Devise a way to reduce CO2 EMISSIONS of future cars from a dataset which shows the characteristics of cars produced in the same year

Author: Jimoh Abdulbasit Adedamola

Introduction .

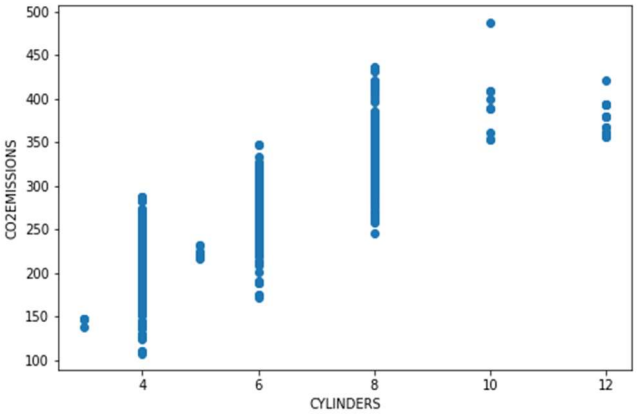
This is a Data Science project to use a dataset of cars produced in the year 2014 to create a model to predict the CO2EMISSIONS of other cars in the same year.

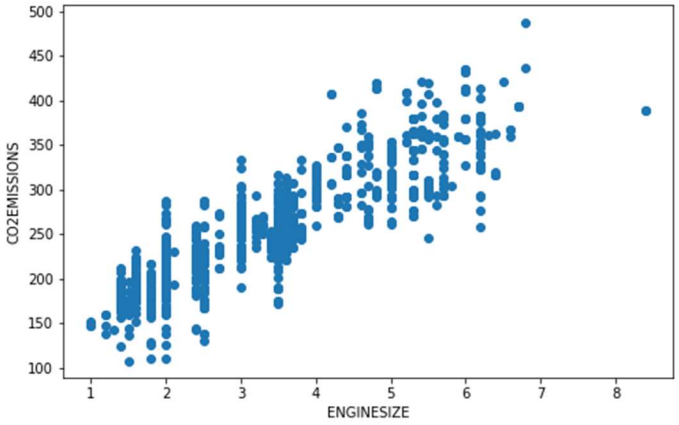
The dataset comprised of 1067 rows and 13 columns, 5 of which are objects and were not used in the analysis. Out of the other 8, 4 are integers while the other four are decimals. A regressional analysis was done on the dataset and I was able to find a relationship between the independent variable and the dependent variable(CO2EMISSIONS), then I chose the independent variable with lowest mean absolute error to build my model.

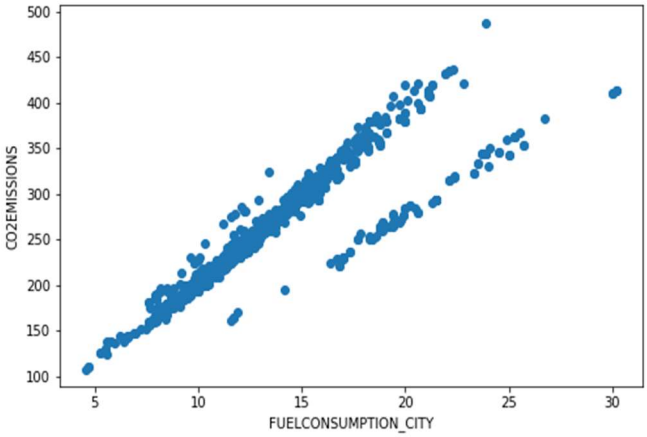
Processing .

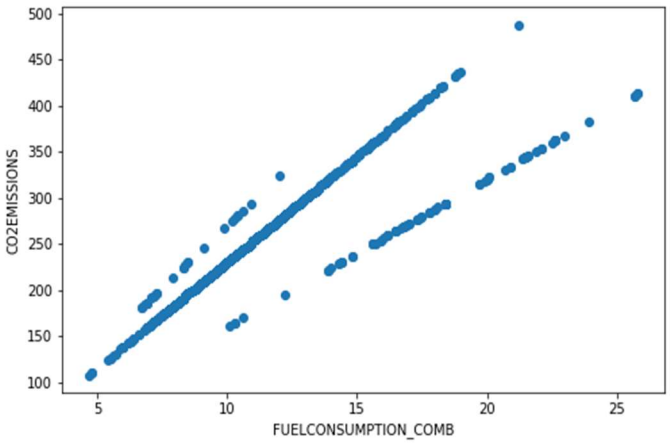
I began by cleaning the dataset, I separated the object type data from the integers and decimals so I could focus on the lather. And I also removed the model year since that was a common feature to all the cars in the dataset.

