Analyzing Vehicle Emissions: Exploring Factors Affecting Environmental Impact

Joby Varghese

26 July 2023

Outline

- Business Problem
- Data
- Methods
- Results
- Conclusions

Disclaimer: *This project utilizes fictitious identities for individuals and institutions, including the names "NSW Transport" and "Blacktown Council," which are purely for academic purposes. It is important to note that neither NSW Transport nor Blacktown Council has any involvement or role in this project, as their names are solely used to provide context and support for the academic project.*

Business Problem

- As part of the National Policy toward Carbon Offset, the Blacktown Council and NSW Transport are very keen on implementing strategies that will reduce vehicle emissions considerably.

- The Council aims to develop a predictive model capable of accurately forecasting emissions from various vehicle types and classes. This proactive approach will empower the Council to implement effective measures to combat carbon emissions and promote sustainability.

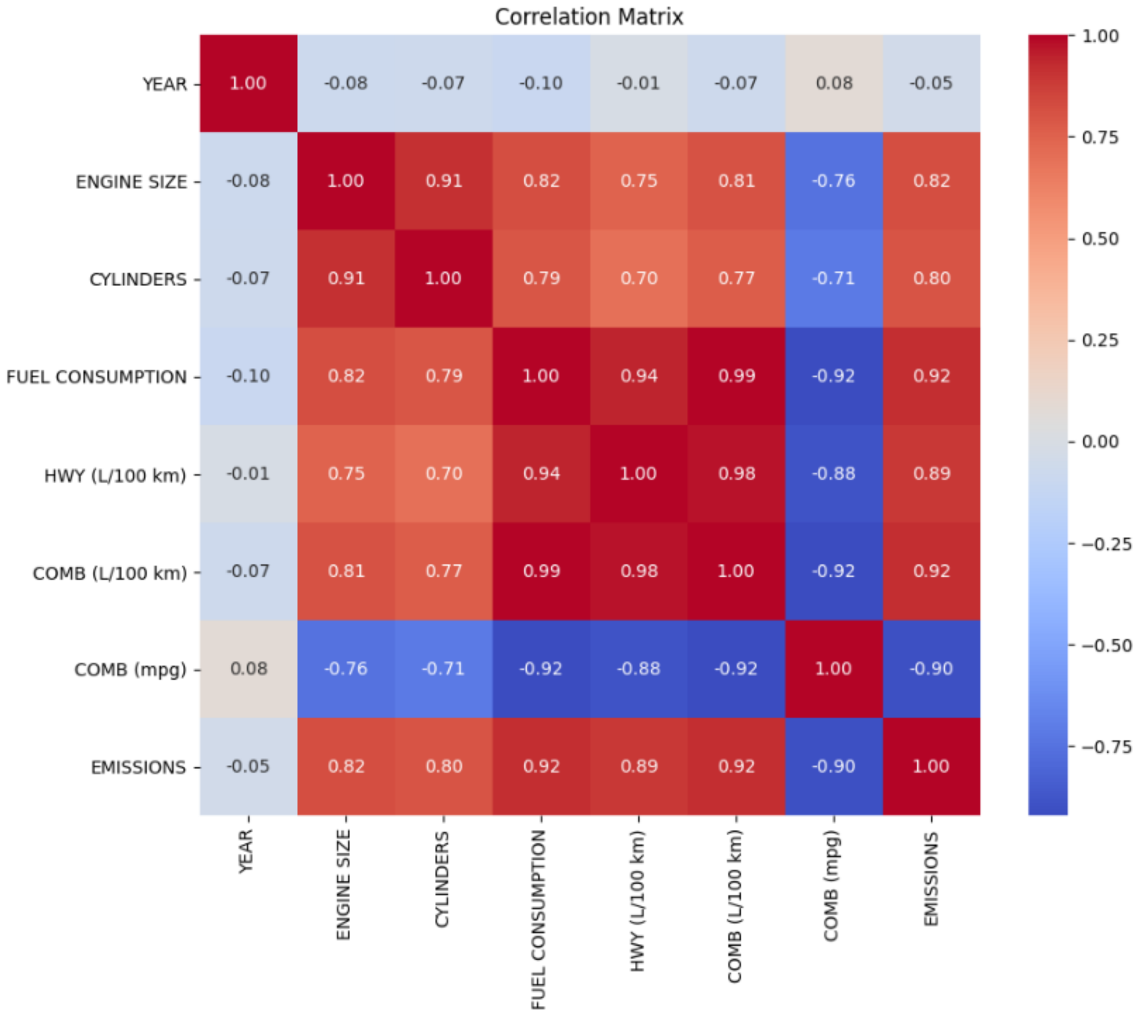
Data

- The analysis utilized a dataset provided by the NSW Transport for the Blacktown Council, containing information on vehicles registered within the council's jurisdiction from 2000 to 2022.
- The analysis focused on several significant variables, including:
  1. Vehicle Class: This variable indicates the category to which the vehicle belongs.
  2. Engine Size: It represents the size of the vehicle's engine.
  3. Cylinders: The number of cylinders in the vehicle.
  4. Transmission: This variable indicates whether the vehicle has an automatic or manual transmission.
  5. Fuel: It specifies the type of fuel used by the vehicle.
  6. Mileage: The mileage of the vehicle per gallon.
  7. Emissions: This variable represents the emissions produced by the vehicle.

Methods Used

- Data Preprocessing: The dataset was thoroughly examined for any missing values.
- Correlation Analysis: To understand the relationships between the dependent variable (Emissions) and the independent variables, a correlation matrix was generated. This analysis provided valuable insights into the factors influencing Emissions and helped identify strong and weak correlations.
- Feature Selection: To simplify the model and remove potential noise or redundancy, certain variables such as ‘Year', ‘Make', ‘Mode’, and ‘Fuel Consumption’ were dropped from the dataset.
- Transformation Techniques: The dataset underwent multiple transformations at different stages to enhance the model's performance. Techniques like log transformations, dummy variable encoding, and normalization were applied to establish more accurate relationships between predictor variables and the target variable, resulting in improved predictions.
- Following the outlined procedure, a Linear Regression Model was constructed. The model was rigorously tested to assess its adherence to the assumptions of residual normality and linearity in regression analysis. Additionally, to validate the model's performance, K-fold cross-validation was performed. The dataset was divided into training and testing sets for this purpose.

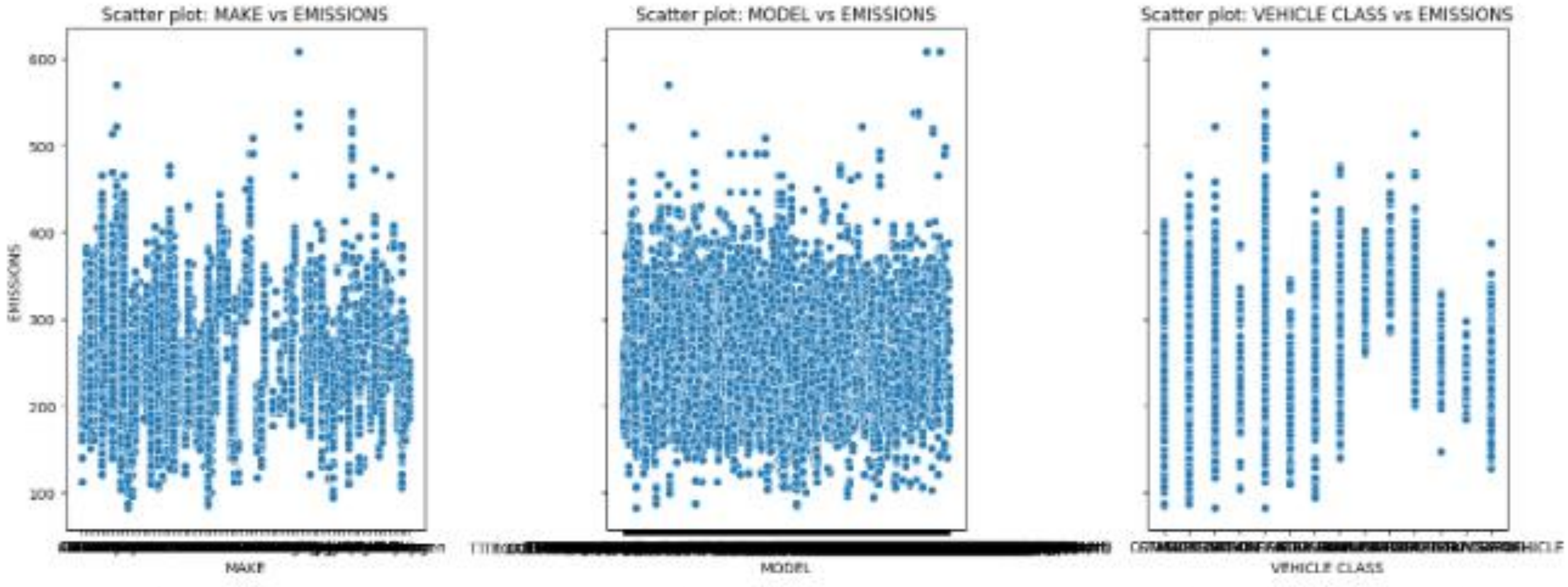
EDA



Observations:

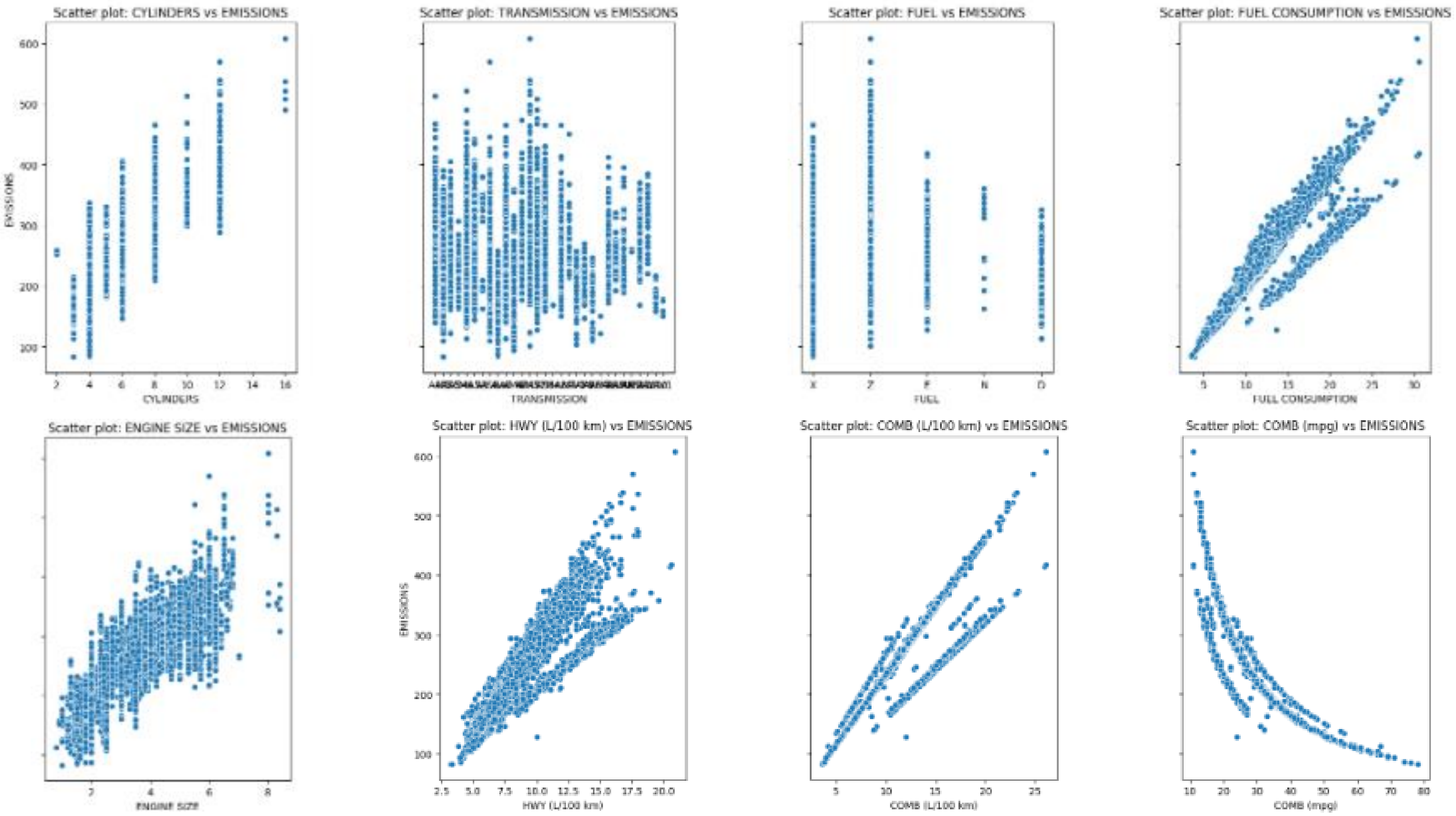
- Year exhibits weak correlation with other variables.
- All the other numeric variables shows positive correlation with Emission except for COMB (mpg)

EDA



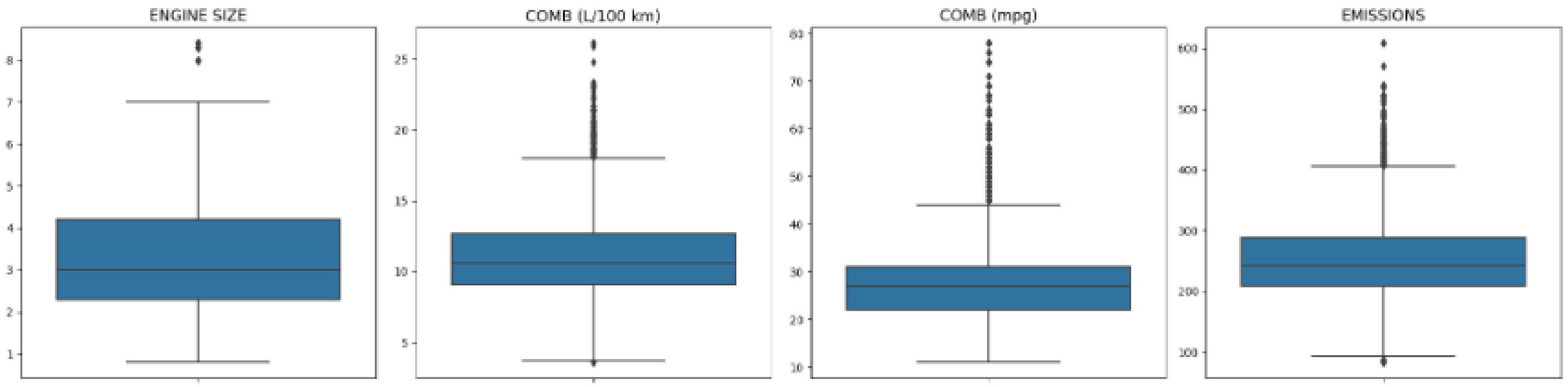
- **'MAKE' and 'MODEL' predictor variables did not exhibit a linear relationship with 'EMISSIONS'**

EDA

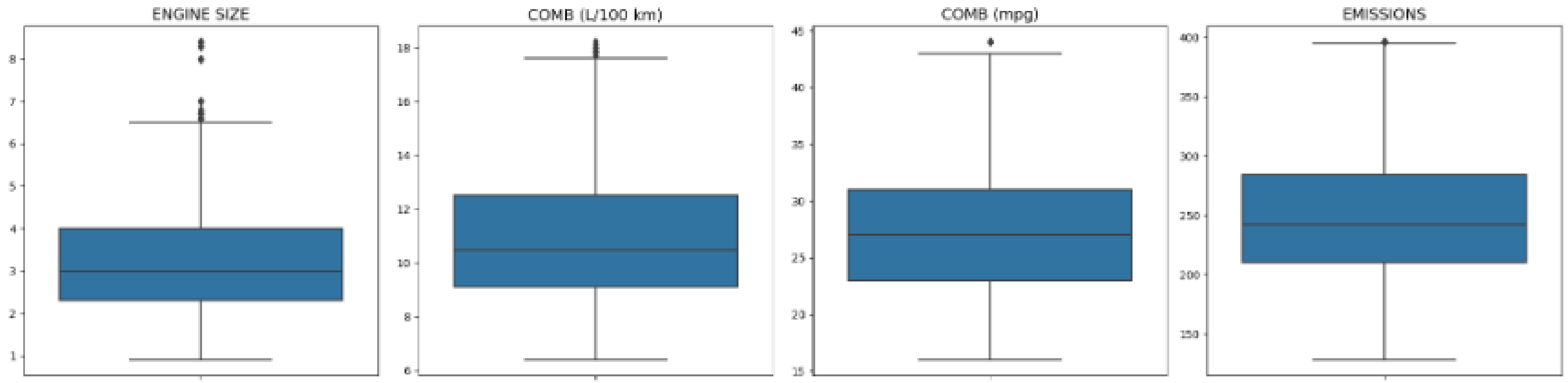

Scatter plot: CYLINDERS vs EMISSIONS
EMISSIONS
CYLINDERS
Scatter plot: TRANSMISSION vs EMISSIONS
TRANSMISSION
Scatter plot: FUEL vs EMISSIONS
FUEL
Scatter plot: FUEL CONSUMPTION vs EMISSIONS
FUEL CONSUMPTION
Scatter plot: ENGINE SIZE vs EMISSIONS
ENGINE SIZE
Scatter plot: HWY (L/100 km) vs EMISSIONS
EMISSIONS
HWY (L/100 km)
Scatter plot: COMB (L/100 km) vs EMISSIONS
COMB (L/100 km)
Scatter plot: COMB (mpg) vs EMISSIONS
COMB (mpg)

Handling Outliers

Before



After



Linear Regression Model

```
                            OLS Regression Results
==============================================================================
Dep. Variable:              EMISSIONS   R-squared:                       0.796
Model:                            OLS   Adj. R-squared:                  0.795
Method:                 Least Squares   F-statistic:                     2030.
Date:                Sun, 23 Jul 2023   Prob (F-statistic):               0.00
Time:                        20:59:19   Log-Likelihood:                 22461.
No. Observations:               21415   AIC:                        -4.484e+04
Df Residuals:                   21373   BIC:                        -4.450e+04
Df Model:                          41
Covariance Type:            nonrobust
============================================================================================================
                                              coef    std err          t      P>|t|      [0.025      0.975]
```

K-Fold Validation Score

```
Mean Train Mean Squared Error: 0.007182024730219057
Mean Test Mean Squared Error: 0.007225135297286781
```

- The adjusted R-squared is 0.795.
- The F-statistic is 2030, and the Prob (F-statistic) is very close to zero (0.00), indicating that the model is statistically significant.
- All the coefficients have P-values close to zero, indicating that they are statistically significant, which means there is a real and meaningful relationship between the variables being studied, and the observed pattern or effect is not just due to random variability in the data.
- The [0.025 0.975] values associated with each coefficient represent the 95% confidence interval.
- The K-Fold Validation Score suggest that the linear regression model is performing well on both the training and testing datasets.
- MSE values for both the training and testing data are relatively close to each other for each train-test split. This suggests that the model is not overfitting.

Observations

- **Cylinder Count**: According to the coefficients for the different cylinder counts, it is evident that fewer cylinders generally result in lower emissions. As a result, vehicles with smaller engines (fewer cylinders) tend to produce fewer emissions. It is possible to reduce emissions by encouraging the use of smaller, more fuel-efficient engines.
- **Fuel Type**: Vehicles using certain fuel types (such as Fuel E - Ethanol E85 & Fuel N - Natural Gas) produce fewer emissions than those using others (such as Fuel Z - Premium Gasoline). It is possible to significantly reduce emissions by promoting the use of cleaner fuels or alternative energy sources, such as electric or hybrid vehicles.
- **Transmission Type**: Based on the coefficients for different transmission types (such as Transmission A3, Transmission A6, Transmission M5), specific transmission technologies may influence emissions. Emissions can be reduced with advanced transmissions (such as automatics with more gears or modern continuously variable transmissions).
- **Vehicle Class**: As shown by the coefficients for different vehicle classes, different vehicle classes have varying impacts on emissions. In comparison to smaller and more compact vehicles, SUVs and pickup trucks produce more emissions. The adoption of fuel-efficient vehicles, such as subcompacts and hybrids, could contribute to the reduction of emissions.
- **Engine Size**: Larger engine sizes produce higher emissions, according to the coefficient for Engine Size. Emissions can be reduced by promoting vehicles with smaller, more efficient engines.

To Summarize...

- Smaller engines and fewer cylinders should be encouraged, especially in urban areas where fuel efficiency is of greatest importance.
- Using cleaner fuels or alternative energy sources, such as electric or hybrid vehicles, can significantly reduce emissions.
- Improve fuel efficiency and reduce emissions through the development and use of advanced transmission technologies.
- Provide incentives for fuel-efficient vehicle purchases and discourage the use of high-emission vehicles in commercial and public fleets.
- Research and develop new technologies to improve fuel efficiency and emissions reduction in vehicles.

Thank you

Joby Varghese

Email: grace.joby@gmail.com

GitHub: @VargheseJoby

LinkedIn: linkedin.com/in/joby-varghese-516