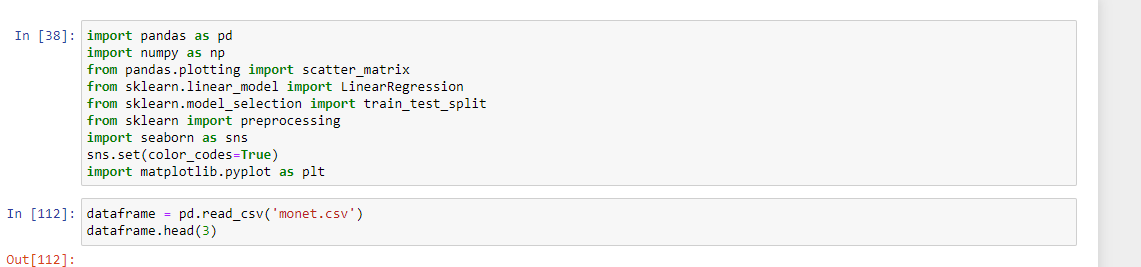
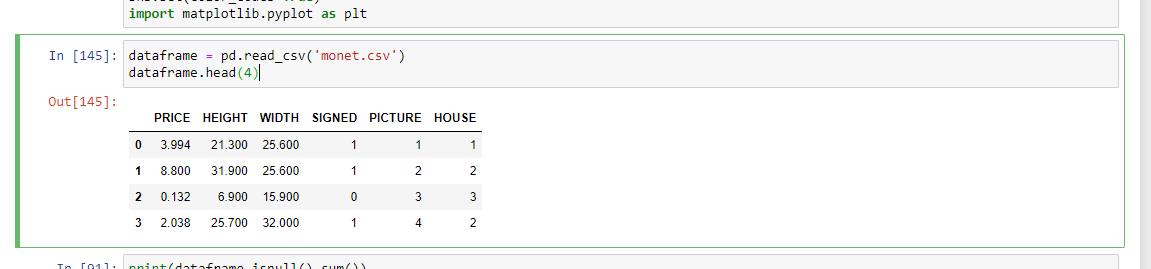
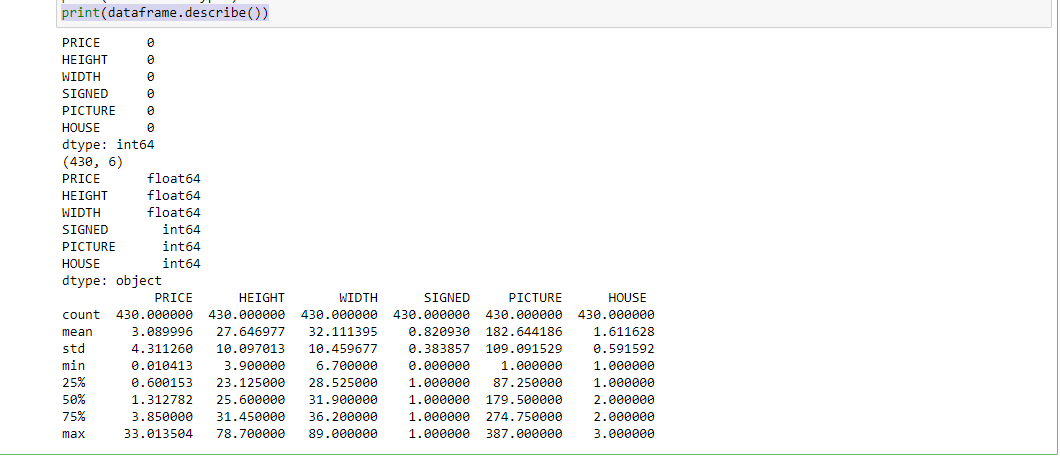
**Linear Regression**

1. I have imported the monet.csv data set into the jupyter note book by using pandas and loaded it into the data frame and named the data frame as dataframe.





2.To perform explotary analysis I have used the function print(dataframe.describe()).

