International Journal of Science and Research (IJSR)
ISSN: 2319-7064
Impact Factor 2024: 7.101

What Affects Fuel Efficiency? A Data-Driven Study of Vehicle Data using ML Algorithms

Kiros Kamaal

[1]Singapore International School, Mumbai
Email: *kiros.kamaal1[at]gmail.com*

Abstract: ***This manuscript takes into account the different factors affecting fuel economy and describes the individual effect of each factor mentioned in the paper. The factors are described and we investigated which factors have the most impactful effect on a vehicle's fuel economy. In contrast to other research papers focusing on the fuel economy of vehicles, this manuscript takes into account fuel economy as a whole. In the paper we used two different ML algorithms to predict the real world effect of the different factors and identify the most effective variable in accurately predicting fuel economy.***

Keywords: Fuel Economy, Car Body Type, Car Weight, Linear Regression, Engine Displacement

1. Introduction

Fuel economy, which refers to the distance a car can cover with a specified amount of fuel (United States Environmental Protection Agency, 1976). Due to growing environmental issues and rising fuel prices, manufacturers, governments, and vehicle consumers need to comprehend the multiple determinants that affect fuel economy. In an internal combustion engine, the burning of fuel occurs within the internals of the engine; the energy released from here is required to overcome the weight of the car and the rolling and static friction. This power demand directly impacts the amount of fuel used, making the design and structure of vehicles one of the primary ways manufacturers can boost efficiency (Vehicle Technologies Office, 2013).

Research highlights vehicle weight and engine displacement (volume) as the two most significant factors of fuel economy (United States Environmental Protection Agency, 1976). 10% reduction in weight leads to 3%-6% increase in a vehicle's fuel economy. These factors cause other design elements, such as body shape, vehicle type and engine cylinder count, to cause further drastic changes in fuel economy when one of them is altered.

The main factors discussed in this manuscript are as follows: Weight, Aerodynamic Drag, Vehicle Body Types, Engine Displacement and Cylinder Count (Morris, 2025); discussed in detail in Section 3. While these factors are acknowledged, their importance toward fuel economy and the combined effect of using these factors to improve fuel economy have not been delved into with much detail. Further, statistical modelling is not straightforward, since fuel economy depends on multiple variables that interact, which means this study requires multiple statistical applications of different types to collect data and reach a conclusion. This paper aims to investigate the real effect of these factors utilising vehicle data, with a focus on understanding how differences in the factors mentioned above translate into differences in fuel efficiency. This manuscript aims to provide a clearer understanding of what most strongly affects fuel economy in everyday vehicles.

Section 2 discusses the literature findings and how they connect to the research questions of this manuscript. While previous studies have studied different relationships by focusing on the basic vehicle characteristics, this paper provides a suitable statistical analysis for understanding real-world fuel efficiency trends using statistical ML algorithms. The remainder of the paper investigates the fuel economy/efficiency in relation to the most important factor affecting fuel economy. These include the vehicle's body type, its aerodynamic drag and engine size, all of which are explained in detail in Section 3 of the paper. Section 4 describes the dataset we will analyse to test the effect of the factors on fuel economy. The analysis will employ Machine Learning (ML) modelling methods. Section 5 contains the results of the analysis using a couple of methods. Finally, Section 6 is a conclusion to this manuscript.

Each vehicle characteristic has its own effect on a specific vehicle's overall fuel economy. The effect of these factors can then be modelled using ML algorithms to understand the contribution each factor has. This then allows vehicle manufacturers to help determine their design of a car, especially by closely measuring the drag coefficient of vehicles to make them more fuel efficient. Vehicle manufacturers can also then compare the trade offs of other things with fuel economy to make sure the car is upto a desired standard but is still efficient. Similarly, Governments all over the world can also use these ML algorithms to create emission policies, for example smaller cars such as hatchbacks have stricter emission laws due to their small weight and hence are more economical whereas pickup trucks may be given more lenient emission laws with a comparatively reduced focus on its fuel economy.

2. Literature Review

Similar to the factors discussed and considered in this manuscript, other research papers state similar observations based on their data and research regarding fuel efficiency.

