Coursera Capstone Project

FINAL PROJECT

(By Gebremedhin Melaku)

Is your commute eco-friendly?

A model to determine SmartWay vehicles at purchase

Introduction

Global warming and climate changes are existential threats to human race. The rate at which warming takes place is much faster than the Permian-Triassic extinction warming event. Extreme warming events always involve the same invisible culprit, one we're familiar with today: A massive dose of carbon dioxide, or CO_2. With this rapid growth in technology; rapid industrialization and the race for being economic superpower, CO_2 emission had increased exponentially.

The number of motorized vehicles (cars and trucks) in 2015 were around 1.5 billion. The number is expected to climb to more than 2 billion by 2025[2,3,4]. With the number of vehicles increasing at this rate, it is recommended to drive environmentally safe vehicles. In this project, the 2020 office of Energy efficiency & renewable energy guide data (https://fueleconomy.gov/feg/download.shtml) will be used to create a model predicting if a given vehicle falls within the category of SmartWay vehicle.

Greenhouse Gas Score and Air Pollution Score (smog score) are the key variables in determining if your drive is "green vehicle". With this era of growing traffic and CO2 emission, it is advisable to use vehicles that qualify for SmartWay. This project will train a model to predict if a given vehicle is SmartWay vehicle based on various independent variables.

Coursera Capstone Project

Data

For this project 2020 fuel economy data from the office of Energy efficiency & Renewable energy were used (https://fueleconomy.gov/feg/download.shtml).The data has 18 columns (variables) with 2513 observations(rows).

The first 10 rows of the data will look like the following.

Table 1: The first 10 rows of the original dataset

R2527

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	Model	Dis	Cy	Trans	Driv	Fuel	rt Region	Stnd	Stnd Description	Underhood ID	Veh Class	Air Pollution Score	City MPG	Hwy MPG	Cmb MPG	Greenhouse Gas Score	SmartWay	Comb CO2
2	ACURA ILX	2.4	4	AMS-8	2WD	Gasoline	CA	L3ULEV125	California LEV-III ULEV125	LHNXV02.4KH3	small car	3	24	34	28	6	No	316
3	ACURA ILX	2.4	4	AMS-8	2WD	Gasoline	FA	T3B125	Federal Tier 3 Bin 125	LHNXV02.4KH3	small car	3	24	34	28	6	No	316
4	ACURA MDX	3	6	AMS-7	4WD	Gasoline	CA	L3ULEV125	California LEV-III ULEV125	LHNXV03.0ABC	small SUV	3	26	27	27	6	No	333
5	ACURA MDX	3	6	AMS-7	4WD	Gasoline	FA	T3B125	Federal Tier 3 Bin 125	LHNXV03.0ABC	small SUV	3	26	27	27	6	No	333
6	ACURA MDX	3.5	6	SemiAuto-9	2WD	Gasoline	CA	L3ULEV125	California LEV-III ULEV125	LHNXV03.5PBM	small SUV	3	20	27	23	5	No	387
7	ACURA MDX	3.5	6	SemiAuto-9	2WD	Gasoline	FA	T3B125	Federal Tier 3 Bin 125	LHNXV03.5PBM	small SUV	3	20	27	23	5	No	387
8	ACURA MDX	3.5	6	SemiAuto-9	4WD	Gasoline	CA	L3ULEV125	California LEV-III ULEV125	LHNXV03.5PBM	small SUV	3	19	26	22	4	No	404
9	ACURA MDX	3.5	6	SemiAuto-9	4WD	Gasoline	FA	T3B125	Federal Tier 3 Bin 125	LHNXV03.5PBM	small SUV	3	19	26	22	4	No	404
10	ACURA MDX A-spec	3.5	6	SemiAuto-9	4WD	Gasoline	CA	L3ULEV125	California LEV-III ULEV125	LHNXV03.5PBM	small SUV	3	19	25	21	4	No	415

Methodology

For this project 2020 fuel economy data from ***the office of Energy efficiency & Renewable energy*** was used (https://fueleconomy.gov/feg/download.shtml).The original data had 18 variables with 2513 rows. Columns containing data irrelevant for the project (such as 'certification regions' and 'underhood ID') were omitted. Rows containing vehicles operating with electricity and hydrogen was also omitted. The entire rows of columns containing missing values were dropped. The final dataset had 18 variables and 2356 observations (rows). Scatter plots were used to explore relationship amongst variables and see the patterns associated with these relationships. Descriptive statistics was also used to summarize the data as a whole. **Describe(include= 'all') method was used to explore the data.**

Data exploration was done, and the mean, Std Dev, minimum and maximum values were documented for each variable using python. A multiple linear regression model was trained taking Greenhouse score and Pollution score as dependent variables against the rest of the variables using **Scikit-learn**. Matplotlib is used as a tool for visualization.

Coefficient and Intercept , are the parameters used for the fit line. Given that it is a multiple linear regression, with 4 parameters, and knowing that the parameters are the intercept and coefficients of hyperplane, sklearn were used to estimate variables from our data. Ordinary Least Squares (OLS) method was used to solve this problem.

Results

A negative linear relationship was observed between Hwy MPG, Combined MPG and CO2 combustion with that of pollution score. The same negative linear relationship was also observed between greenhouse score and CO2 combustion. A positive linear relationship was observed between greenhouse score and the independent variables.