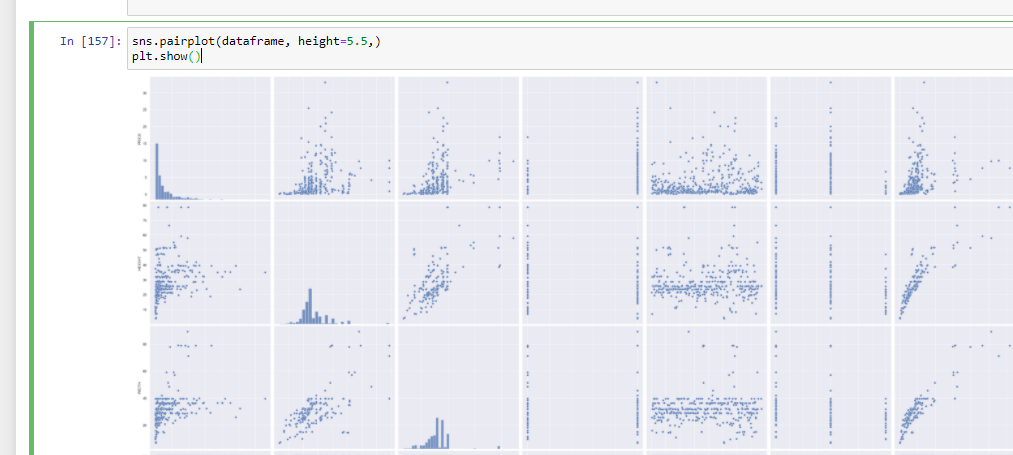


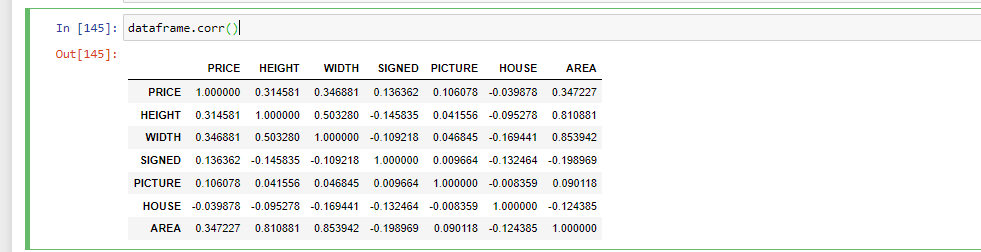
As shown as example in the assignment I have combined two Variable, height and width and named that as AREA.

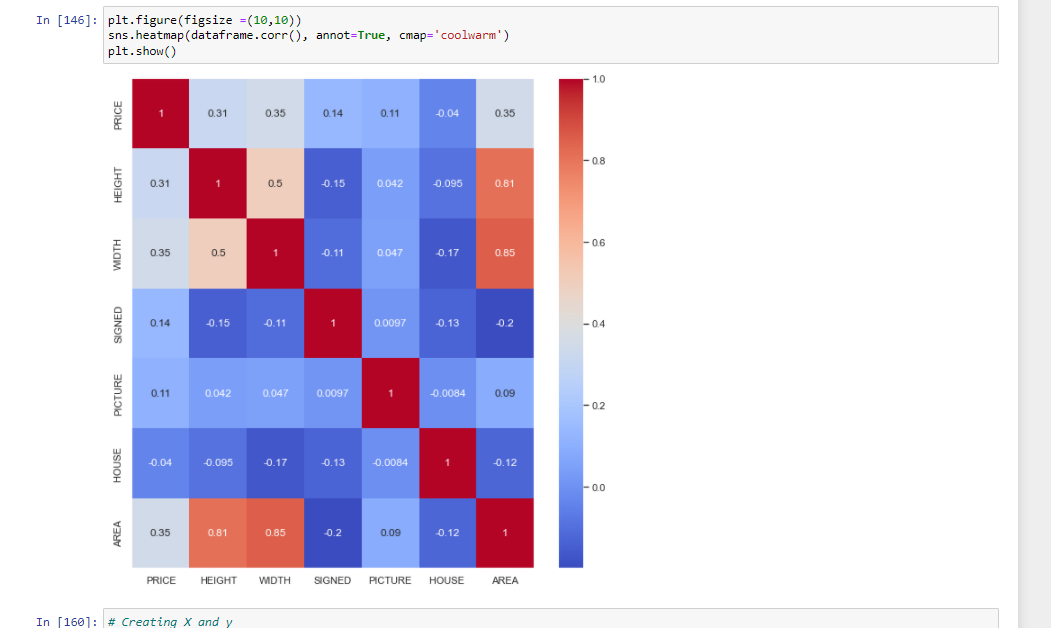


The above image shows that there are no null and missing values in the data set hence we can proceed with our linear regression model.

