# PREDICTING CO2 EMISSIONS BY CAR FEATURES

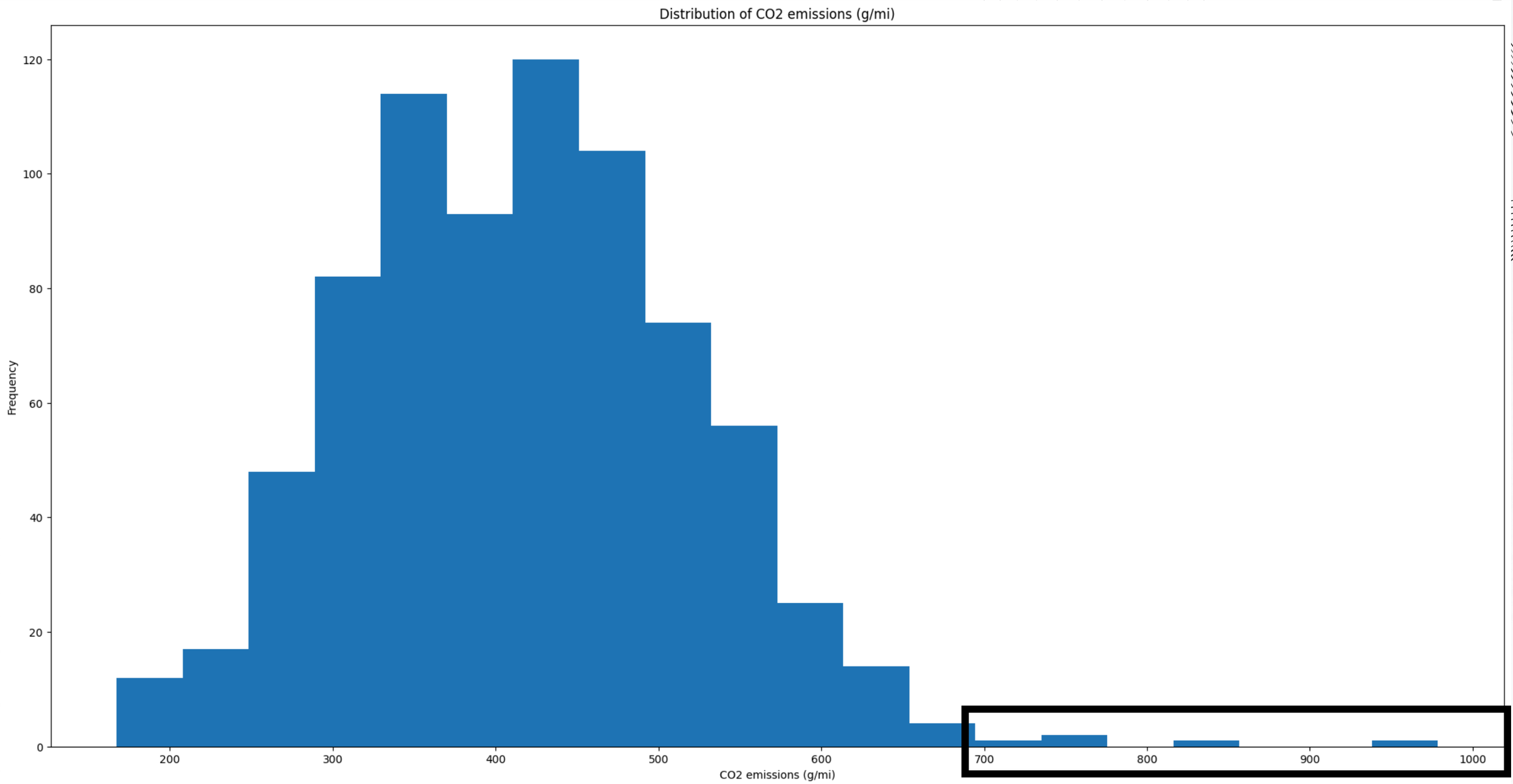
# INTRODUCTION

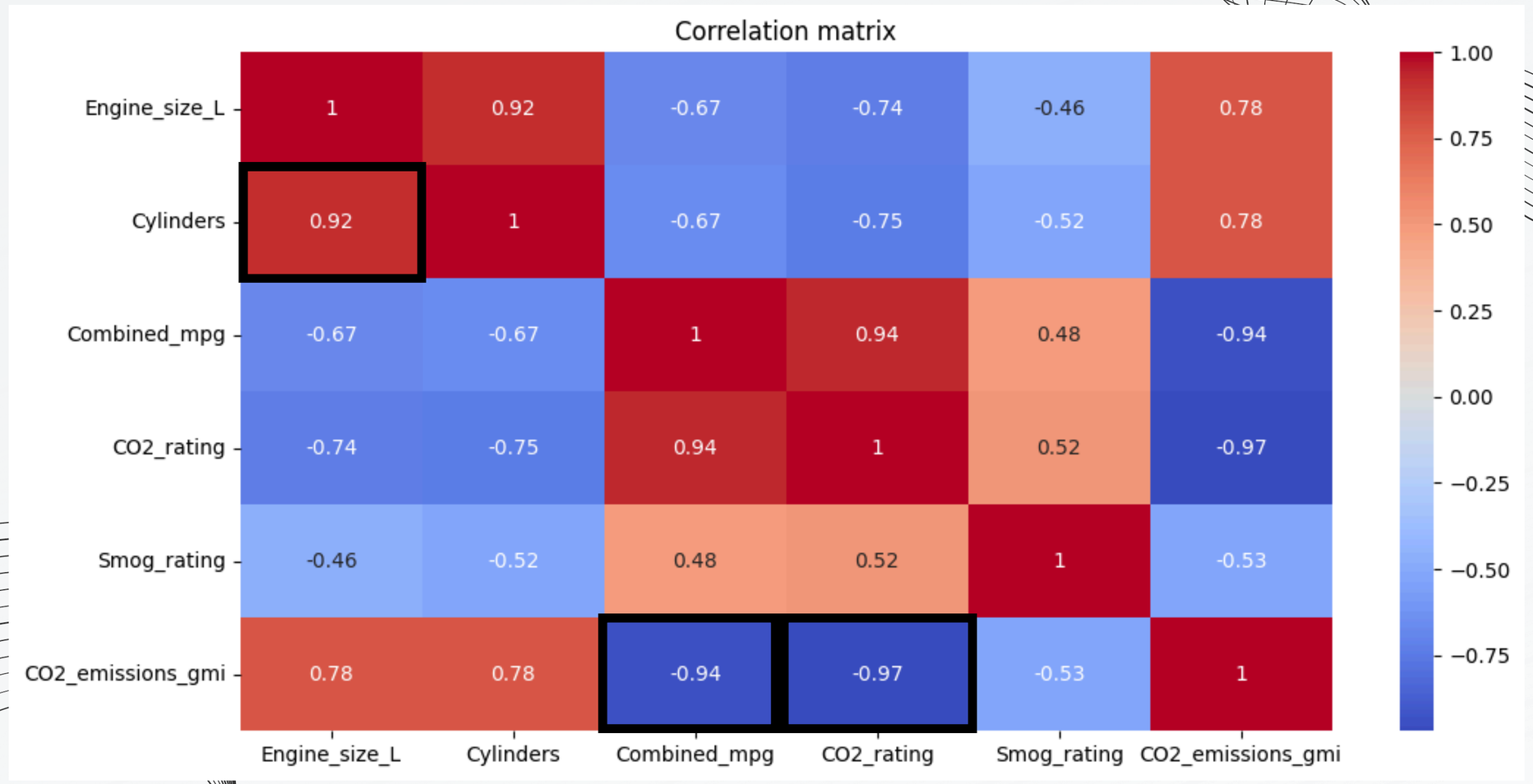
- **Problem:** Cars produce almost 16% of total carbon dioxide ($CO_2$) emissions in the United States

- **Project goal:** Determine which features of a car predict its $CO_2$ emissions the best

# DATA CLEANING/EDA

- **Dataset:** Canada govt. database of 2024 vehicles
- **Key variables:** Gas mileage, transmission, fuel type
- **Data cleaning:** drop outliers and irrelevant columns, rename columns, convert $CO_2$ emissions to g/mi


Distribution of CO2 emissions (g/mi)
Frequency
0
20
40
60
80
100
120
200
300
400
500
600
700
800
900
1000
CO2 emissions (g/mi)

