NAME: MARTIN OLOO NGESA

INSTITUTION: AFRICA DATA SCHOOL

DATA SCIENCE AND MACHINE LEARNING PROJECT

PRICE PREDICTION OF USED VEHICLES

Introduction

The project is based on Linear regression to predict the price of used cars.

We would like to predict the price of used vehicles based on their specifications. The specifications provided in the data set include:

* Brand e.g. BMW, Mercedes.
* Body e.g. Sedan, Wagon.
* Mileage
* Engine Volume
* Engine Type e.g. Petrol, Diesel.
* Registration
* Year of manufacture
* Model

The first potential regressor is *Brand*, as it is known that a BMW is generally more expensive than a Toyota. The second relevant variable is *Mileage*, as the more a car is driven, the cheaper it should be. Third is the *Engine Volume.* Sports cars have larger engines than economy cars. The final variable is *Year of manufacture*. The older the car, the cheaper it is.

The rest are categorical variables which will be dealt with when need be.

The steps followed in this project are:

* Cleaning of the data set.
* Relaxing of the Ordinary Least Square assumptions.
* Log transformation of feature variables.
* Creation of a model.
* Use of dummy variables
* Training and testing the data
* Making predictions