Bento et al. (2017) talks about the effect of Corporate Average Fuel Economy/Efficiency (CAFE) regulations on weight distribution in cars and empirically analyses how this affects vehicle fuel economy. Bento et al. (2017) do this by using 17 million accident records by unconditional quantile regression states the differences in safety that have caused differences in vehicle fuel economy. According to the same reference as above, CAFE is associated with a lower mean weight and an increase in dispersion. He states that every decrease of 40-50 lbs in accordance with CAFE in a car causes a rise of 1 mile per gallon. Bento et al. (2017) also state that there are a lot of fuel cost savings due to CAFE, which saves $1300-$2000 over a vehicle's lifetime.

According to Atabani et al. (2011), air drag increases by the square of the speed increase. Which means it is a significant factor of drag, and it requires high amounts of energy to overcome i.e more fuel is used. Atabani et al. (2011) state that 10% reduction in the drag coefficient leads to 23.5% decrease in fuel consumption- aligning with the weightage of aerodynamics regarding vehicle fuel economy given in this paper. Moreover, Atabani et al. (2011) explain how rolling friction can determine the mileage of a vehicle; it is mentioned that this rolling friction is determined by tire pressure and what is identified as the most significant factor of fuel economy in this paper-weight. Atabani et al. (2011) also mention different engine technologies such as variable valve timing, that alter timing chains and Valve Timings depending on the style of driving. It helps save fuel by allowing less fuel into the engine when it is not needed and provides significantly better air flow through air intake and free-flow exhaust systems. Early closing of the intake valve reduces stress on engine components and restricts fuel flow, thereby making it more economical. First invented by Alfa Romeo in 1983 and majorly improved on to production by Honda with the VTEC branding in 1989. Findings state that there is a 2%-5% improvement in fuel economy with variable valve timings in production vehicles proving that engine technology is also a factor affecting vehicle fuel economy (Atabani et al., 2011). Additionally, according to Atabani et al. (2011) turbocharging and supercharging with smaller engine displacement and cylinder blocks can also help improve fuel efficiency. This makes another factor of fuel economy being engine displacement and cylinder count as mentioned in Sections 3.4 and 3.5 of this paper.

With regards to engine technology, Johnson & Joshi (2018) stated the implementation of a technology called Gasoline Direct Injection (GDI), downsized turbo GDI engines have scope of up to 10%-15% reduction in fuel consumption compared to traditional gasoline engines (Atabani et al., 2011). Lastly, Johnson & Joshi (2018) state that another factor affecting vehicle fuel economy are government policies and emission regulations. Their paper states how different emission laws in the USA and Europe cause engine sizes to differ and make engines more economical to align with the emission laws. This means engines are usually downsized and/or turbocharged to follow the emission laws. The fuel flow may also be restricted to reduce emissions to abide by regulations, in turn affecting the fuel economy too. Johnson & Joshi (2018) mention the USEPA standards to reduce greenhouse gases, which in turn means the US government intends to reduce the tailpipe emissions of cars on the road, which means they will be made to be more economical. This proves that government policies are an overlooked factor of fuel economy.

3. Factors Affecting Vehicle Fuel Economy

Previous research has identified several key vehicle characteristics that influence fuel economy through physical and mechanical mechanisms. These factors affect how much energy a vehicle requires to accelerate, maintain speed, and overcome different forms of resistance during operation. Although these characteristics often interact in real-world driving, examining them individually helps clarify their primary effects and makes comparisons more straightforward. In this section, the most commonly cited factors affecting fuel economy are discussed, along with the physical reasoning behind their influence, providing a foundation for later data analysis and modeling.

3.1 Vehicle Weight and Energy Demand

This means more mass requires more energy for a car to move. Reduced weight has been a proven technique to increase fuel economy over multiple decades. Cheah (2010) states every 10% reduction in weight can lead to a fall in fuel consumption by a roundabout of 7%. However vehicle fuel reduction is not as simple as thought of, this is because safety standards still have to be followed and with more demand for bigger cars this makes it even harder of a task to reduce fuel consumption (Cheah, 2010). Additionally, vehicle weight is an overlooked dynamic of fuel consumption as it is variable.Weight also causes tyre deformation leading to further more drag and being an even more significant factor of fuel economy as it indirectly affects drag too (Fan et al., 2024). The rate of fuel consumption also almost linearly increases with weight (Essenhigh et al., 1979).

