Car Fuel Economy

Jeffrey He, Northwestern University
Pranav Swaminathan, Vernon Hills High School

Summer 2021 Mechanistic Data Science Course – Northwestern University

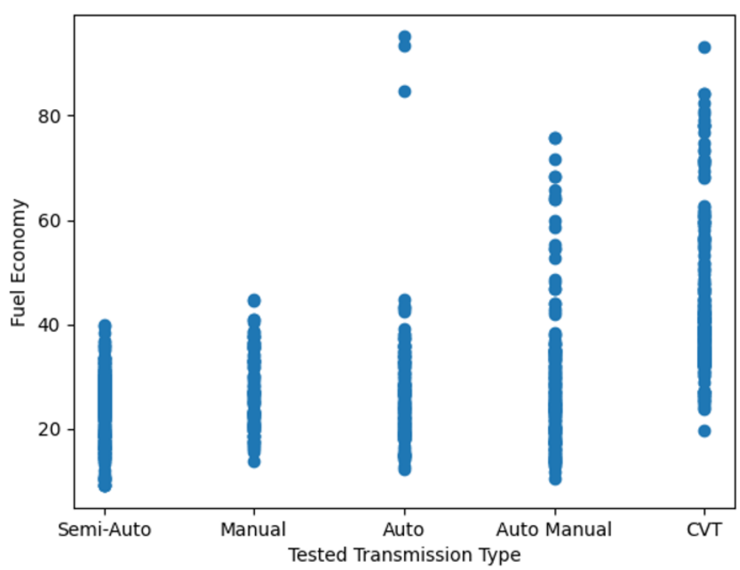
July 10, 2021

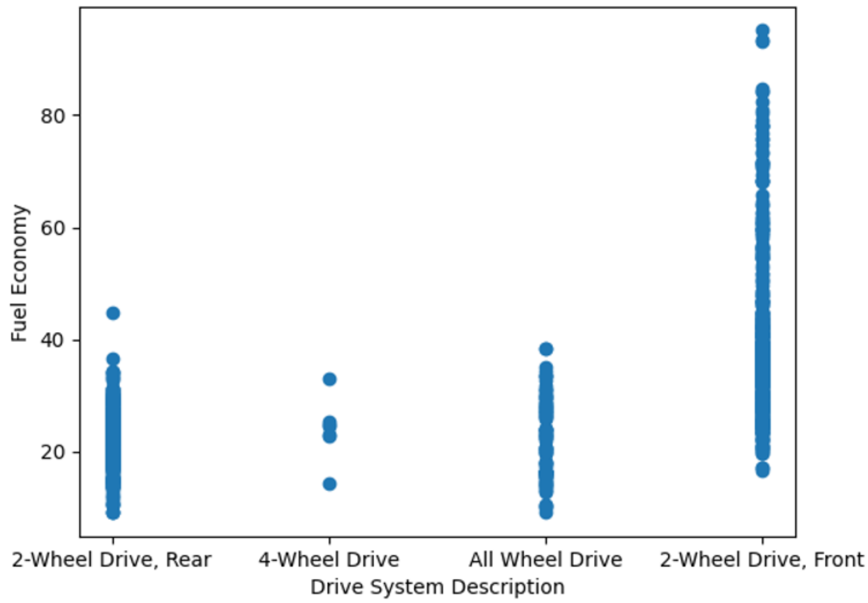
Objective and Goal

Car fuel economy is a measure of how far a car can travel per unit of fuel. In most countries, fuel economy is represented in miles per gallon or mpg. The average US citizen spends around $3000 per year on gas. This cost is an enormous amount which can be reduced with better fuel economy. Not only cost, but by decreasing gasoline usage which also reduces carbon dioxide emissions as well. Due to this, we can create a model that can predict fuel economy for a vehicle based off of several of its attributes through regression analysis and neural networks. The model will be able to help determine if a newer car model is more or less fuel efficient compared to other cars at a similar price.

From the dataset provided by the EPA, there is an excel sheet or csv file from 1984 to 2021. This data is obtainable from: https://www.epa.gov/compliance-and-fuel-economy-data/data-cars-used-testing-fuel-economy.
Inside the excel sheet are numerous features. Some of the more important ones are vehicle model, rated horsepower, transmission type, test weight, fuel economy, and gas emissions.

One piece of data that is not included which may be important is the drag force. From other research, it is known that the drag force will affect fuel economy by 50% if it is on the highway and around 20% if it is doing city driving. Although this information is not included in the dataset itself, additional engineering principles can be used in conjunction with the raw data for mechanistic data science. Another potential aspect that should be considered is price. We are currently attempting to obtain a source for car prices to add to our list of important features since it can help provide an aspect of which cars have better fuel economy for its value. Also if we could identify how much fuel efficiency decreases per year that would also be a useful feature. In terms of current progress, we have been working on extracting certain features which most consumers care about from the datasheet, and then identifying common trends compared to the fuel economy. Some examples of this are below.





As shown above, auto-manual and continuously variable transmission are better than other transmission types for fuel efficiency. Also the 2-wheel front drive is clearly more fuel efficient compared to the other driving systems. This is due to 4-wheel and all wheel drive requiring many more parts to drive. The 2-wheel rear drive does have advantages over the 2-wheel front drive such as handling and maintenance, however, does decrease in fuel economy due to distance from the engine.

We are currently working with highway fuel economy tests and cars only, however, we will consider both highway and city driving as well as other vehicles such as trucks. The data above is for only cars on highways. As the database used does not contain prices, we are currently trying to obtain prices for these different car makes and models from other sources to compare fuel efficiency across multiple prices.

After particular features are extracted with regards to consumer importance, regression analysis and neural networks can be used to create a model that inputs the car price and important features, and then outputs an expected fuel economy in miles per gallon. A user can then input the features of an upcoming car, to see if the fuel economy listed is better or worse than the average car in the past 10-20 years for a given price point.

6 step process

Multimodal data generation and collection: Utilize the EPA dataset to obtain car features for different cars throughout the past 35 years.

Feature engineering: Narrow the large EPA dataset into only considering useful features that impact fuel economy through sensitivity analysis of different features compared to fuel economy. Use python to extract the useful features and compile them alone. Add price from an outside source as a feature.

Knowledge-Driven dimensional reduction: Utilize known properties about fuel economy such as how the aerodynamic drag force impacts highway and city driving differently. Potentially identify reasons for certain features having behavior.

Surrogate models: Split up the data into highway and city driving fuel economy ratings. Model the effects of different features on fuel economy.

Neural networks & regression: Model the relationships between chosen impactful features towards fuel economy using regression analysis and neural networks.

System Design: Obtain a predictive model that can help predict fuel economy based on certain features it has, and potentially see if it has above or below average fuel economy compared to other cars within its price range or throughout time.