**Case Study of Computer Data**

**Problem Statement:**

Predict Price of the computer

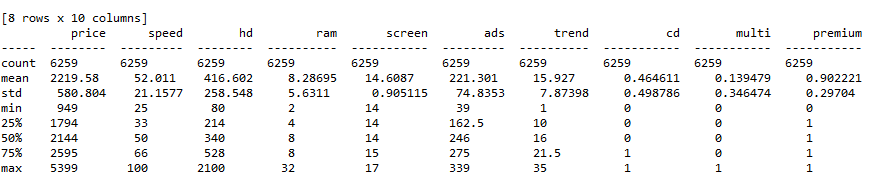
**Data**: Computer\_Data.csv {6259 observation with 10 variables}

Y: Dependent Variable => price (Continuous)

X: Independent Variable => speed, hd, ram, screen, ads & trend (Continuous)

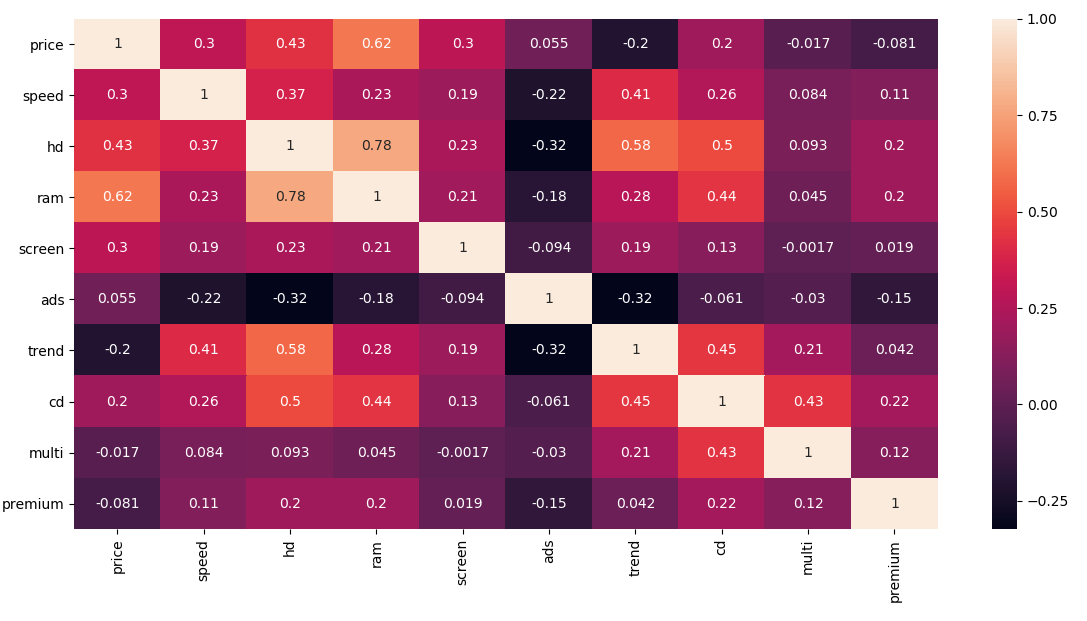
=> cd, multi & premium (Discrete)

**Exploratory data analysis:**



The above is the descriptive statistics of the Computer dataset. There are 6259 observations and 10 variables/features in the data frame.