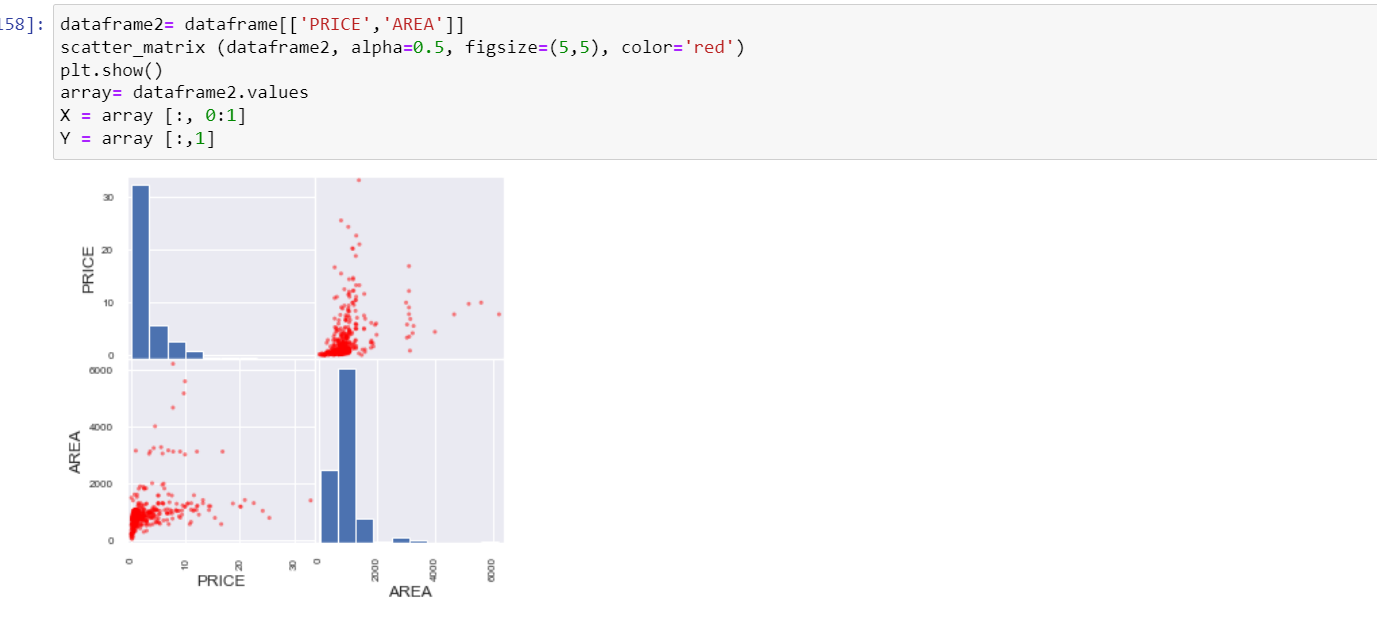
This shows image below shows the relationship between the variables that are present in the data set



The below image shows the correlation of the newly formed data set.   




The above pic shows the heat map for the data set



The above screenshot shows the visualisation for the model 1 I have taken price and area under consideration.

I have taken the test size as 30% and train size to be as 70% for my model 1 linear regression. Hence I have split the data set

