Helping your car selection help our planet

A linear regression story by Celina Plaza

AGENDA

01 OVERVIEW & OBJECTIVE

02 METHODOLOGY

03 THE MODEL

04 KEY TAKEAWAYS

05 FUTURE OPPORTUNITIES

Transportation is the largest contributor of the United States' greenhouse gas emissions.

- United States Environmental Protection Agency

KNOW YOUR CAR'S IMPACT

Through linear regression modeling, identify **features of non-electric cars** that are **correlated** with the car's **greenhouse gas emissions (gge).**

METHODOLOGY

1: COLLECT DATA

All data was web-scraped from FuelEconomy.gov = **~5,200 samples**

2: EXPLORATORY DATA ANALYSIS

Examined features relationship to greenhouse gas emissions

3: BUILD MODEL

Tested various feature shifts to get strongest results from a model

4: VALIDATION

Conducted validation & cross-validation tests on model's performance

5: PUT MODEL TO THE TEST

Use model to create predictions and compare to real value

METHODOLOGY

1: COLLECT DATA

All data was web-scraped from FuelEconomy.gov = ~5,200 samples

2: EXPLORATORY DATA ANALYSIS

Examined features relationship to greenhouse gas emissions

3: BUILD MODEL

Tested various feature shifts to get strongest results from a model

4: VALIDATION

Conducted validation & cross-validation tests on model's performance

5: PUT MODEL TO THE TEST

Use model to create predictions and compare to real value

METHODOLOGY

1: COLLECT DATA

All data was web-scraped from FuelEconomy.gov = ~5,200 samples

2: EXPLORATORY DATA ANALYSIS

Examined features relationship to greenhouse gas emissions

3: BUILD MODEL

Tested various feature shifts to get strongest results from a model

4: VALIDATION

Conducted validation & cross-validation tests on model's performance

5: PUT MODEL TO THE TEST

Use model to create predictions and compare to real value

METHODOLOGY

1: COLLECT DATA

All data was web-scraped from FuelEconomy.gov = ~5,200 samples

2: EXPLORATORY DATA ANALYSIS

Examined features relationship to greenhouse gas emissions

3: BUILD MODEL

Tested various feature shifts to get strongest results from a model

4: VALIDATION

Conducted validation & cross-validation tests on model's performance

5: PUT MODEL TO THE TEST

Use model to create predictions and compare to real value

METHODOLOGY

1: COLLECT DATA

All data was web-scraped from FuelEconomy.gov = ~5,200 samples

2: EXPLORATORY DATA ANALYSIS

Examined features relationship to greenhouse gas emissions

3: BUILD MODEL

Tested various feature shifts to get strongest results from a model

4: VALIDATION

Conducted validation & cross-validation tests on model's performance

5: PUT THE MODEL TO THE TEST

Use model to create predictions and compare to real value

THE DATA

MAKE
+
MODEL
+
YEAR

ENGINE CAPACITY

CYLINDERS

TRANSMISSION TYPE

TRANSMISSION SPEED

FUEL TYPE

MILES-PER-GALLON

GREENHOUSE GAS EMISSIONS (from tailpipe)

THE FEATURES

Numerical

ENGINE CAPACITY

Categorical

YEARS

TRANSMISSION TYPE

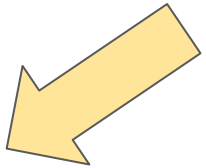
FUEL TYPE

GREENHOUSE GAS EMISSIONS (from tailpipe)

THE FEATURES

Numerical

ENGINE CAPACITY



Categorical

YEARS

TRANSMISSION TYPE

FUEL TYPE

GREENHOUSE GAS EMISSIONS (from tailpipe)

WHY YEARS MATTERS

The first U.S. greenhouse gas vehicle standards began with 2012 models.

‘Pre-Regulations’ Period

car model years

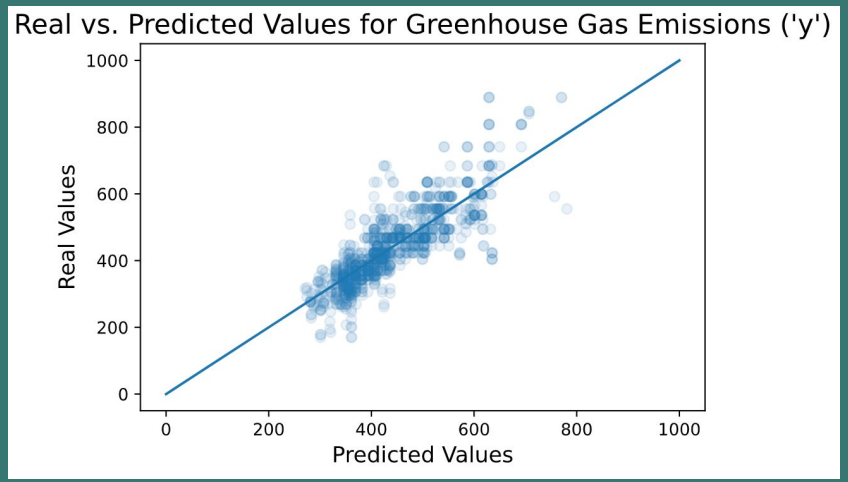
1984 - 2011

Regulations Period

car model years

2012 - 2021

THE MODEL's MEASUREMENTS



R^2 = **69%**

Mean Absolute Error (mae) = **47.2**

What the model tells us:
Signs that your car has lower greenhouse gas emissions:

TAKEAWAYS

ENGINE CAPACITY

Lower engine capacity. Data showed range from 1 to 8.4 liters, so look for lower end of that range

YEAR

Cars with model years on or after **2012 at least until 2020**. Time will tell of effects of loosened regulations starting 2021.