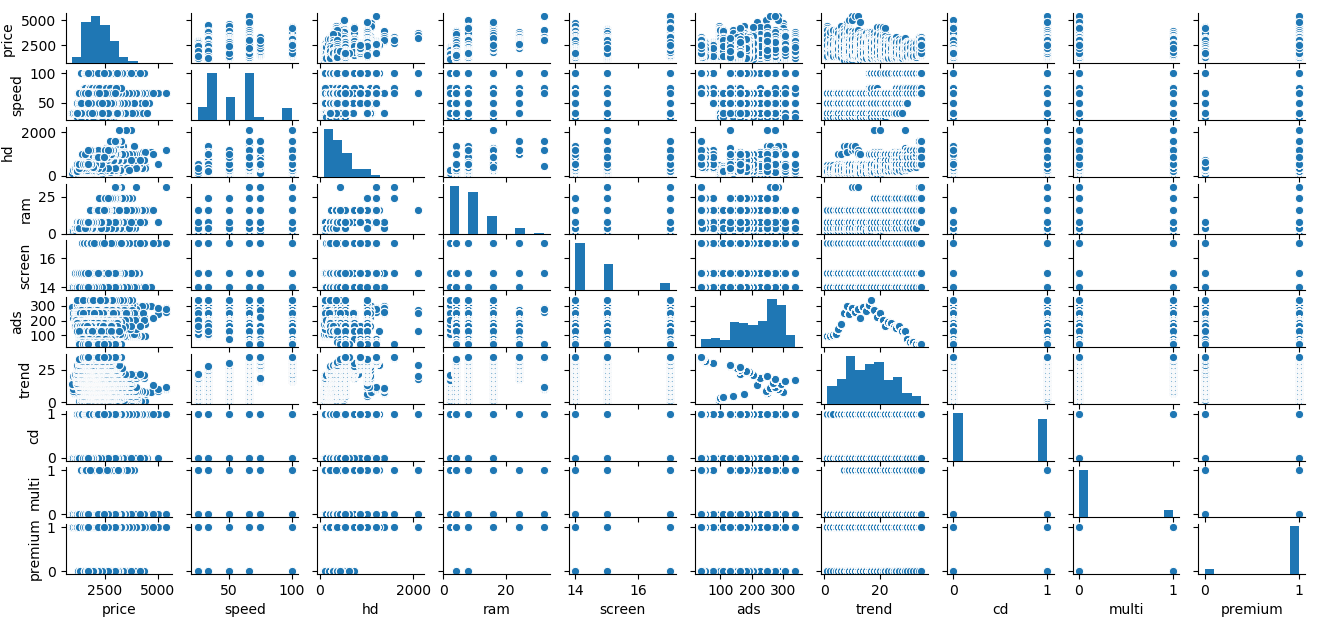
The heat map of the correlation between different variables



From the above heat map we can estimate the following:

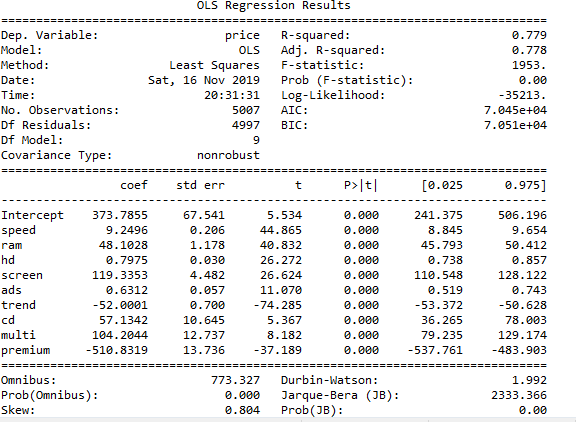
* Comparatively higher correlation between Price and ram
* Comparatively higher correlation between Price and hd
* The correlation between independent variables hd and ram have high correlation that may led to multi collinearity issue.

The following is the pairplot of the dataframe:



Building the regression model:

The following is the model considering all the variables and after splitting the data into train and test data



From the above model summary.

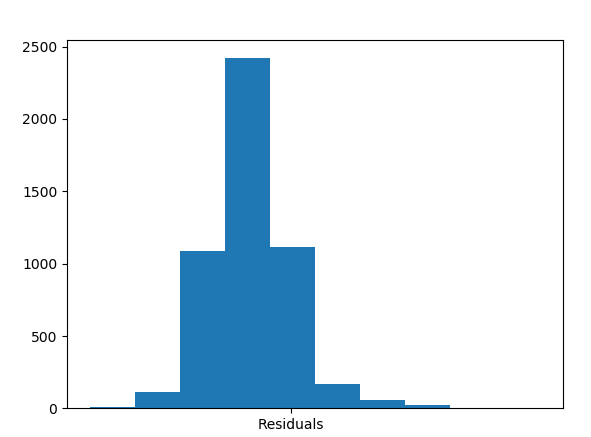
* The variables or the features are statistically significant for the outcome variable of price.
* The R-squared and Adjusted R-squared values of .779 and 0.778 indicates that the model is very good.
* The F-statistics value of less than 0.05 also indicates the good health of the model.