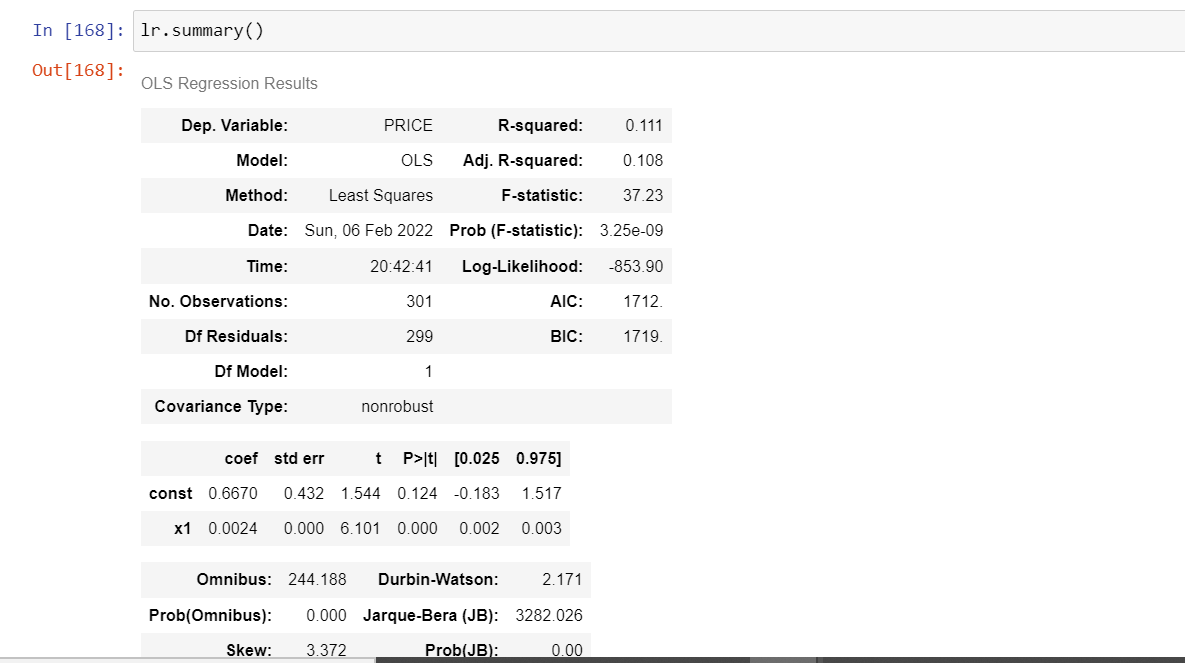


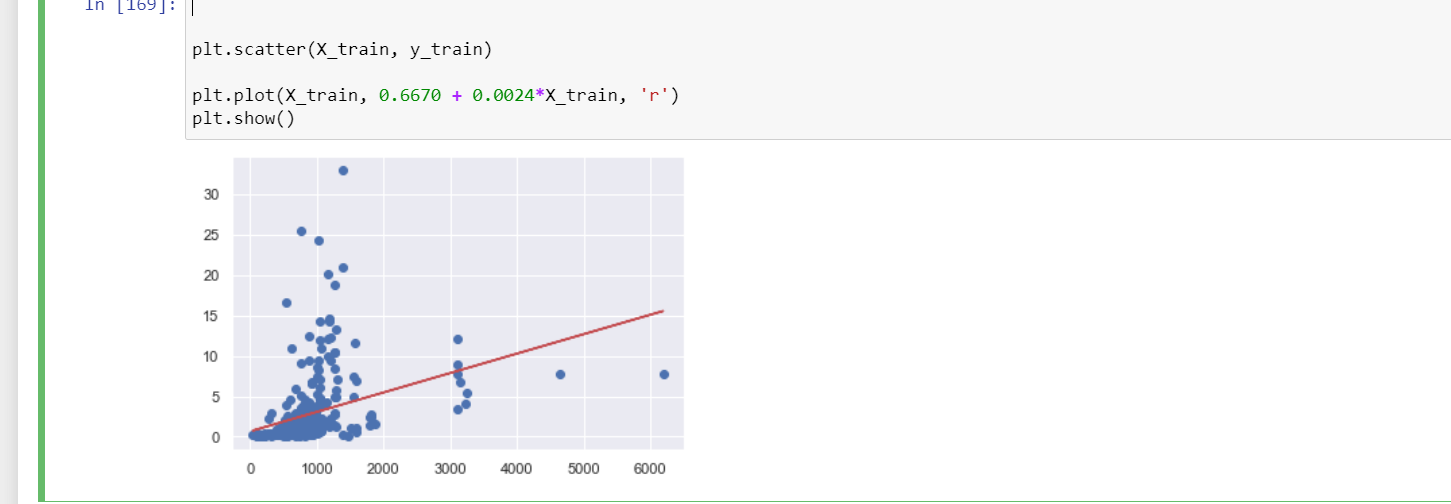
As we know that the linear regression is Y=mx+C it has an intercept as c so to find out the constant we have to add it manually which I have done in the above screenshot.

I have used the OLS method that is ordinary least square method that is present in statsmodel.

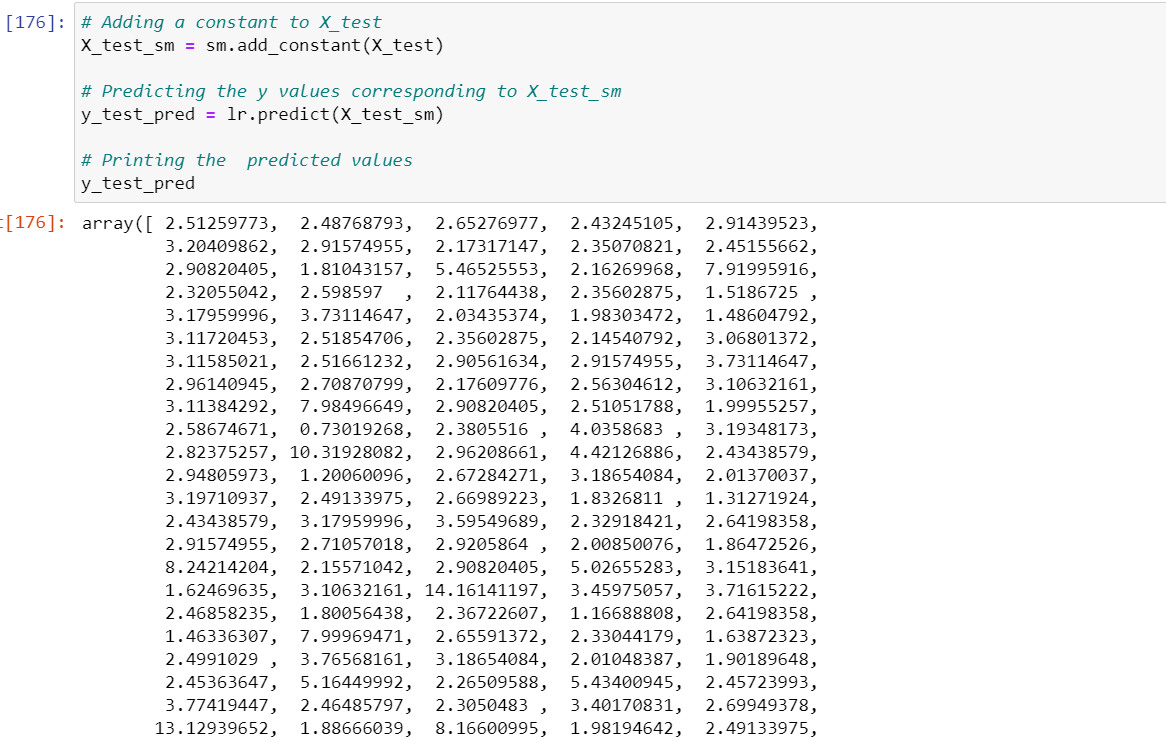
As we see we got the parameters for our model constant=25.280.