3.2 Aerodynamic Drag and Speed-Dependent Efficiency

Aerodynamic drag is also a significant factor that affects fuel consumptions. This is because it acts as an opposing force to motion. It is also not proportional to speed as air resistance is seen to increase by the cube of the speed increase as found through the mathematical formula (Mirmahdi et al. , 2021). Aerodynamics has a major effect on trucks as they have a boxy shape, with multiple panel gaps causing further increases in drag as the air then becomes highly turbulent, significantly reducing the fuel economy as the air drag may multiply to act together. Aerodynamic drag accounts for about 33.333% of losses in fuel economy for diesel trucks (Cooper, 2003). Further research by Mirmahdi et al. (2021) states that 36% reduction in the aerodynamic drag coefficient can lead to fuel savings of 16% or more. At speeds of 80km/h i.e., the cruising speed for semi-trucks in europe air drag can affect fuel economy with a weightage of almost 50% (Mirmahdi et al., 2021).

3.3 Body Type Classification and Structural Characteristics

Vehicle body type significantly influences weight distribution, aerodynamic shape, and engine size requirements, all of which affect fuel consumption. Hatchbacks, the smallest category, usually achieve between 18 and 24 km/l, benefiting from their compact size, lower weight, and smaller engines. Sedans, the middle category, deliver between 14 and 20 km/l with balanced designs: their streamlined shapes provide better aerodynamics than SUVs, but their increased weight and power needs surpass those of hatchbacks. SUVs use noticeably more fuel, generally around 10 to 18 km/l, due to larger engines, heavier weight, and less efficient aerodynamic profiles that feature larger frontal areas and boxy shapes, which increase drag. Trucks face a similar efficiency challenge, as their designs prioritise cargo space over aerodynamics. The aerodynamic drawbacks of larger vehicles worsen their weight issue, leading to a greater efficiency penalty. When comparing similar platforms, differences in vehicle weight often outweigh aerodynamic benefits; however, a sedan's lighter weight does not completely balance out an SUV's greater cargo capacity, highlighting the engineering trade-offs between efficiency and utility (Ahmed & Stater, 2023).

3.4 Engine Displacement and Power Generation

Increasing fuel consumption is not directly linear to increasing engine displacement, as it differs by a specific amount when larger engines create more efficiency due to less strain on the engines. Fuel consumption is inversely proportional to engine displacement. Fuel consumption of smaller engines is better as the smaller cylinders minimise heat losses (Essenhigh et al., 1979). Smaller engines also pull in significantly lesser fuel due to lesser displacement making them even more fuel efficient. Larger engines have a higher face-to-volume ratio which increases fuel consumption but may reduce the fuel consumption at idle RPMs. This especially affects fuel consumption at lower speeds(Essenhigh et al., 1979). Not just this but engine displacement is closely related to vehicle weight which is a significant factor of fuel economy (Lam, 1985). This means fuel economy may be reduced just due to the added weight of a larger engine. Lam (1985) states that there is a linear relation between engine size and fuel consumption (as mentioned before). There is also a slight correlation between engine size and power generation, which implies that larger displacement means more fuel is used and hence fuel economy will reduce with increasing displacement (Lam, 1985). Essenhigh et al. (1979) state that a 10% increase in displacement leads to a 3%-6% change in fuel economy.

3.5 Cylinder Count and Friction Losses

The number of cylinders affects fuel consumption in several ways, beyond just displacement. With a fixed total displacement, more cylinders create smaller individual combustion events happening more frequently. This leads to smoother power delivery and allows the engine to run at lower throttle settings while cruising. However, having more cylinders also raises internal friction losses: additional pistons, connecting rods, complex valvetrain setups, and bearing friction all increase mechanical losses. This friction penalty can offset the efficiency gains from requiring less throttle. Thus, the relationship between cylinder count and efficiency is not straightforward: a turbocharged three-cylinder engine, a naturally aspirated four-cylinder engine, a six-cylinder engine, and an eight-cylinder engine with the same displacement would demonstrate different efficiency results based on the balance between reduced throttle losses and higher friction losses. Current trends favor fewer, smaller cylinders with forced induction, reflecting the understanding that friction losses often outweigh the benefits of smoother operation in designs focused on efficiency (Xu, 2025).

4. Data and Methods

Our data comes from a publicly available fuel economy dataset that is based on official vehicle testing data collected by the U.S. Environmental Protection Agency (EPA). The original dataset has over 81 variables explaining different things about the car. A limitation of the dataset is that it does not include the weight of the vehicles, which we have theoretically identified as the main contributor to affecting fuel economy. For this reason, the analysis will focus on other variables, excluding the weight effect.

The 81 variables have been filtered into simply 4 variables of Miles per Gallon (MPG), displacement, Number of Cylinders in the engine and the vehicle group, which was split into Compact Cars, Midsize/Large Sedans, SUVs, Pickups and Other special vehicle categories. The data was edited to ignore any blank spaces where data was unavailable. This changed the total data set from over 38000 observables to 37977 observables, the variables measured (as mentioned previously) also changed from 81 variables to just 4 variables.

Miles per Gallon (MPG) was taken as the variable that measures fuel efficiency since the dataset is based on vehicles from the USA, where the imperial system is used instead of the metric system. Noting that weight was not included in the data set, we can still make assumptions of its effect on fuel economy as weight is usually directly proportional the number of cylinders and engine displacement as well, furthermore the weight of the vehicle can be estimated by seeing the vehicle body type too; i.e., it is simple and straightforward to understand the effect fuel economy due to weight by inferring its relationships with the other variables being measured.

In the dataset, there are 11605 Compact Cars, 9290 Midsize/Large Cars, 5231 SUVs, 5706 Pickup Trucks, 4652 Vans, and 1493 vehicles of other types. The small number of vehicles of other types helps make the data readings more accurate, and these vehicle groups are neither too broad nor too narrow for this research paper, making it perfect for the data analysis of this manuscript. The MPG ranges between 7mg to all the way to 58mpg, the displacement (measured in Litres) ranges from as tiny as 0.6L to a monstrous 8.4L, the number of cylinders also vary vastly from just 2 cylinders (as little as an entry sport motorcycle) to 16 cylinders (found more commonly in marine or diesel locomotives) proving the dataset is diverse enough for the manuscript.

In addition to the descriptive statistics above, we focused on visually exploring MPG against the vehicle body types. MPG seems to be larger on average for the Other Vehicle category, then for Compact,then for Midsize/Large, and then SUVs, Trucks, and Vans. Furthermore, the Midsize/Large vehicle body type has the most outliers. These can be observed from the boxplot in Figure 1. The boxes in Figure 1 represent the Interquartile Range (IQR) which is the middle 50% of observations, the line inside the box represents the median which is the middle value if we sort the observations. In other words, the line inside the box represents the middle value if the observations are sorted in ascending order. The individual points represent values that have unexpected behaviour (outliers).

In general, the boxplot shows how the fuel economy is distributed between each category. In compact cars, the average fuel efficiency is approximately 23MPG, Midsize/Large cars have an efficiency of about 20MPG, the other category has an efficiency of about 24MPG, the SUVs have an efficiency of about 17MPG, and lastly the Trucks and Vans have an efficiency of about 16MPG. From this trend we can infer firstly, how different vehicle types really affect the fuel economy but also how weight and fuel economy have a negative correlation, as when the vehicle gets larger, i.e., the weight increases the fuel economy drops. Furthermore, the drop in fuel economy between SUVs and Pickup Trucks potentially shows how aerodynamics also has an effect on fuel economy, the pick- up trucks open bed creates more drag as the air flow is not as smooth- hence shown by the difference in fuel economy in the data.