TRANSMISSION

Now's the time to learn how to stick shift! Opt for **manual transmission** over automatic.

FUEL TYPE

Regular is best. Then diesel. Then premium. Who wants to pay more anyways?

TAKEAWAYS

Signs that your car has lower greenhouse gas emissions:

ENGINE CAPACITY

Lower engine capacity. Data showed range from 1 to 8.4, so look for lower end of that range

YEAR

Cars made on or after **2012 at least until 2020**. Time will tell of effects of loosened regulations sta

TRANSMISSION

Now's the time to learn how to sti
manual transmission over auton

FUEL TYPE

Regular is best. Then diesel. Then pr
And who wants to pay more anyway?

Or better yet, go electric!

FUTURE OPPORTUNITIES

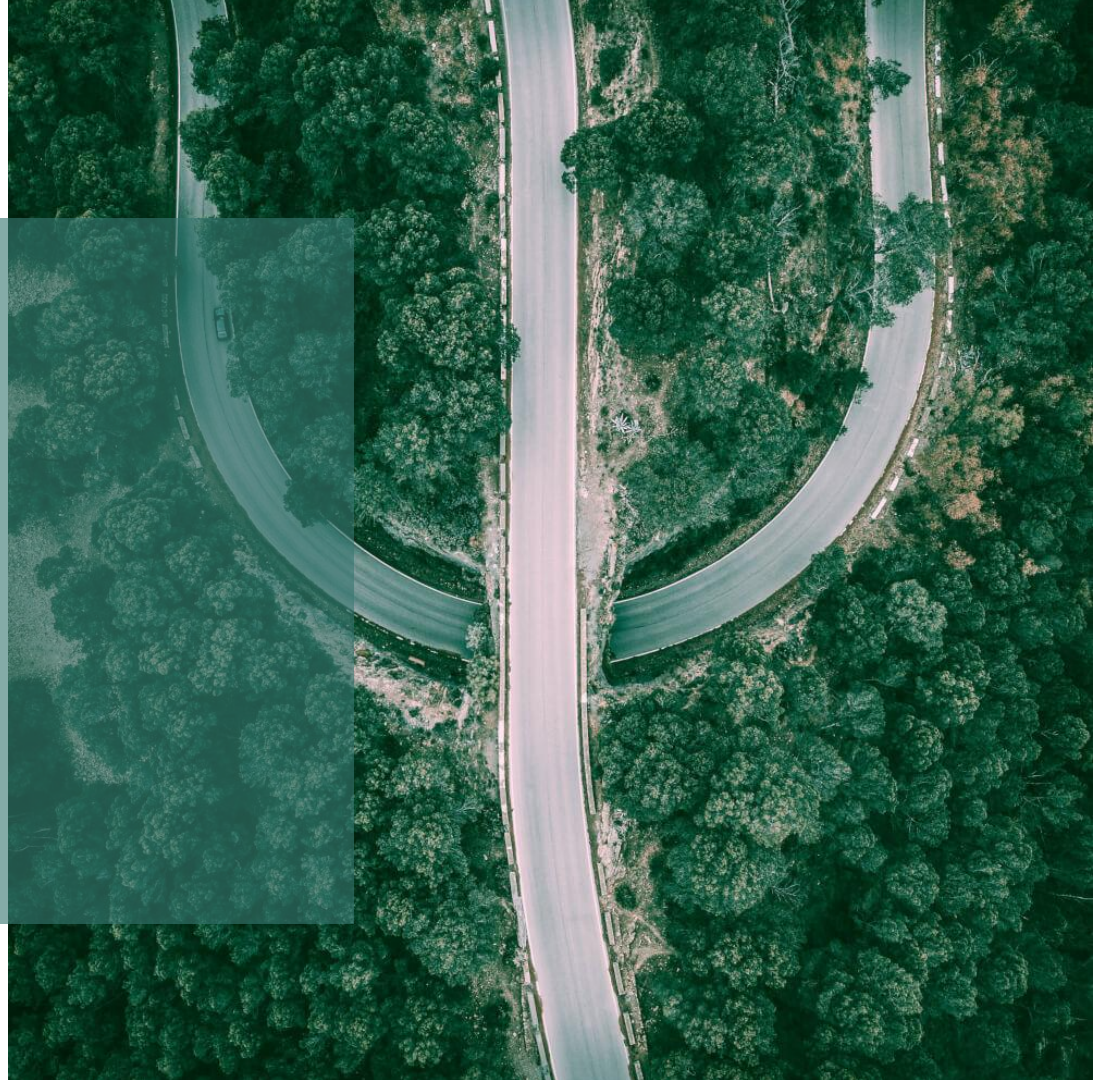
- Re-examine after 2021 and beyond to see if model changes for period when regulations were loosened

- Broaden scope of project to include ‘full-cycle’ of emissions related to car, not limited to but including:
  - manufacturing of car
  - delivery of cars to sales shop
  - if electric, power source’s fuel & related emissions

THANK YOU

Does anyone have any questions?