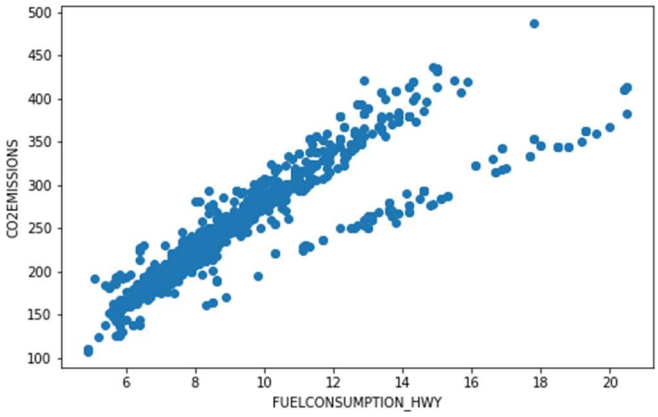
After cleaning I went straight to draw a scatter plot of all the independent variable against the dependent variable

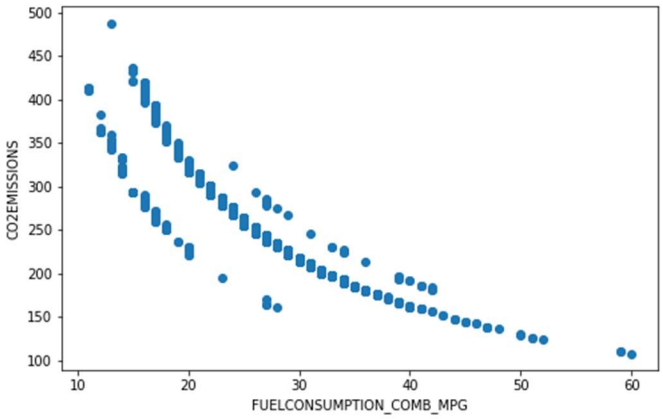




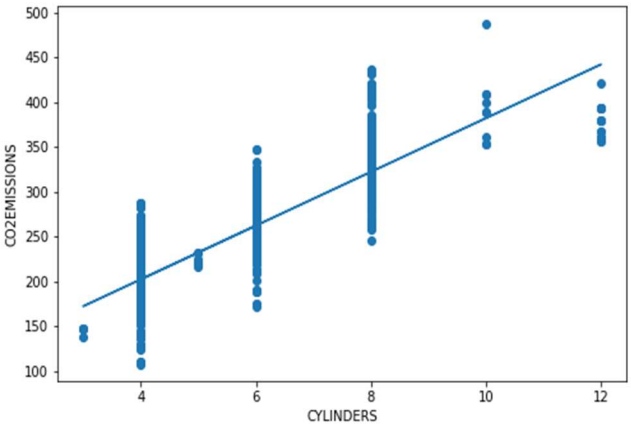


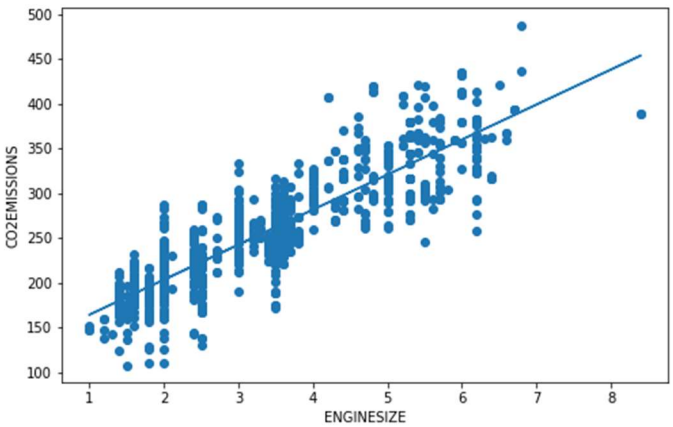


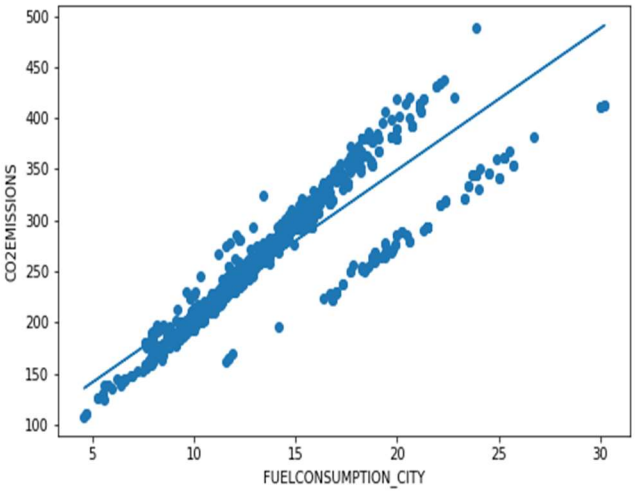


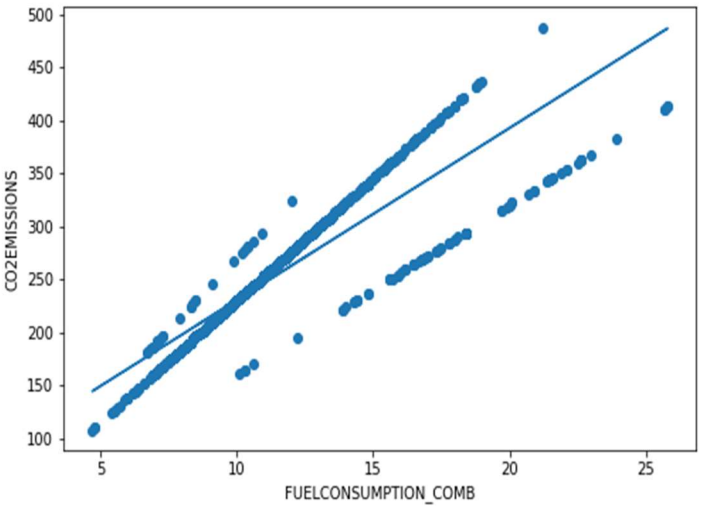


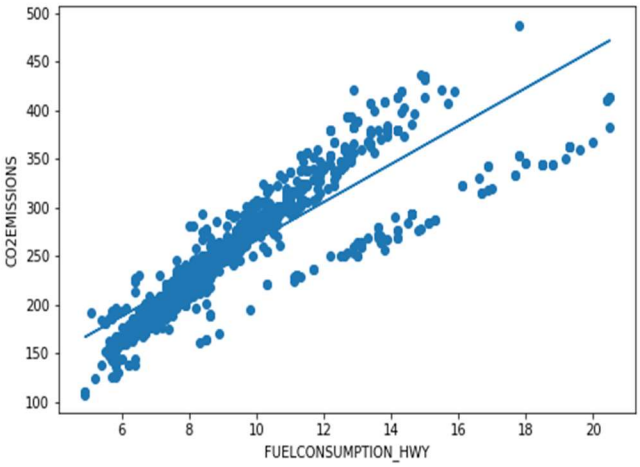
From the Scatter plot we see that 5 out of the 6 have a linear shape so we are going to take a linear approach then the last one has a curve shape which means we are going to take a polynomial approach. Now we plot our line on the scatter plot

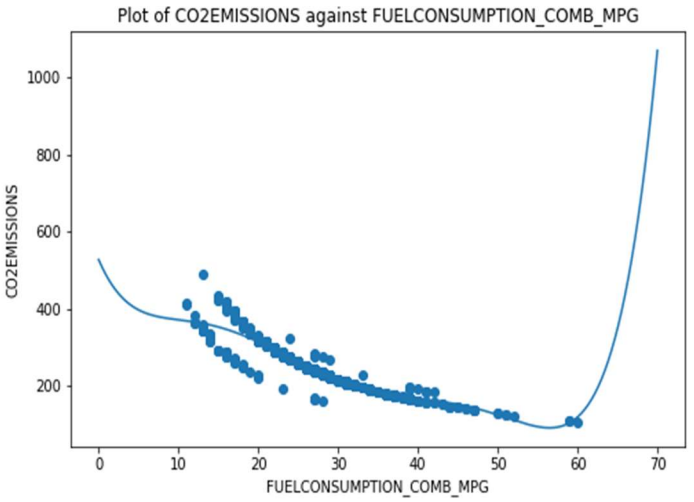












After analyzing the graphs, I find out that the polynomial graph is the one with the closest relationship and the one with the best closeness of data. Then I analyzed the polynomial graph with a machine learning model by training it with part of the data given to us so we will be able to test it with the remaining data given to us and we are going to get the relationship. I then got the relationship to be over 80%.

To be further sure I calculated the mean Absolute Error. I got a mean absolute error of 11.843 which is satisfiable for a dataset of just above 1000 rows.

Conclusion .

I started by cleaning then processed it to get a relationship that is above 80%, then I created my model. After processing I found out that the bigger the engine size of the car the more CO2 it emits. And also, the more fuel the car consumes the more CO2 it emits because CO2 is emitted during combustion in a car, and the combustion is caused by the usage of fuel in the car. So, the relationship between the engine size of the car, the fuel consumed by the car and the CO2 emitted by the car is directly proportional that is, the bigger the engine size of the car the more fuel the car consumes and the more fuel the car consumes means the car is going to emit more CO2.

So we can see that to reduce the CO2 emissions of the future cars we have to reduce the size of the engines.