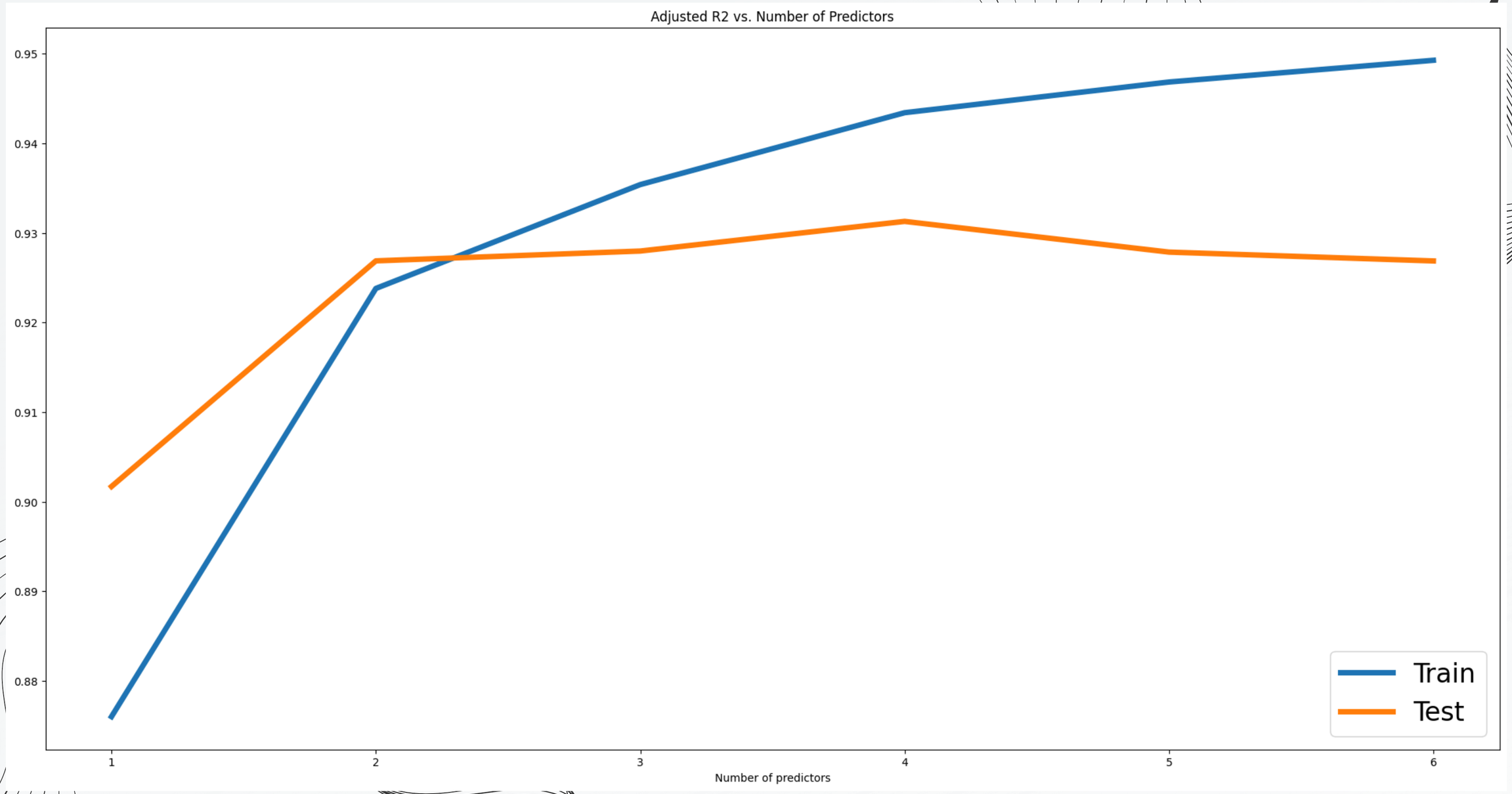

Correlation matrix
Engine_size_L
Cylinders
Combined_mpg
CO2_rating
Smog_rating
CO2_emissions_gmi
1
0.92
-0.67
-0.74
-0.46
0.78
0.92
1
-0.67
-0.75
-0.52
0.78
-0.67
-0.67
1
0.94
0.48
-0.94
-0.74
-0.75
0.94
1
0.52
-0.97
-0.46
-0.52
0.48
0.52
1
-0.53
0.78
0.78
-0.94
-0.97
-0.53
1
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75

# MODEL BUILDING

- **ML approach:** Multilinear regression
- Used forward selection to choose best variable to add before comparing using adjusted R2 value
- Checked for collinearity using variance inflation factors and adding interaction terms on final model

# FORWARD SELECTION

| Number of Predictors | New Predictor | Adjusted R2 |
|---|---|---|
| 1 | Combined Gas Mileage (MPG) | 0.876 |
| 2 | Number of Cylinders | 0.924 |
| 3 | Transmission | 0.935 |
| 4 | Fuel Type | 0.943 |
| 5 | Vehicle Class | 0.947 |
| 6 | Engine Size (L) | 0.949 |


Adjusted R2 vs. Number of Predictors
0.95
0.94
0.93
0.92
0.91
0.90
0.89
0.88
1
2
3
4
5
6
Number of predictors
Train
Test

# CONCLUSION

- **Final model with two predictors:**

  Combined gas mileage (mpg) = -9.65

  Number of cylinders = 15.84

  Adjusted R2 = 0.923

- This model highlights the car features that have the largest impact on CO2 emissions