# Multiple Linear Regression (Module -7)

**Problem Statement: -**

With the growing consumption of Avacado, in USA, a freelance company would like to do some analysis on the patterns of consumption in different cities and also would like to come up with a prediction model of price for Avocado. For this to be implemented build a prediction model using multilinear regression and provide your insights on it.

Snapshot of the dataset is given below: -

There are five basic steps when you’re implementing Multiple linear regression:

1. Import the packages and classes you need.
2. Provide data to work with and eventually do appropriate transformations.
3. Create a regression model and fit it with existing data.
4. Check for collinearity between independent variables and between independent and dependent variables .
5. Check for overfitting issue
6. Apply the model for predictions.

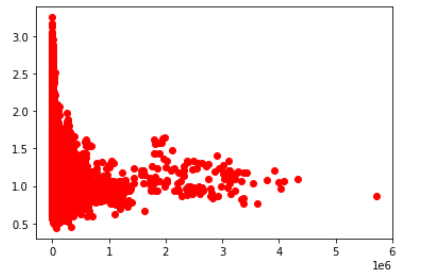
The first step is to import the package numpy and

import numpy as np

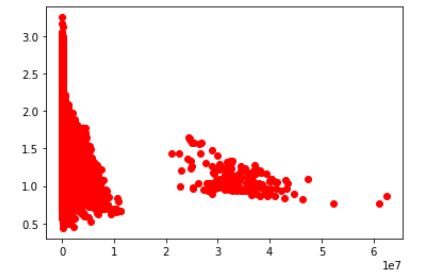
Now, you have all the functionalities you need to implement linear regression.

relavancy check of dependent variable with independent variables by scatter plot

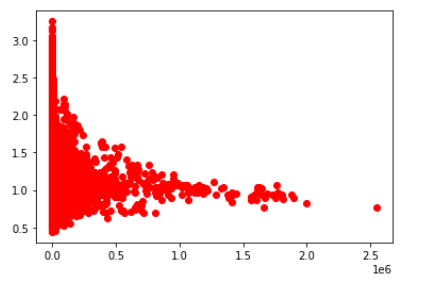
#scatterplot large bags vs average price



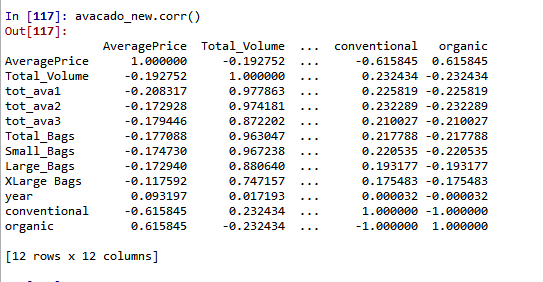
# scatter plot of total volume vs average price



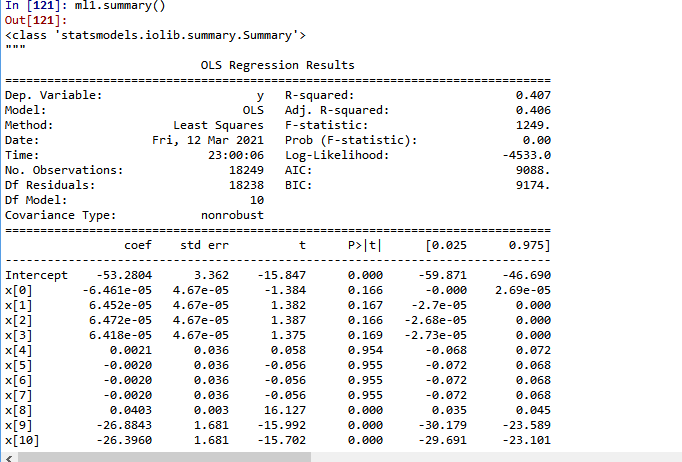
#scatterplot tot\_ava3 vs average price



correlation between variables is obtained by correlation matrix

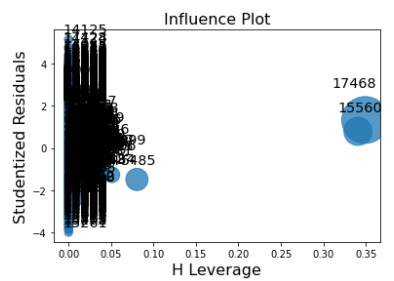


The class statsmodel.formula.api  will be used to perform linear regression and make predictions accordingly.



from statsmodels.api import sm

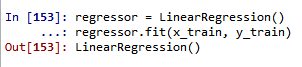
influence\_plot is plotted for model



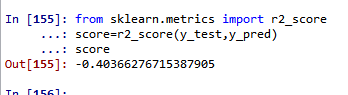
Data is splitted in to test and train with 20% test data



from sklearn.linear\_model import LinearRegression



r2\_score value for test data is obtained by



score is less than zero which means model is not good