Train RMSE : 274.2036418059921

Test RMSE : 279.030642493059

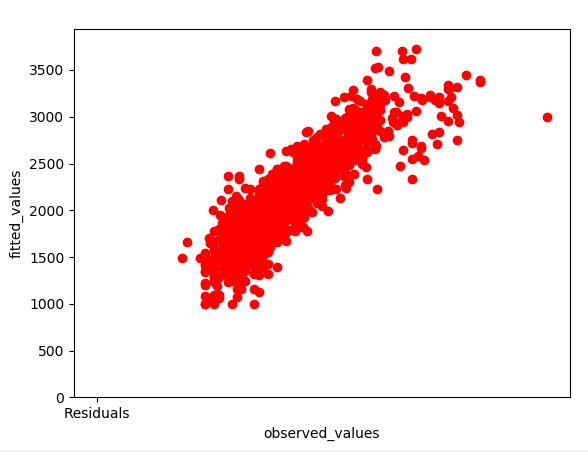
**Residual Analysis:**

Checking the normality of the residuals or errors:



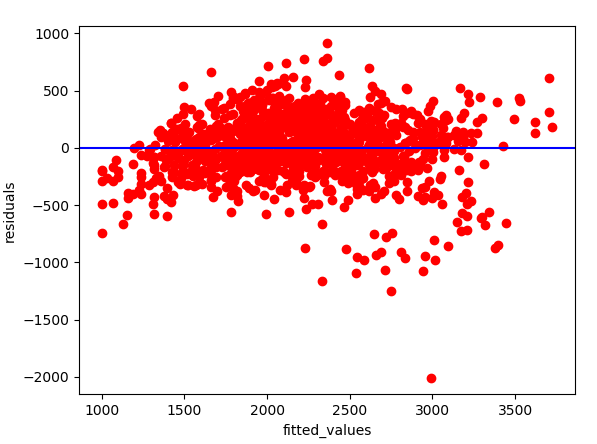
There is almost normal graph of the residuals.

**Checking for Fitted Vs Observed Values**

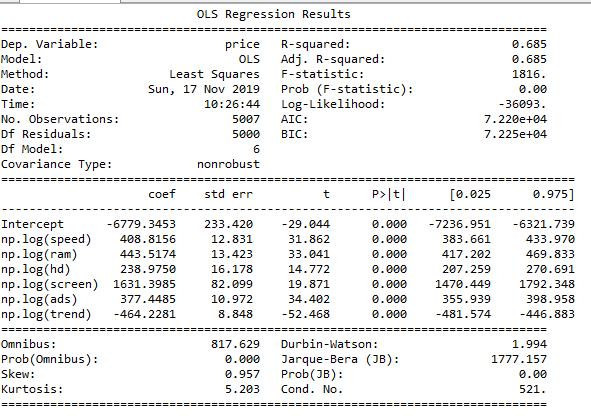


The above graph shows the conformance of the fitted and observed values

Checking for Residuals Vs Fitted values

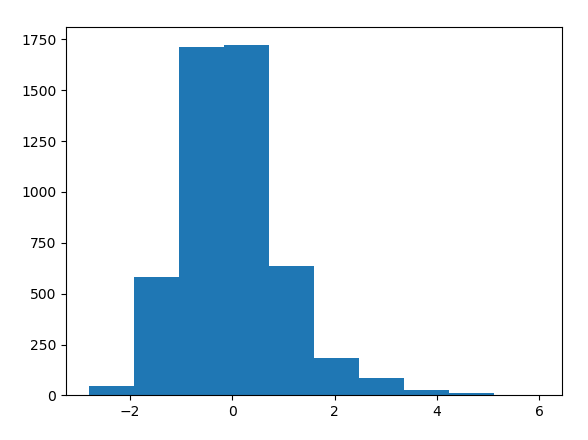


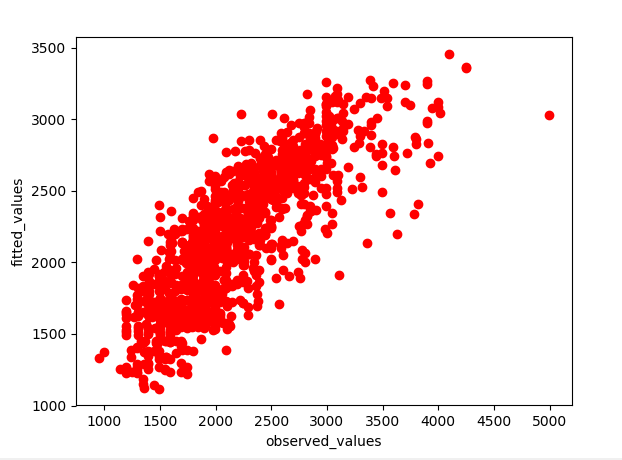
**Applying Log Transformation for the training data :**

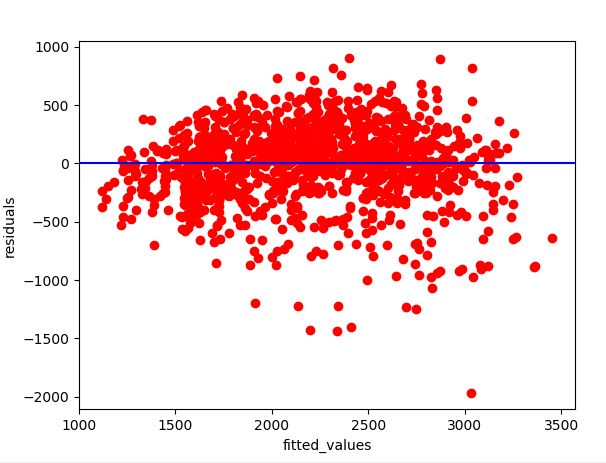


The categorical variables were dropped for log transformation.

