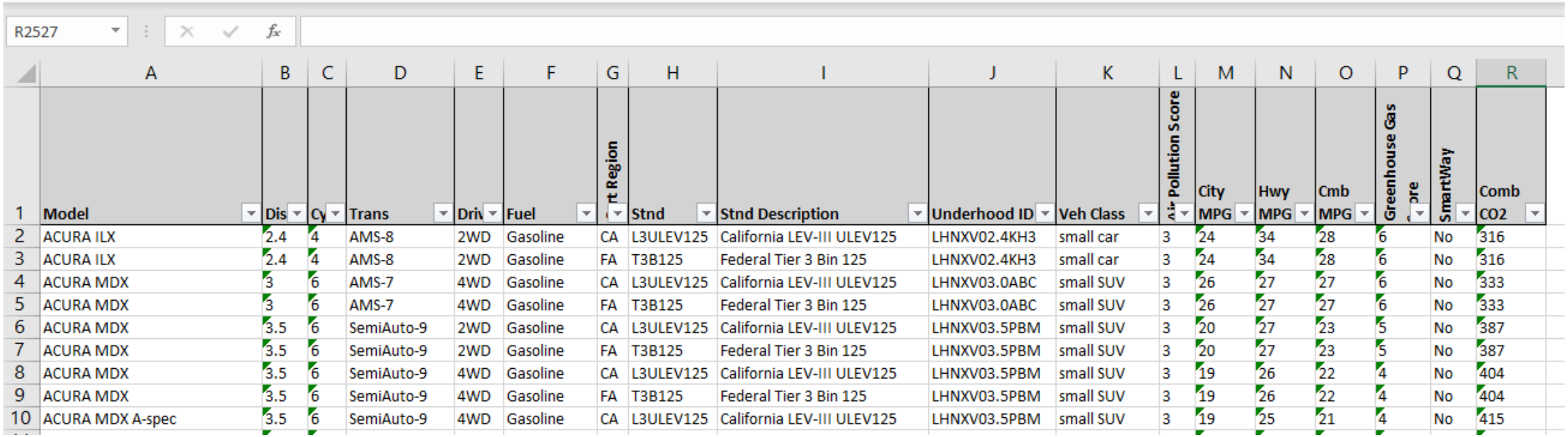
Coursera Capstone Project

- Data Science
- Is your vehicle eco-friendly??

Introduction

- Global warming and climate changes are existential threats to human race.
- CO2 emission is the major culprit in the process
- Motor vehicles are the main source of CO2 emission
- The number of Motor vehicle is expected to be higher than 2 billion by 2020.
- Policy makers should devise a way to regulate the amount of CO2 emission
- More “Green vehicles” on the road means less CO2 emission.



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 core	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

Data

- Fuel economy data from department of renewable energy is used.
- File can be found at (https://fueleconomy.gov/feg/download.shtml)
- Original data has 18 variables and 2513 rows

Methodology

- Data cleaning
- Descriptive statistics
- Scikit-Learn to train/test the model
- Matplotlib used for visualization
- Ordinary Least Squares (OLS) method was used to solve this problem.

Result

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

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/Conclusion

1

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.

2

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

3

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

Thank you!