GMC Sierra is the most commonly encountered vehicle and SemiAuto-8 is the most common transmission type. Gasoline is the most commonly used fuel and 2WD is the most common drive type. The average CO2 combustion in all vehicle is 403. The City, Hwy and Combined MPG are 20.787351, 27.489813 and 23.293294 respectively.

Table 2: Coefficients

Independent variables	***Pollution score coefficients***	***Greenhouse Score Coefficients***
City MPG	0.07835427	0.07835427
Hwy MPG	-0.08694681	-0.08694681
Combined MPG	-0.02363349	-0.02363349
CO2 combustion	-0.01155652	-0.01155652

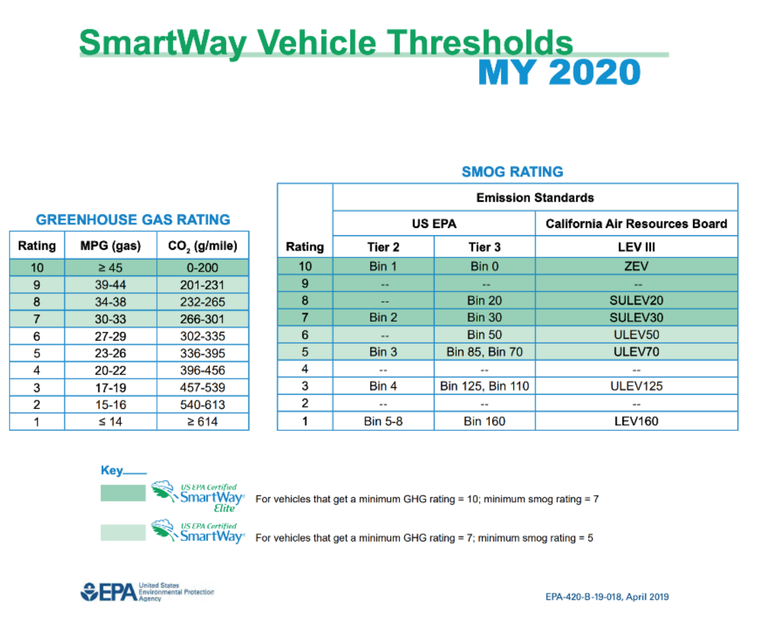
Discussion

City MPG, Hwy MPG, Cmb MPG and Comb CO2 is better in predicting Greenhouse score with a variance score of 0.71 compared to Pollution Gas score with a variance score of 0.23.

Given the City MPG, Hwy MPG, Combined MPG and CO2 combustion of vehicle, we can calculate the pollution and Greenhouse score of vehicles. The vehicles pollution score (also known as 'SMOG' score) and Greenhouse score will then be compared to the smartWay vehicle chart (figure 2).

Given the variance of the two models, the precision of prediction is better for Greenhouse Gas score than pollution score.

Figure 1: SmartWay Vehicle Threshold for 2020



SmartWay Vehicle Thresholds
MY 2020

GREENHOUSE GAS RATING

Rating	MPG (gas)	CO_2 (g/mile)
10	≥ 45	0-200
9	39-44	201-231
8	34-38	232-265
7	30-33	266-301
6	27-29	302-335
5	23-26	336-395
4	20-22	396-456
3	17-19	457-539
2	15-16	540-613
1	≤ 14	≥ 614

SMOG RATING

	Emission Standards		
	US EPA		California Air Resources Board
Rating	Tier 2	Tier 3	LEV III
10	Bin 1	Bin 0	ZEV
9	--	--	--
8	--	Bin 20	SULEV20
7	Bin 2	Bin 30	SULEV30
6	--	Bin 50	ULEV50
5	Bin 3	Bin 85, Bin 70	ULEV70
4	--	--	--
3	Bin 4	Bin 125, Bin 110	ULEV125
2	--	--	--
1	Bin 5-8	Bin 160	LEV160

Key
US EPA Certified SmartWay Elite For vehicles that get a minimum GHG rating = 10; minimum smog rating = 7
US EPA Certified SmartWay For vehicles that get a minimum GHG rating = 7; minimum smog rating = 5

EPA United States Environmental Protection Agency
EPA-420-B-19-018, April 2019

Conclusion

The number of vehicles is increasing worldwide, and it is vital for lawmakers to enforce the use of more and more SmartWay vehicles.

Vehicles fuel economy (measure by city MPG, Hwy MPG and combined city-Hwy MPG), amount of CO2 exhaust, the number of cylinders and the size of the vehicle have an impact to determine the Greenhouse and Pollution score of the vehicles. Elite vehicles tend to be smaller in size and are fuel efficient.

Coursera Capstone Project

Supervised machine learning using Scikit-learn were able to predict the Greenhouse Gas Score of vehicles after training with fuel economy data.

References:

1. Fuel Economy Guide for model yea 2020 (https://fueleconomy.gov/feg/pdfs/guides/FEG2020.pdf)

2. Sperling, D., and D. Gordon. Two Billion Cars: Driving Toward Sustainability. Oxford University Press, 2009.

3. Davis, S., and S. Diegel. Transportation Energy Data Book: Edition 26. ORNL-6978, Oak Ridge National Laboratory, 2007.

4. Intergovernmental Panel on Climate Change. Climate Change 2007: Mitigation of Climate Change. In IPCC Fourth Assessment Report, Working Group III Summary for Policymakers, April 12, 2007.