# Prediction of CO2 Emissions

Poojaa Ravindra Guptha - U1083384

# INTRODUCTION

Since it is a vital metric for understanding how well we are doing in slowing down global climate change, the CO2 Prediction is performed out in this project. There are various attributes that increase the accuracy of the model. The prediction is done using Multiple linear regression.

# DATA COLLECTION

The dataset that I have used in this project is Carbon dioxide emission based on various attributes. Totally there are 500 rows with 11 input columns and 1 output column. The input columns of the model are

1. Make
2. Model
3. Vehicle Class
4. Engine Size(L)
5. Cylinders
6. Transmission
7. Fuel Type
8. Fuel Consumption City (L/100 km)
9. Fuel Consumption Hwy (L/100 km)
10. Fuel Consumption Comb (L/100 km)
11. Fuel Consumption Comb (mpg)

# DATA PREPROCESSING

**Removing Null values:**

The datasets are cleaned by removing null, nan values.

**Readable format:**

The column names of the datasets are changed into an understandable format.

**Label Encoding:**

Finding the numeric and non-numerical values from the datasets. The non-numerical values are changed to numerical values by performing Label encoding.

**Outlier Detection:**

The outliers are detected for all the attributes in the datasets.

# DATA VISUALIZATION

**Box Plot:**

The box plot is constructed to visually represent the attributes of the data.

**Bar Plot:**

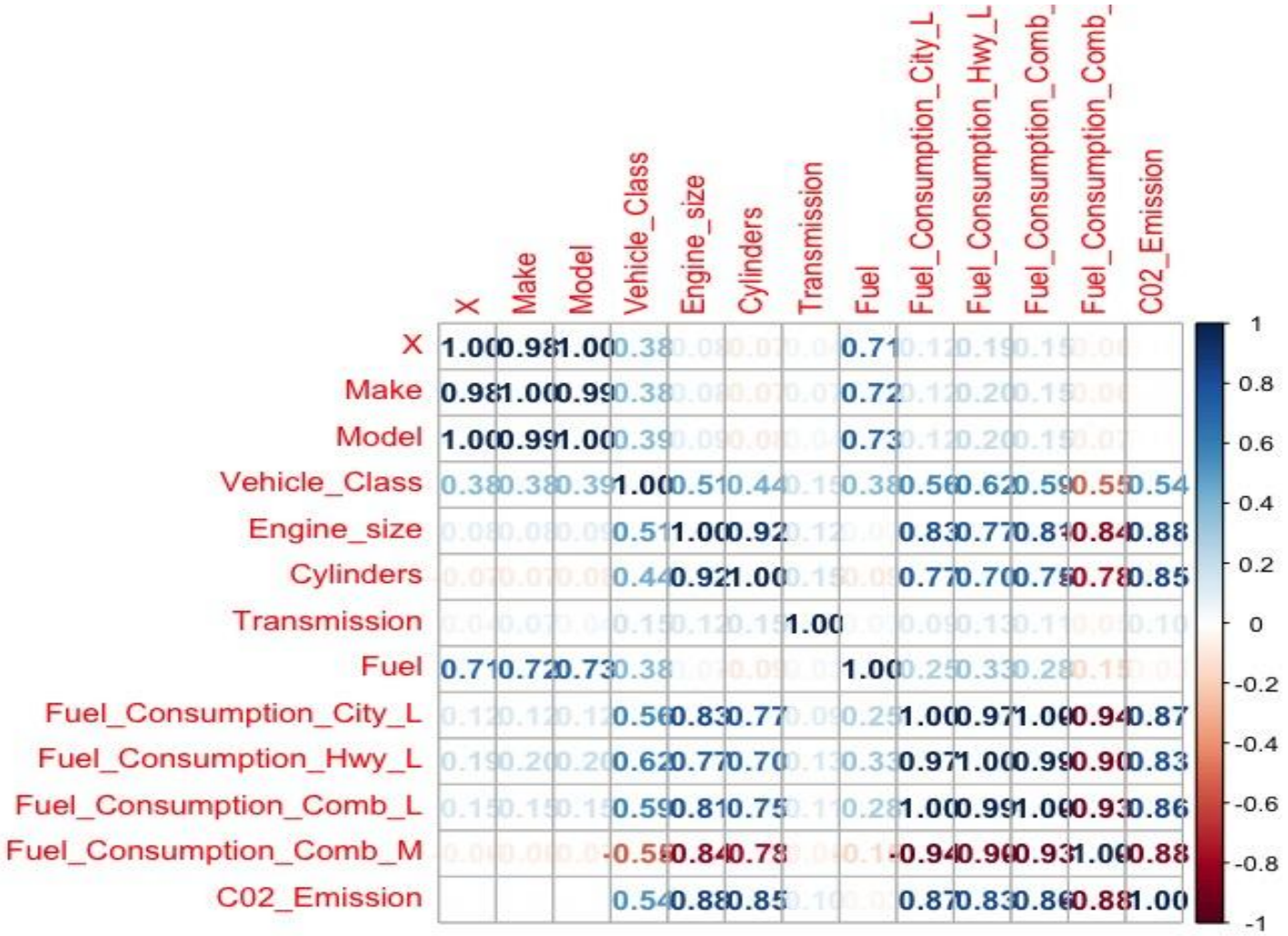
The bar plot is created to find the frequency of the attributes in the dataset.