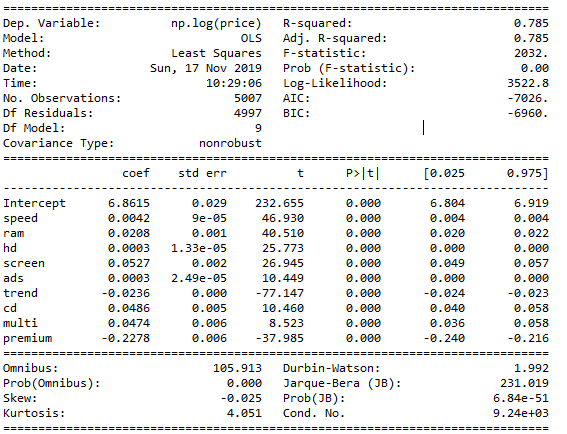
The above model provides a good R-squared values and there are no insignificant variables:

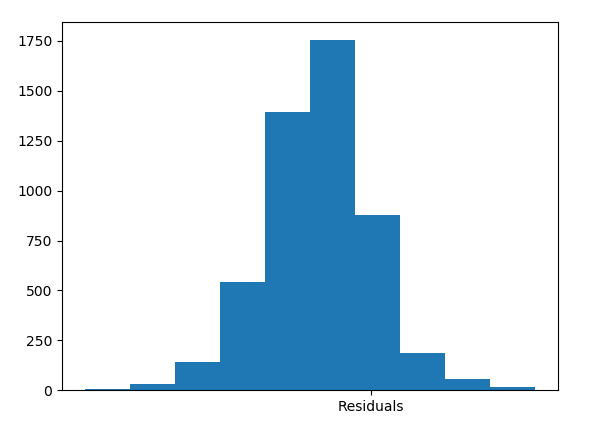


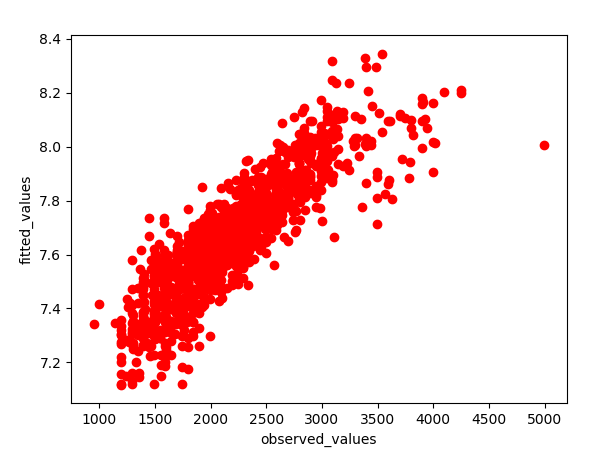




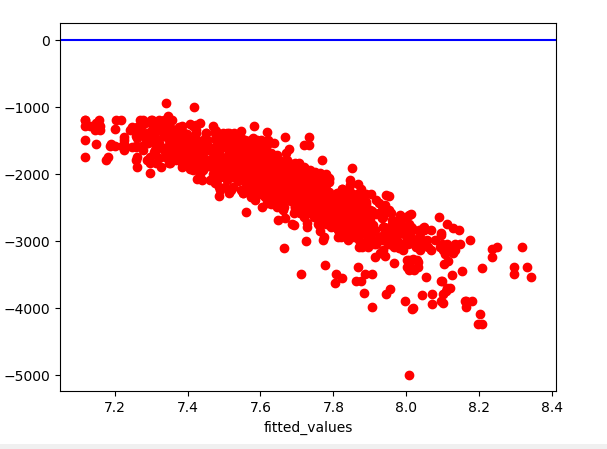
**We apply exponential transformation to the model:**







The exponential model still maintains good linearity with the fitted values



Exponential transformation has heteroscedasticity.

RMSE of all the models:

