- **Main objective of the analysis that specifies whether your model will be focused on prediction or interpretation.**

The dataset I have chosen is a car and its fuel consumption in different conditions. So, for this I am going to predict what is going be the carbon dioxide emitted from the particular vehicle

- **Brief description of the data set you chose and a summary of its attributes.**

- *MODELYEAR* e.g. 2014 – made year
- *MAKE* e.g. Acura - manufacturer
- *MODEL* e.g. ILX – model name
- *VEHICLE* CLASS e.g. SUV – model type
- *ENGINE SIZE* e.g. 4.7
- *CYLINDERS* e.g 6
- *TRANSMISSION* e.g. A6
- *FUEL CONSUMPTION in CITY (L/100 km)* e.g. 9.9
- *FUEL CONSUMPTION in HWY (L/100 km)* e.g. 8.9
- *FUEL CONSUMPTION COMB (L/100 km)* e.g. 9.2
- *CO2 EMISSIONS (g/km)* e.g. 182 --> low --> 0

- **Brief summary of data exploration and actions taken for data cleaning and feature engineering.**

Based on the corelations, 'ENGINESIZE','CYLINDERS','FUELCONSUMPTION_COMB','CO2EMISSIONS' are taken as our final features and are visualised to analyse. The data set is mostly cleaned and only needs to be scaled. The above-mentioned columns are in number format. So, they are not encoded.

- **Summary of training at least three linear regression models which should be variations that cover using a simple linear regression as a baseline, adding polynomial effects, and using a regularization regression. Preferably, all use the same training and test splits, or the same cross-validation method.**

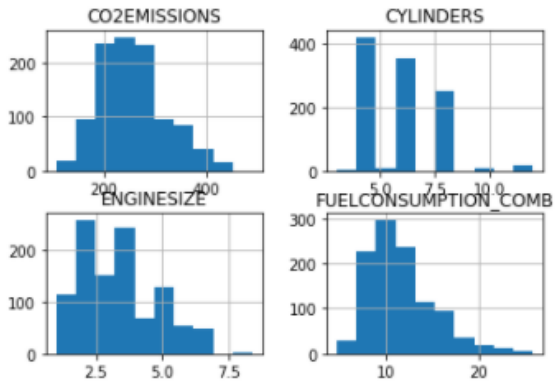
The model is trained on multiple linear regression, polynomial regression and adding a regularization term. The model performed best on the multiple linear regression, next polynomial regression with regularization and worst on polynomial.

- **A paragraph explaining which of your regressions you recommend as a final model that best fits your needs in terms of accuracy and explainability**.

I would definitely choose to using the multiple linear regression since, it is more simpler compared to other models and also it performed well while predicting.

- **Summary Key Findings and Insights, which walks your reader through the main drivers of your model and insights from your data derived from your linear regression model.**

The emissions are mostly based on engine type and the fuel it is consuming while driving. Mostly, vehicles with bigger engines and cylinders are producing more co2 compared to small engines.



- **Suggestions for next steps in analysing this data, which may include suggesting revisiting this model adding specific data features to achieve a better explanation or a better prediction.**

The model can achieve more performance and more in depth review if more data about the other vehicles is collected and weather it is a diesel or petrol or ev can also be useful for more analysing of our model.

By,

SAI GOWTHAM BABU AMBURI