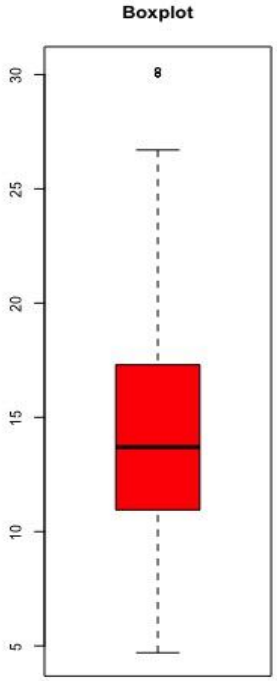
**Correlation matrix:**

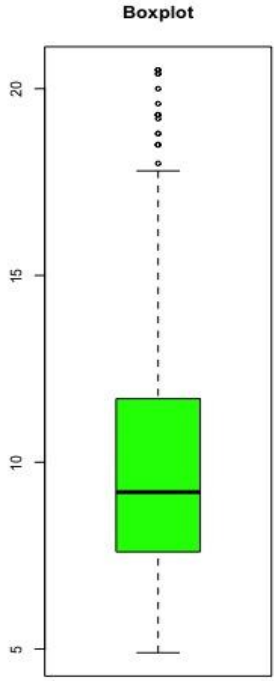
The correlation matrix graphically represents the features that are important for the prediction.

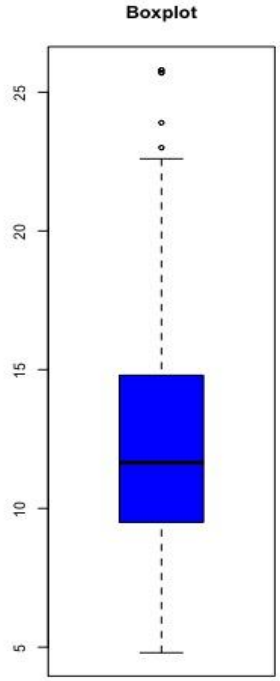
**Pair-wise plot:**

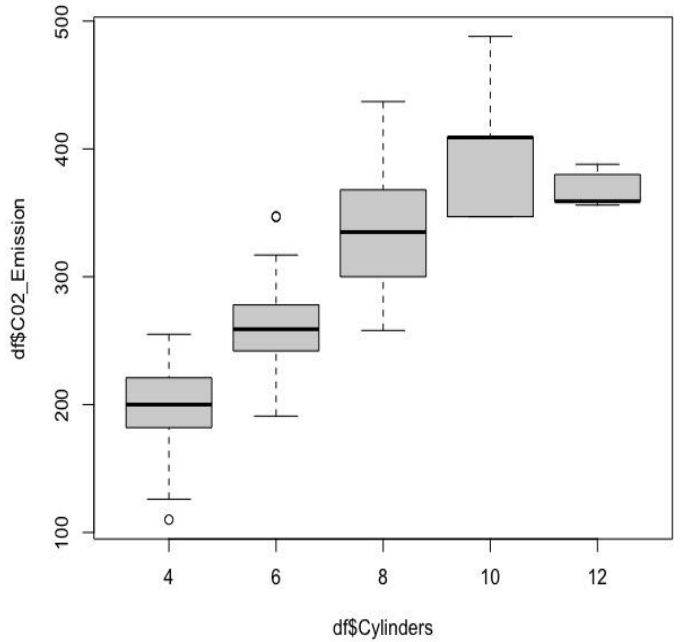
The pair-wise plot represents the relationship between each and every attributes in the dataset.