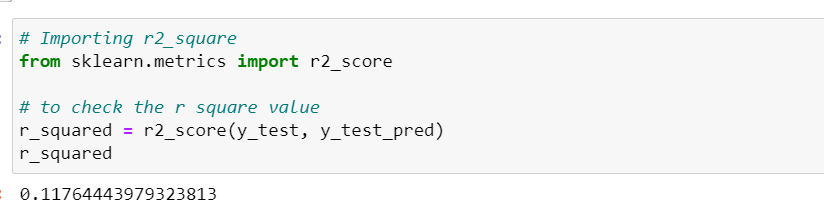
M=0.685



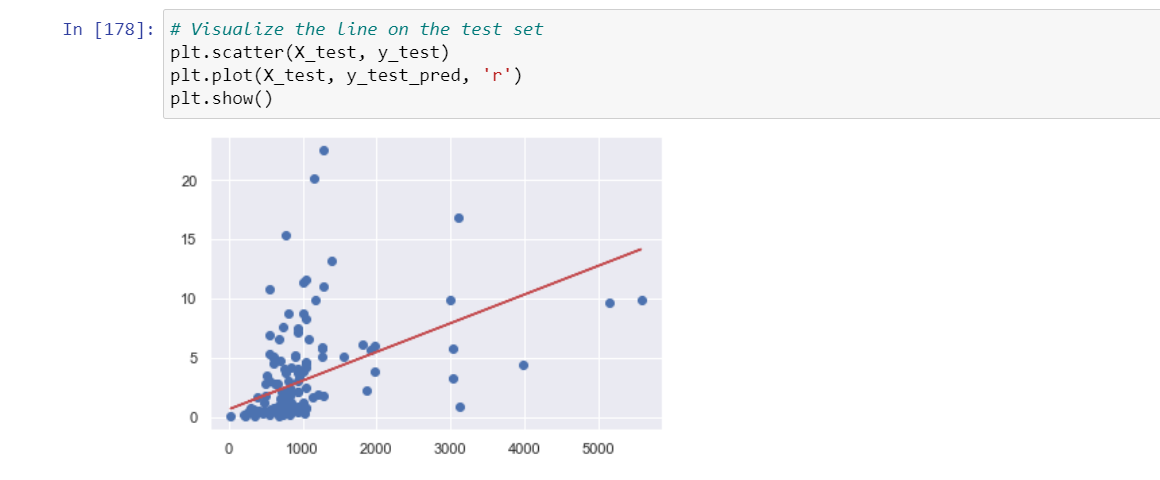


Linear regression line for the first model.

The predicted values are shown in the below picture   




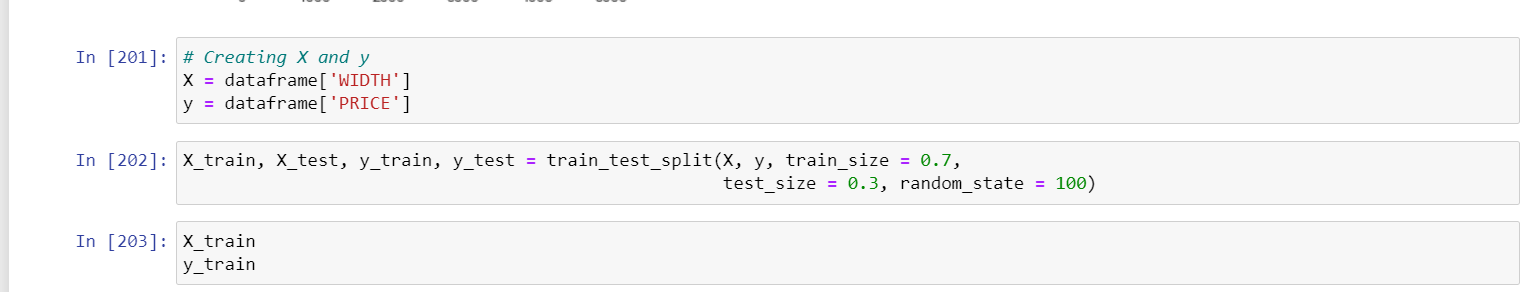
The r square value is 0.117 for the model 1.



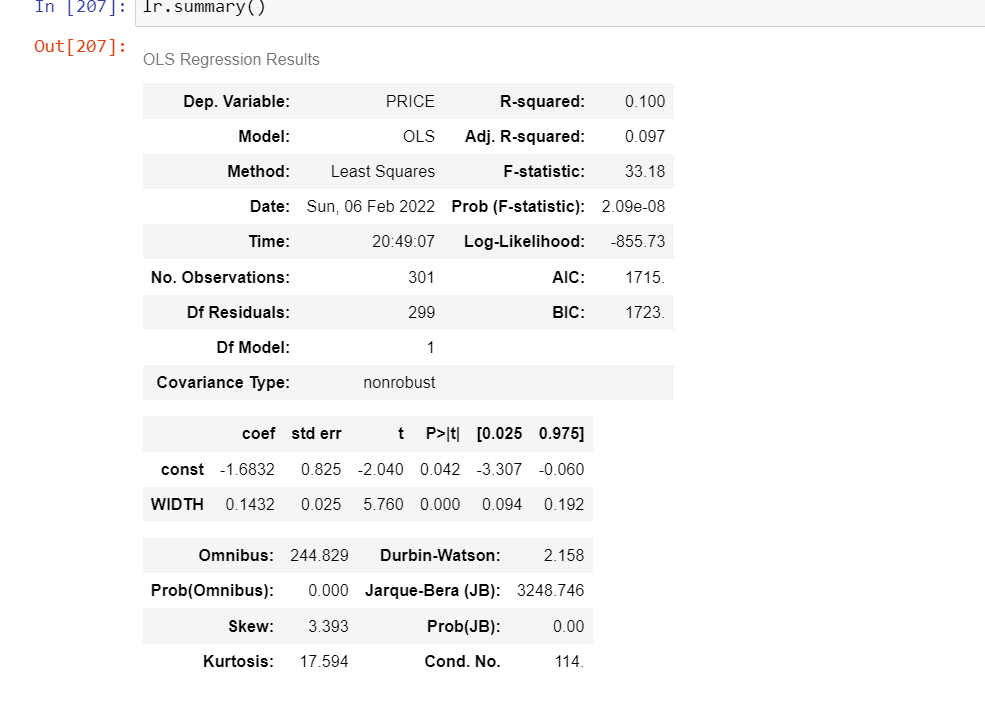
The above picture shows the visualization for the test set on the model 1.

Creating a simple linear regression model 2 for the monet data set

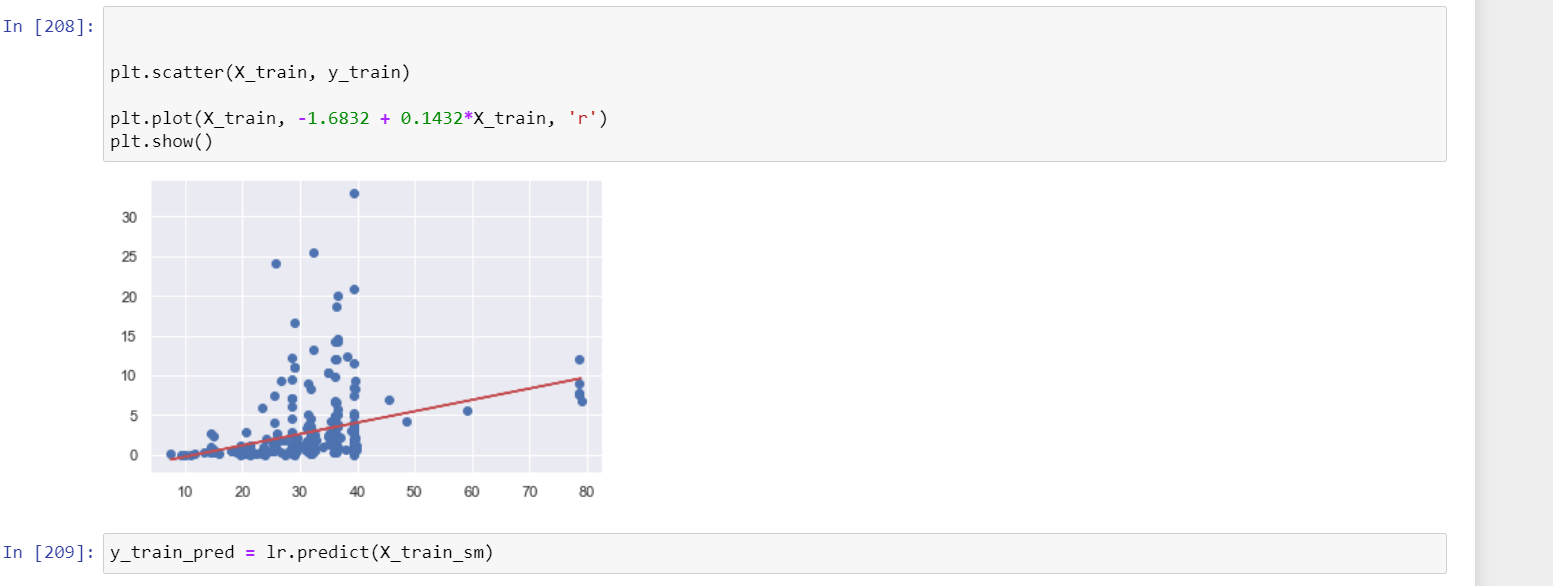
Using width as independent variable and price as dependent variable



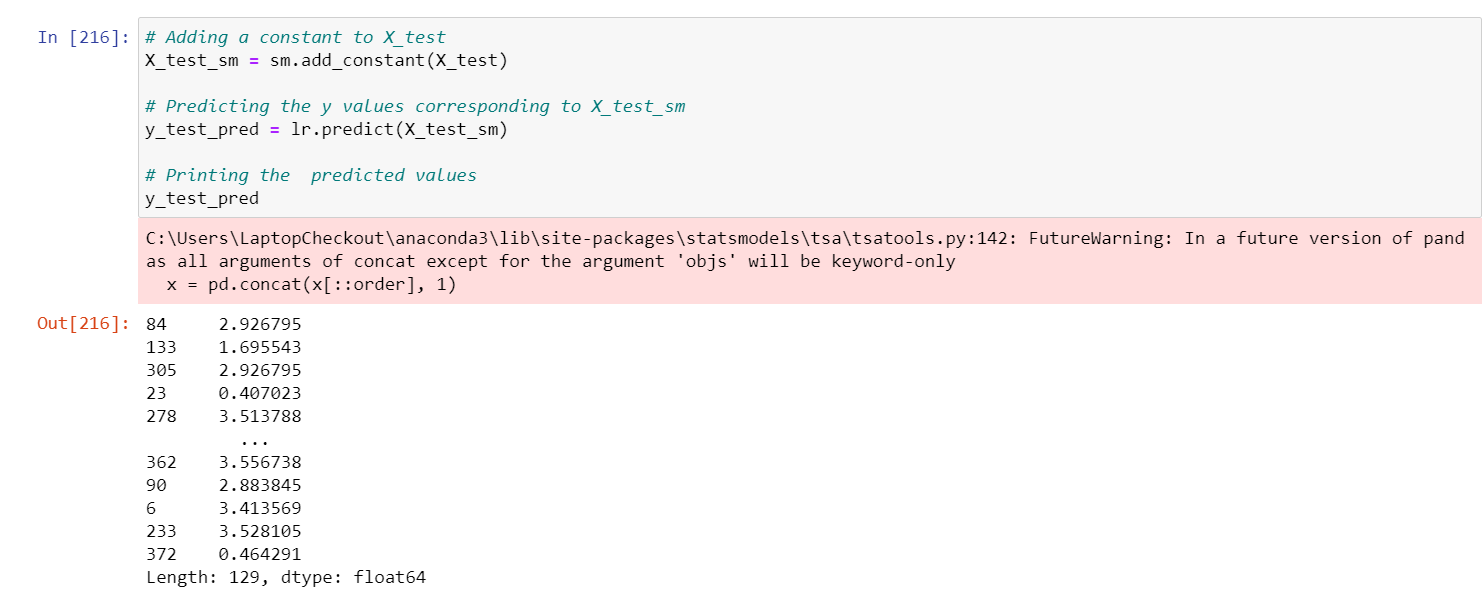
I have split the data set into 70-30 and the train size is of 70% and test size is 30%.



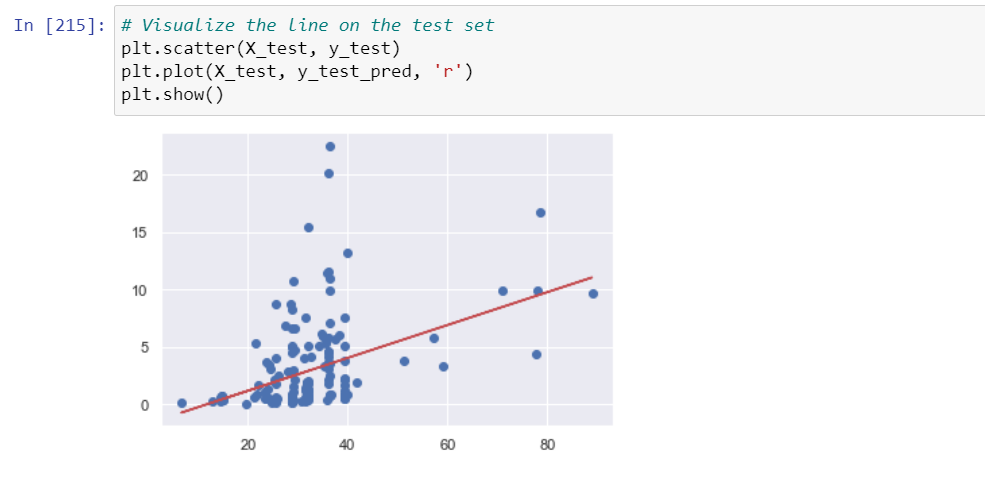
The above screen shot shows the summary of OLS regression linear regression model



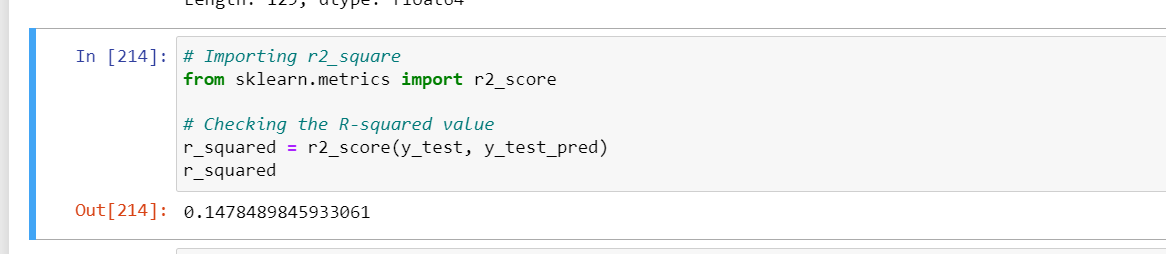
I have plotted the train plot which shows the linear regression line for train plot



The above screen shot shows the predicted values for the model 2 that is price.



