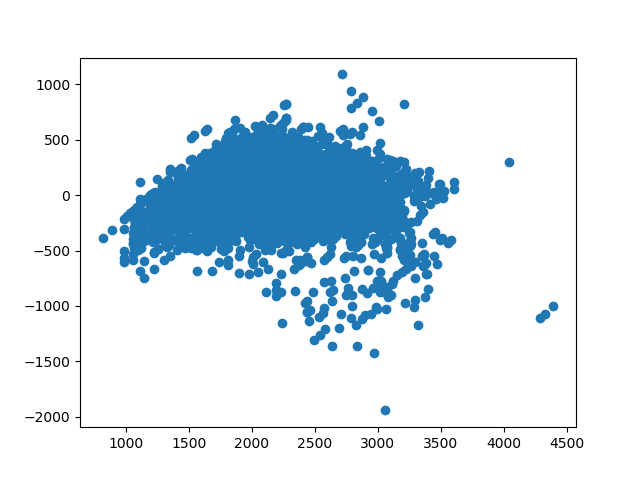
* **Intersect(B0): -** -1803
* **Slope(B1): -** 484.55, 0.80, 669.8, 110.2, 0.7, -51, 36.35, 105.3, -514.5
* **Coefficient of**

**Determination (R2): -** 0.79 - Moderate Correlation

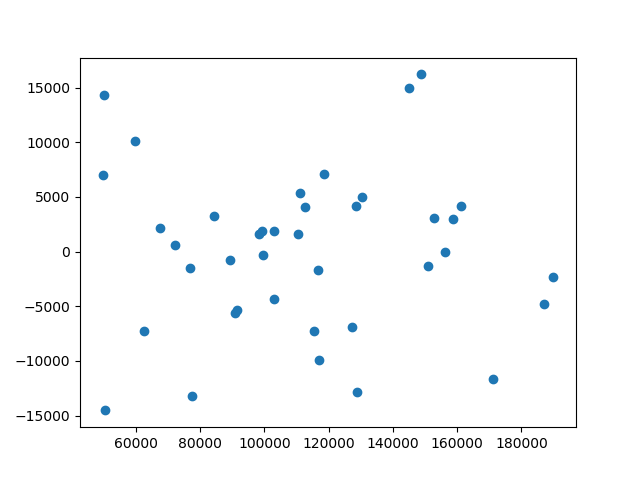
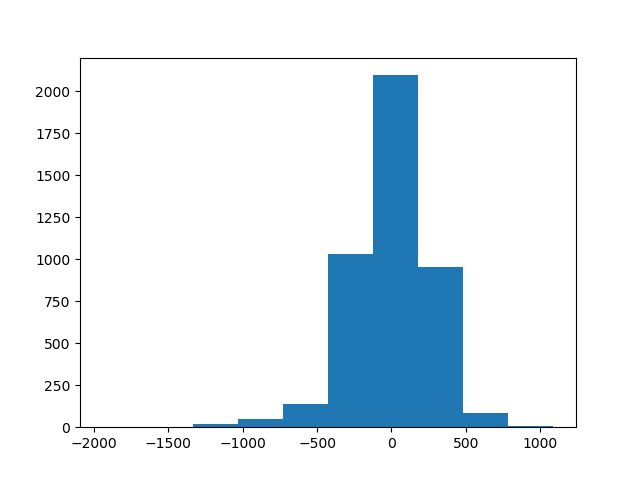
* **Pvalue: -** Pvalue less than 0.05, we can use intercept & slop for model building.
* **Model: - ﻿Price =** -1803+484.55log(speed)+0.8(hd)+669.8(ram)+110.2(screen)+

0.7(ads)-51(trend)+36.35(cd)+105.26(multi)-514.5(premium)

* **RMSE:** - 266

**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.

****

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 = -**1957.9+426.74log(speed)+0.73(hd)+636.75(ram)+101(screen)+0.6(ads)-

52(trend)+14.4(cd)+79.23(multi)-542.837(premium)

* **Price =** -1648,16+506.36log(speed)+0.863(hd)+702.84(ram)+119.34(screen)+0.81(ads)-

49.6(trend)+58.3(cd)+131.3(multi)-486.17(premium)

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