# Multiple Linear Regression (Module -7)

**Problem Statement: -**

An online car sales platform would like to improve its customer base and their experience by providing them an easy way to buy and sell cars. For this, they would like an automated model which can predict the price of the car if user inputs the required factors. Help the business achieve the objective by applying multilinear regression on the given dataset.

Please use the below columns for the analysis purpose.

Price,Age\_08\_04, KM,HP,cc,Doors,Gears,Quarterly\_Tax,Weight

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.

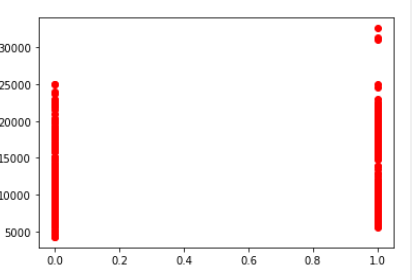
The first step is to import the package numpy and

import numpy as np

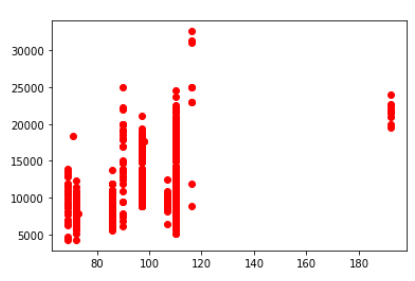
Now, you have all the functionalities you need to implement linear regression.

relavancy check of dependent variable with independent variables by scatter plot

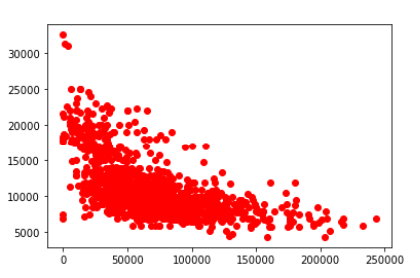
#scatterplot sport\_model vs price



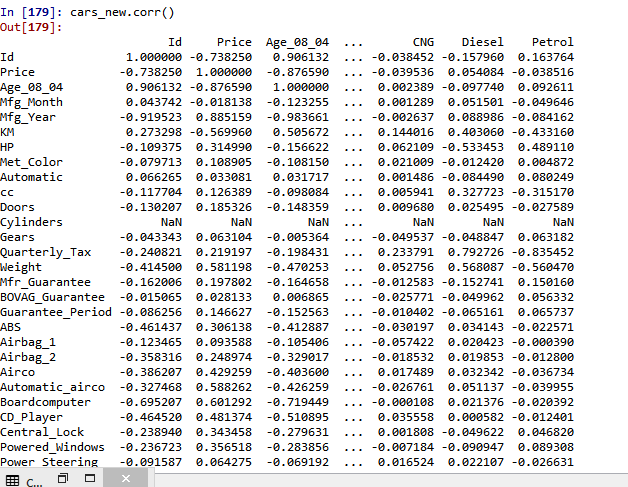
# scatter plot of HP vs price



#scatterplot KM price



correlation between variables is obtained by correlation matrix



from statsmodels.api import sm

influence\_plot is plotted for model

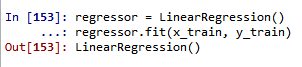
before building the model data is spllitted in to 20percent of test



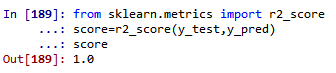
Data is splitted in to test and train with 20% test data



from sklearn.linear\_model import LinearRegression



r2\_score value for test data is obtained by



score is one which means model is good