Celina Plaza



CREDITS

This is where you give credit to the ones who are part of this project.

- Presentation template by Slidesgo
- Icons by Flaticon
- Infographics by Freepik
- Images created by Freepik

APPENDIX

MODEL'S FULL FORMULA

$$Y_p = 227.88 + 60.2x_1 + 13.46x_2 + 47.5x_3 + 32.66x_4 - 63.29x_5$$

- x_1: **Engine Capacity** in **liters** (numeric)
- x_2: **Automatic Transmission** (dummy = 1)
- x_3: **Premium gas** (dummy = 1)
- x_4: **Regular gas** (dummy = 1)
- x_5: **Model year** between **2012 - 2021** (dummy = 1)

Real Example Tests:

1. **2020** Honda CR-V AWD has **engine capacity of 1.5 liters**, **automatic transmission**, and takes **regular gas.**.
   > Model's predict gge**: 300.93**, Real gge: **305**

2. **2006** Chevrolet SSR Pickup 2WD has **engine capacity of 6 liters**, **manual transmission**, and takes **premium gas**.
   > Model's predict gge: **636.58**, Real gge: **635**

GREENHOUSE GAS EMISSIONS DATA BASICS

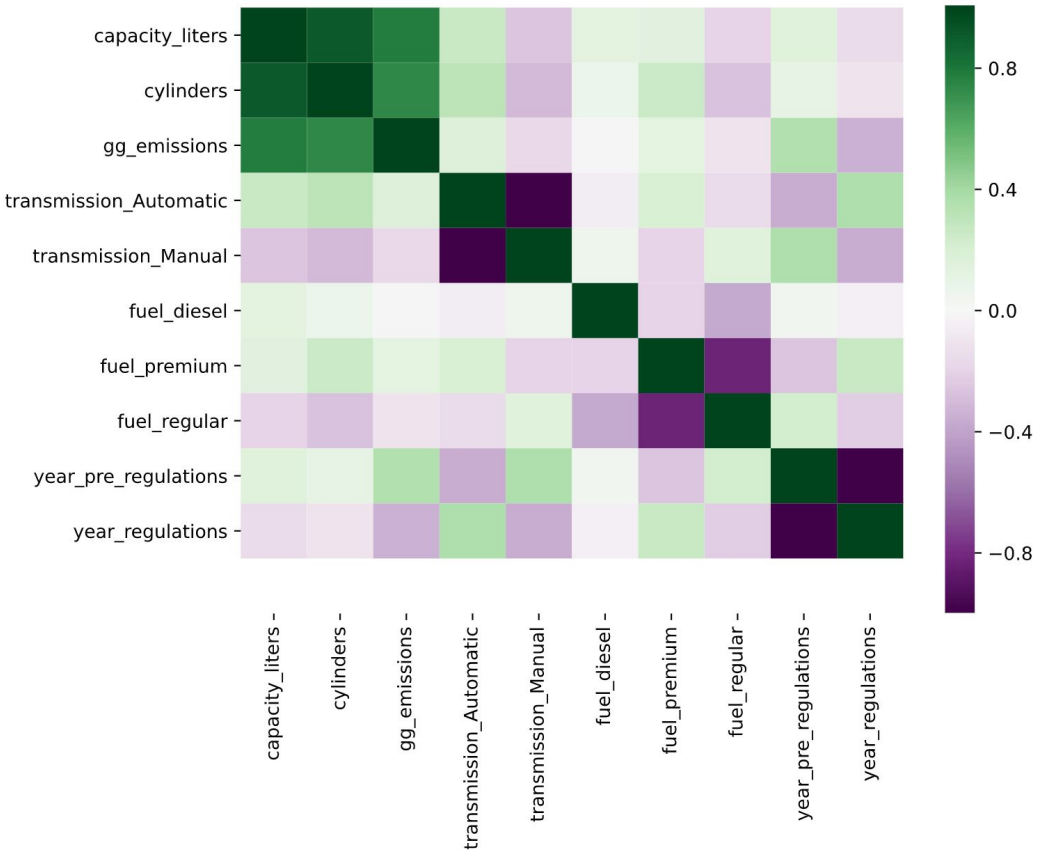
```
There are 335 values of actual greenhouse gas emissions -- the "y" values.
The values for the greenhouse gas emissions range from 168.0 to  889.0 grams/mile
```

CORRELATION LIST TO GREENHOUSE GAS EMISSIONS (y)