The Above graph shows the linear regression line for the test data which was split.



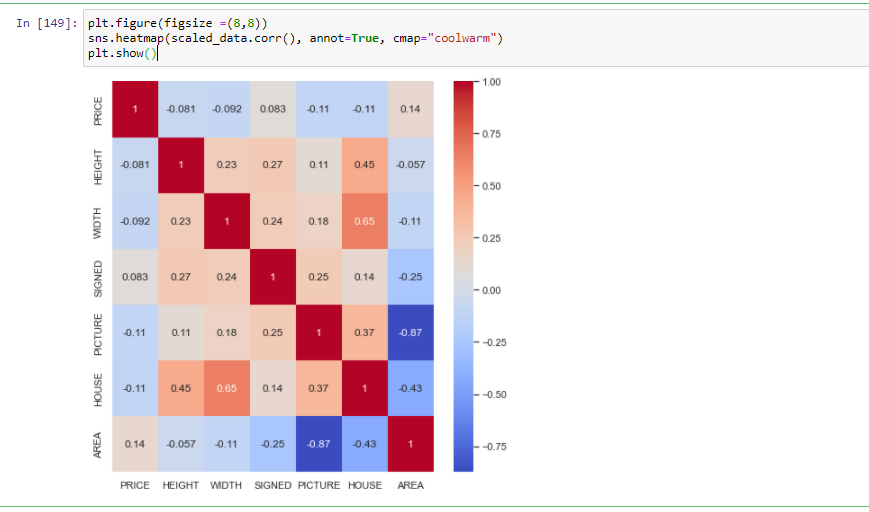
The Rsquare 0.1478 for the linear regression of model 2.

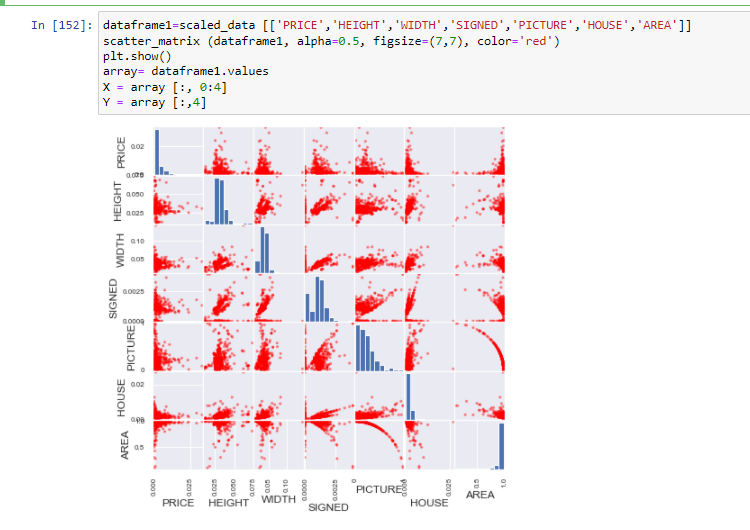
LINEAR REGRESSION OF MULTIVARIATE MODEL.

Normalisation of the model using preprocessing method.

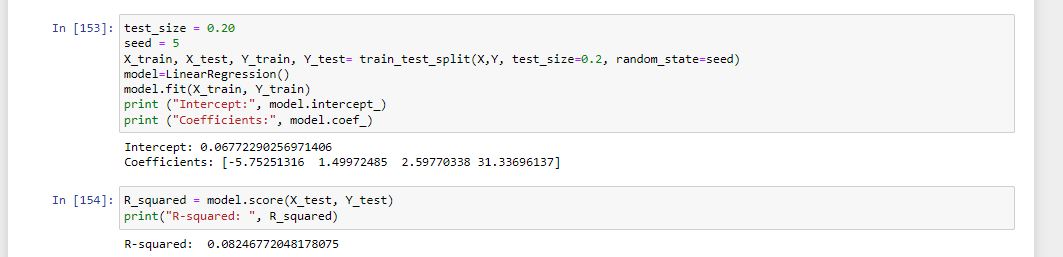
I have used the below code to normalize the below data in order to obtain the results







The above picture shows the pair plot for the multi variate linear regression model.



I have taken the test size for the data set as 20% and seed of 5.

The value of r square for multivariate linear regression model is 0.08246.

**Conclusion**:- Comparing all the 3 models we can say that the model with highest r value is considered to be best model , here in my case which is model 2 with Rsqaure value of 0.147 which is more than model 1 and multi linear regression model.