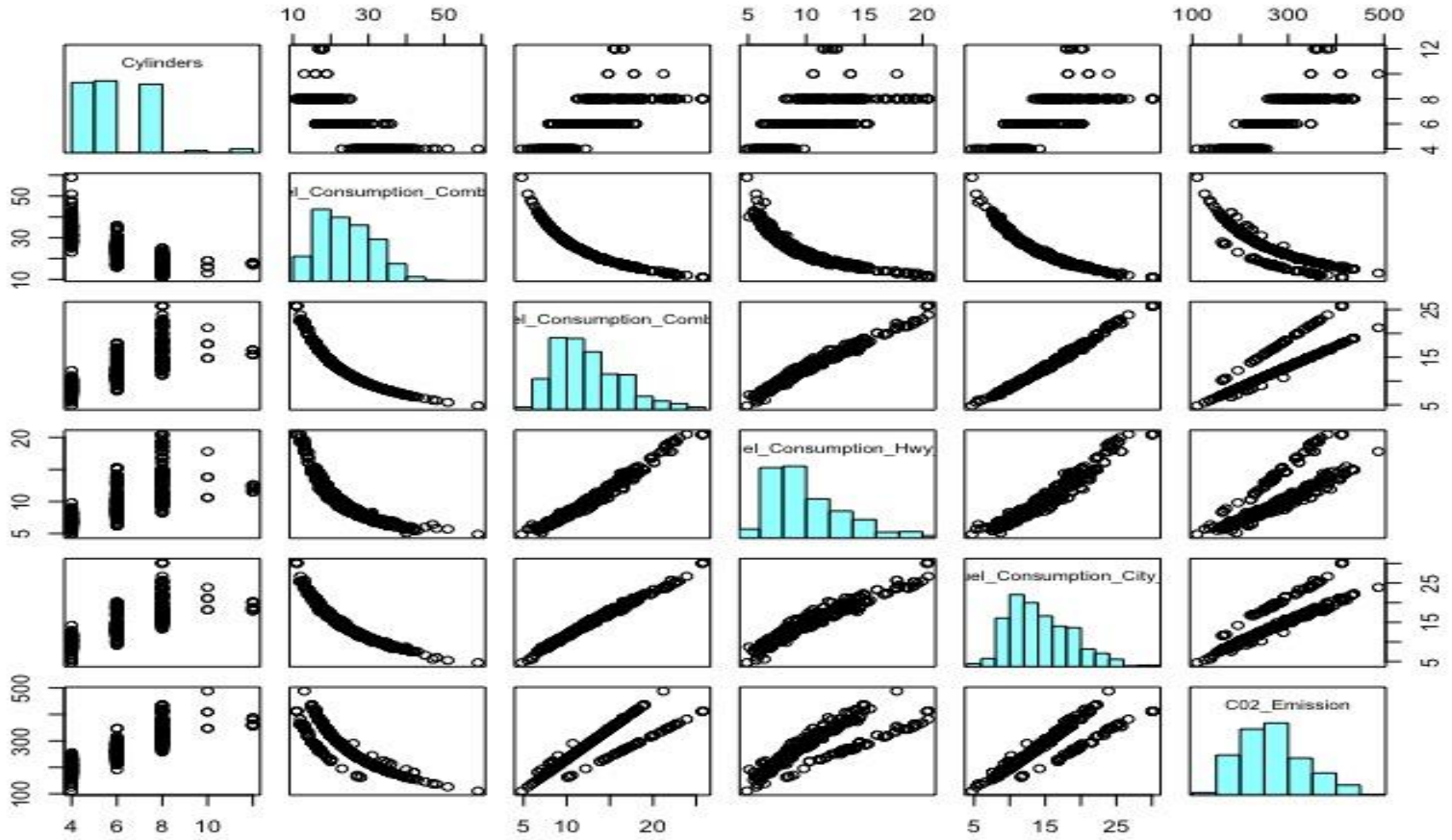

X
Make
Model
Vehicle_Class
Engine_size
Cylinders
Transmission
Fuel
Fuel_Consumption_City_L
Fuel_Consumption_Hwy_L
Fuel_Consumption_Comb_
Fuel_Consumption_Comb_
C02_Emission
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1


Boxplot
30
25
20
15
10
5


Boxplot
20
15
10
5


Boxplot
25
20
15
10
5


df$C02_Emission
500
400
300
200
100
4
6
8
10
12
df$Cylinders


Cylinders
el_Consumption_Comb
el_Consumption_Comb
el_Consumption_Hwy
el_Consumption_City
C02_Emission

# CONCLUSION

Thus the prediction of CO2 Emissions is performed using Multiple linear regression. This projects includes various data visualization techniques to understand the dataset in a correct manner. Also anova testing is performed. Finally the accuracy of the dataset is 88%.

# THANK YOU