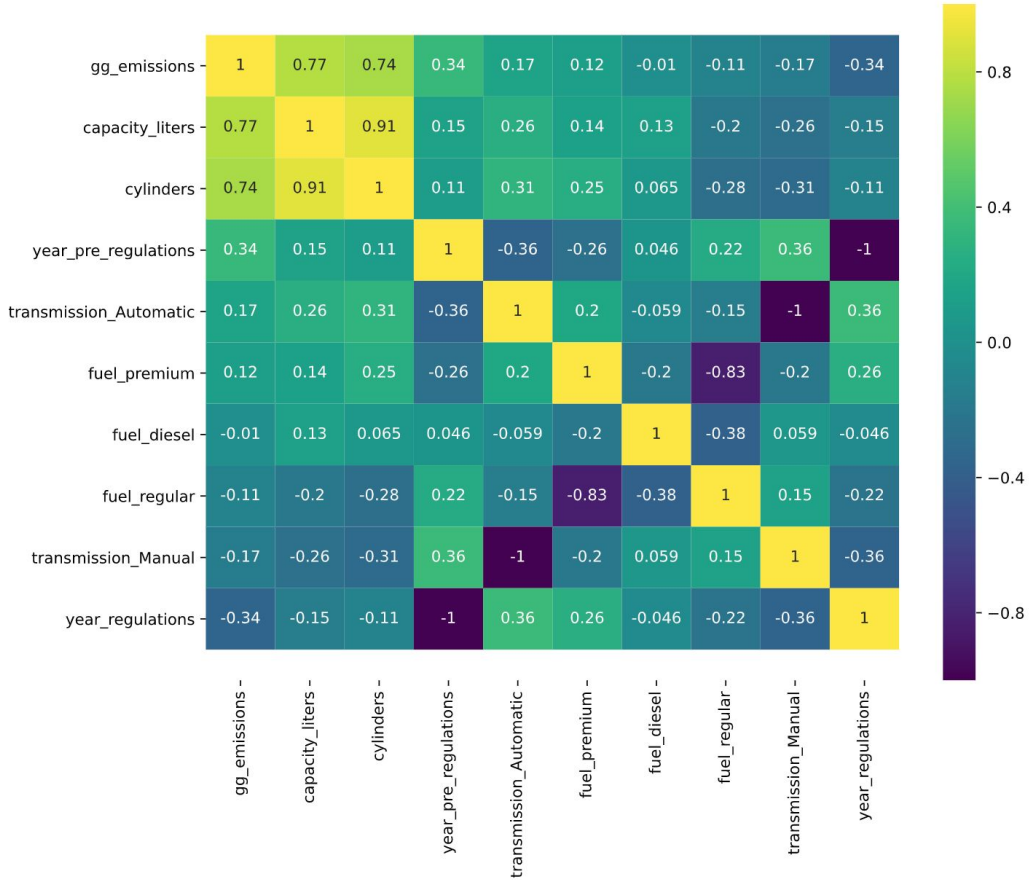
```
capacity_liters              0.773934
cylinders                    0.740521
gg_emissions                 1.000000
transmission_Automatic       0.169507
transmission_Manual         -0.169507
fuel_diesel                 -0.010432
fuel_premium                 0.117899
fuel_regular                -0.105576
year_pre_regulations         0.343847
year_regulations            -0.343847
Name: gg_emissions, dtype: float64

-
```

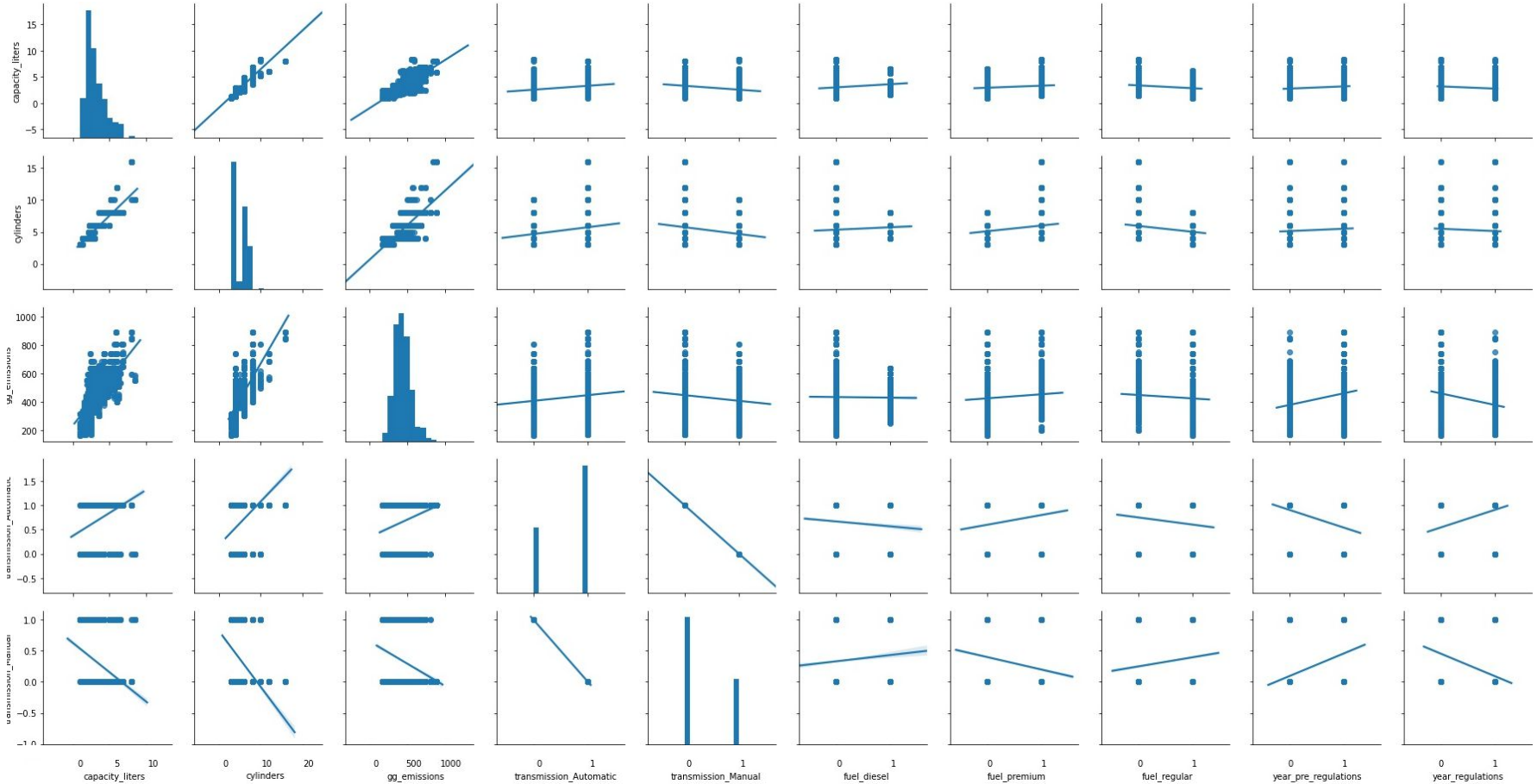
CORRELATION HEATMAP



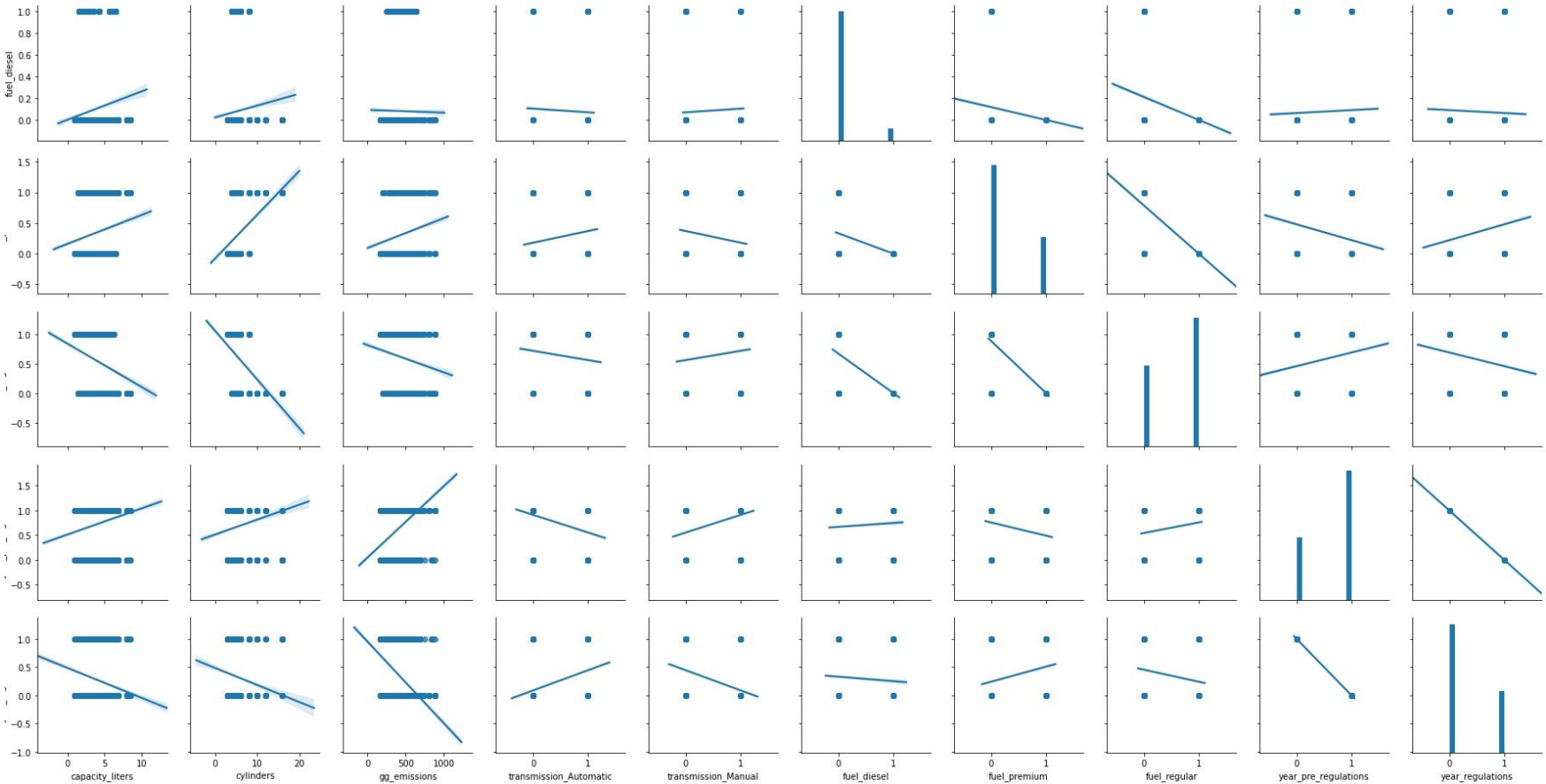
GGE CORRELATION MATRIX



PAIR PLOTS (1 of 2)

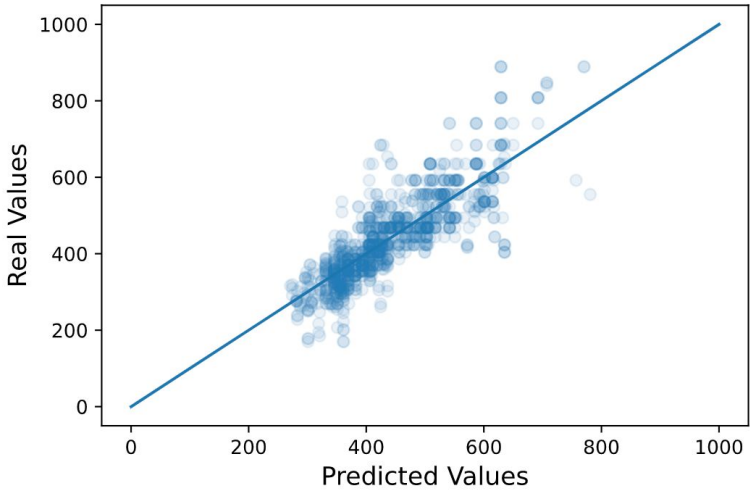


PAIR PLOTS (2 of 2)



RIDGE MODEL REGRESSION RESULTS - FINAL MODEL

Real vs. Predicted Values for Greenhouse Gas Emissions ('y')

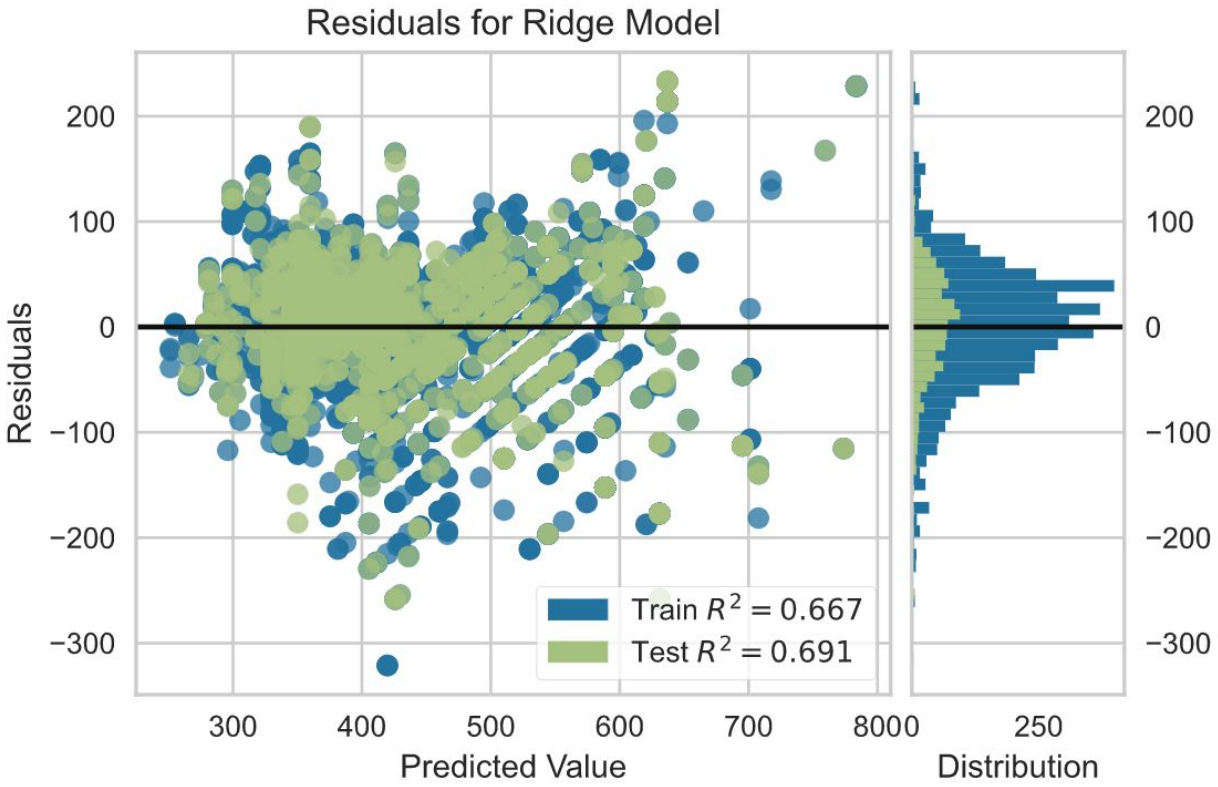



```
COEFFICIENTS for RIDGE Model:
[('capacity_liters', 60.19222009830541),
 ('transmission_Automatic', 13.464090054103353)
 ('fuel_premium', 47.49635509672385),
 ('fuel_regular', 32.65820777307923),
 ('year_regulations', -63.29001086355502)]
```

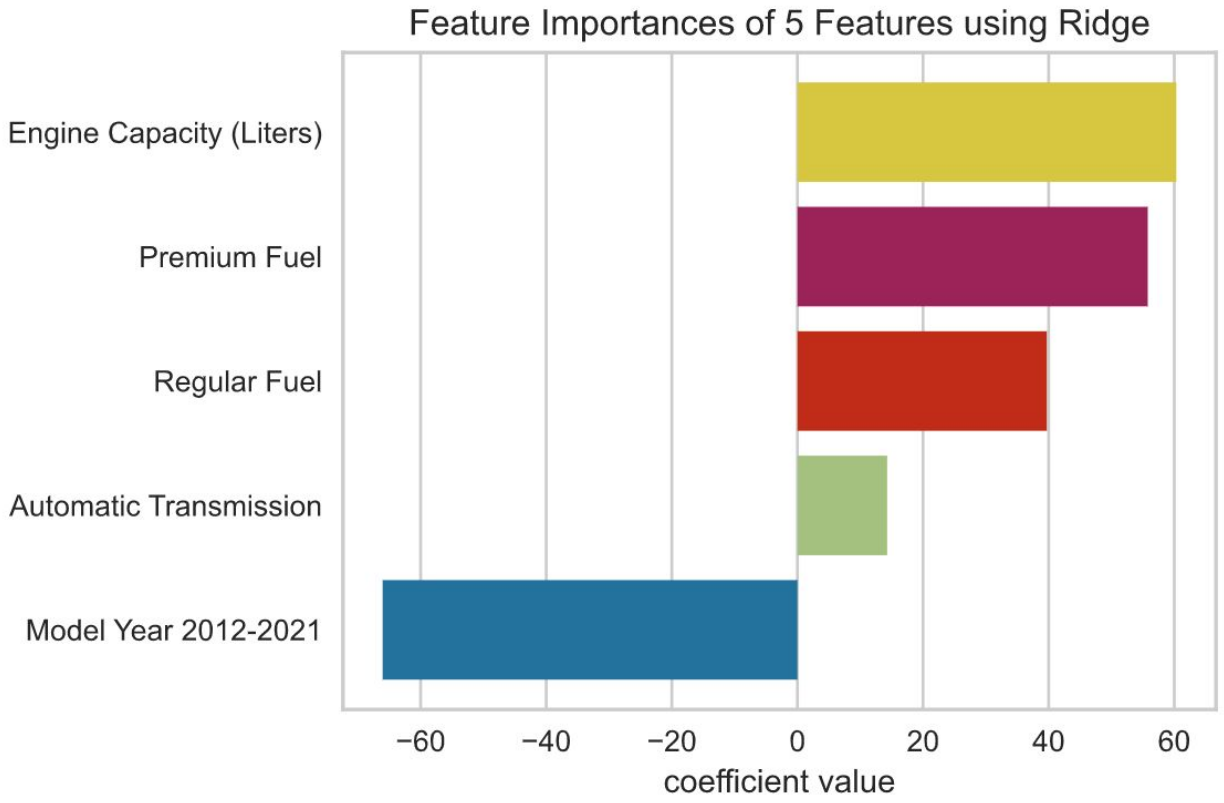
```
INERCEPT for RIDGE Model:
227.88099122750089
```

```
...
RIDGE MODEL:
R^2: 0.6903722423162104
Mean Absolute Error: 47.19790875474612
```

RESIDUALS OF MODEL



PLOTTING COEFFICIENTS FOR RIDGE MODEL



OLS MODEL RESULTS FOR SELECTED MODEL

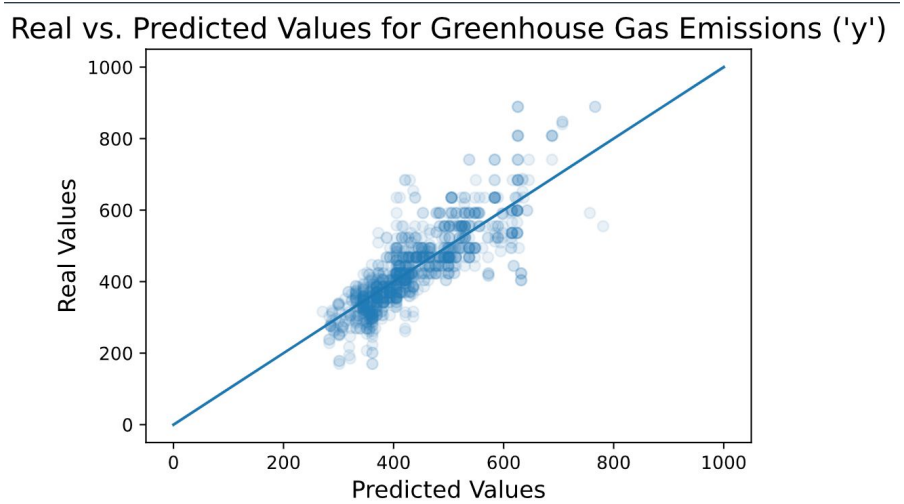
OLS Regression Results

Dep. Variable:	gg_emissions	**R-squared:**	0.667
Model:	OLS	**Adj. R-squared:**	0.666
Method:	Least Squares	**F-statistic:**	1680.
Date:	Thu, 16 Jul 2020	**Prob (F-statistic):**	0.00
Time:	13:14:33	**Log-Likelihood:**	-23460.
No. Observations:	4208	**AIC:**	4.693e+04
Df Residuals:	4202	**BIC:**	4.697e+04
Df Model:	5		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	220.6943	4.515	48.879	0.000	211.842	229.546
capacity_liters	60.3358	0.830	72.683	0.000	58.708	61.963
transmission_Automatic	14.3127	2.384	6.003	0.000	9.638	18.988
fuel_premium	56.1192	3.924	14.302	0.000	48.427	63.812
fuel_regular	40.0015	3.705	10.797	0.000	32.738	47.265
year_regulations	-66.1480	2.422	-27.310	0.000	-70.897	-61.399

Omnibus:	337.500	**Durbin-Watson:**	1.997
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	857.741
Skew:	0.467	**Prob(JB):**	5.54e-187
Kurtosis:	5.005	**Cond. No.**	23.3

LASSO MODEL REGRESSION RESULTS - NOT SELECTED

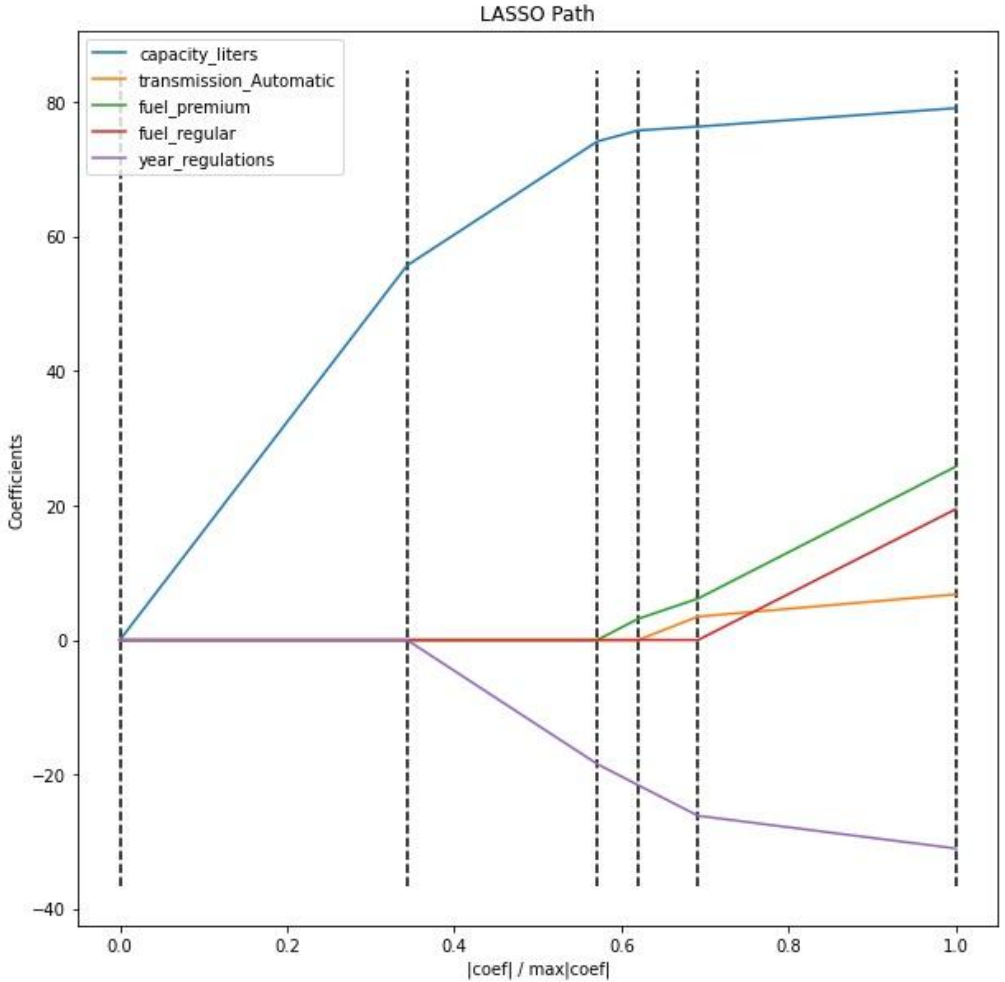



```
COEFFICIENTS for LASSO Model:
{'capacity_liters': 60.23716584200582,
 'transmission_Automatic': 9.552830288197704,
 'fuel_premium': 33.16088112384506,
 'fuel_regular': 19.203145188683628,
 'year_regulations': -59.56451385881141}
```

```
INERCEPT for LASSO Model:
241.71179090637253
```

```
LASSO MODEL:
R^2: 0.6875868818662549
Mean Absolute Error: 47.43280666154661
```

LASSO PATH


LASSO Path
capacity_liters
transmission_Automatic
fuel_premium
fuel_regular
year_regulations
Coefficients
80
60
40
20
0
−20
−40
0.0
0.2
0.4
0.6
0.8
1.0
|coef| / max|coef|