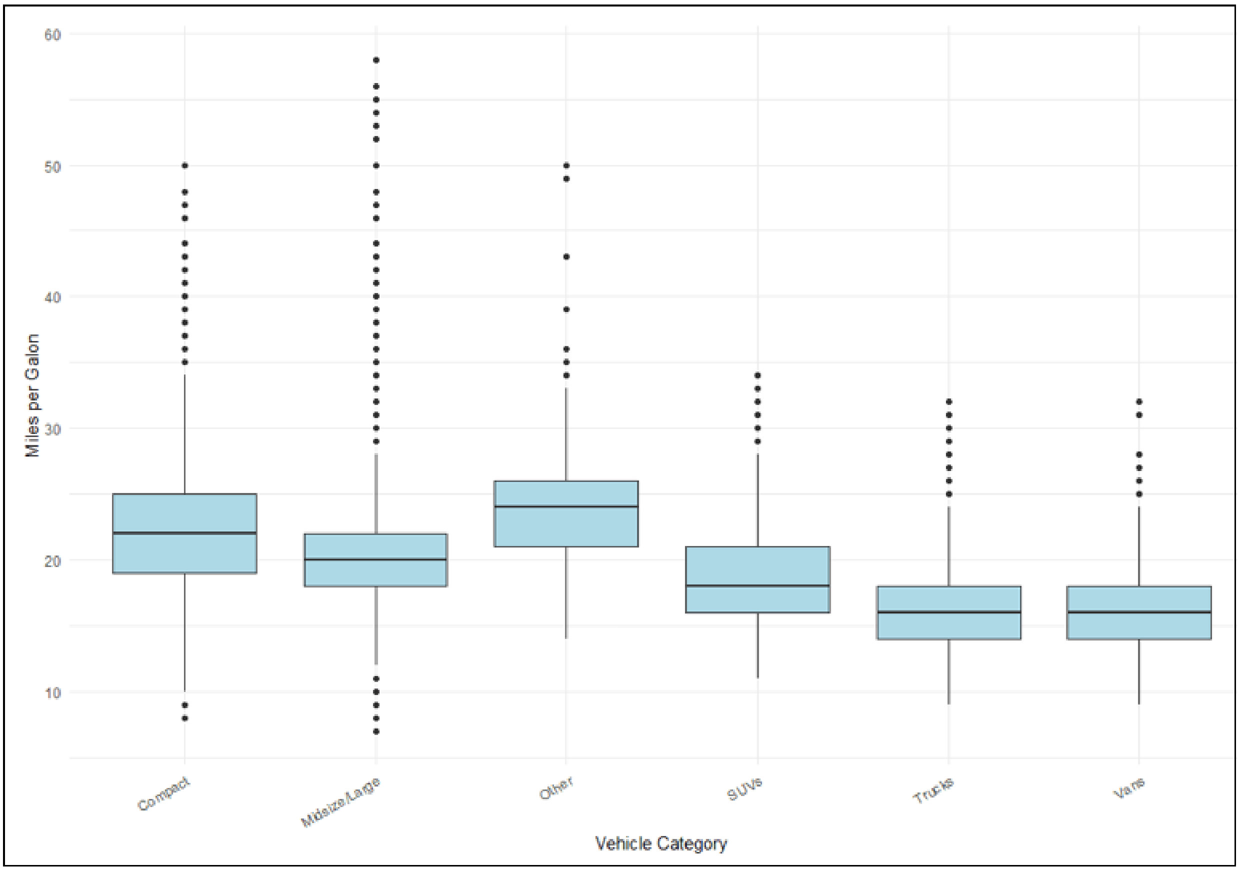

Figure 1: Boxplot of Miles Per Gallon for each Vehicle category.

Another visual investigation we went through is for MPG against Engine displacement. MPG and Engine displacement are negatively correlated, meaning as Engine Displacement increases, MPG drops, hence fuel efficiency decreases. This is observed in Figure 2. Furthermore, Figure 2 also depicts how the number of cylinders affects fuel efficiency, as Engine Displacement and Number of Cylinders have a strong Positive Correlation.

International Journal of Science and Research (IJSR)
ISSN: 2319-7064
Impact Factor 2024: 7.101

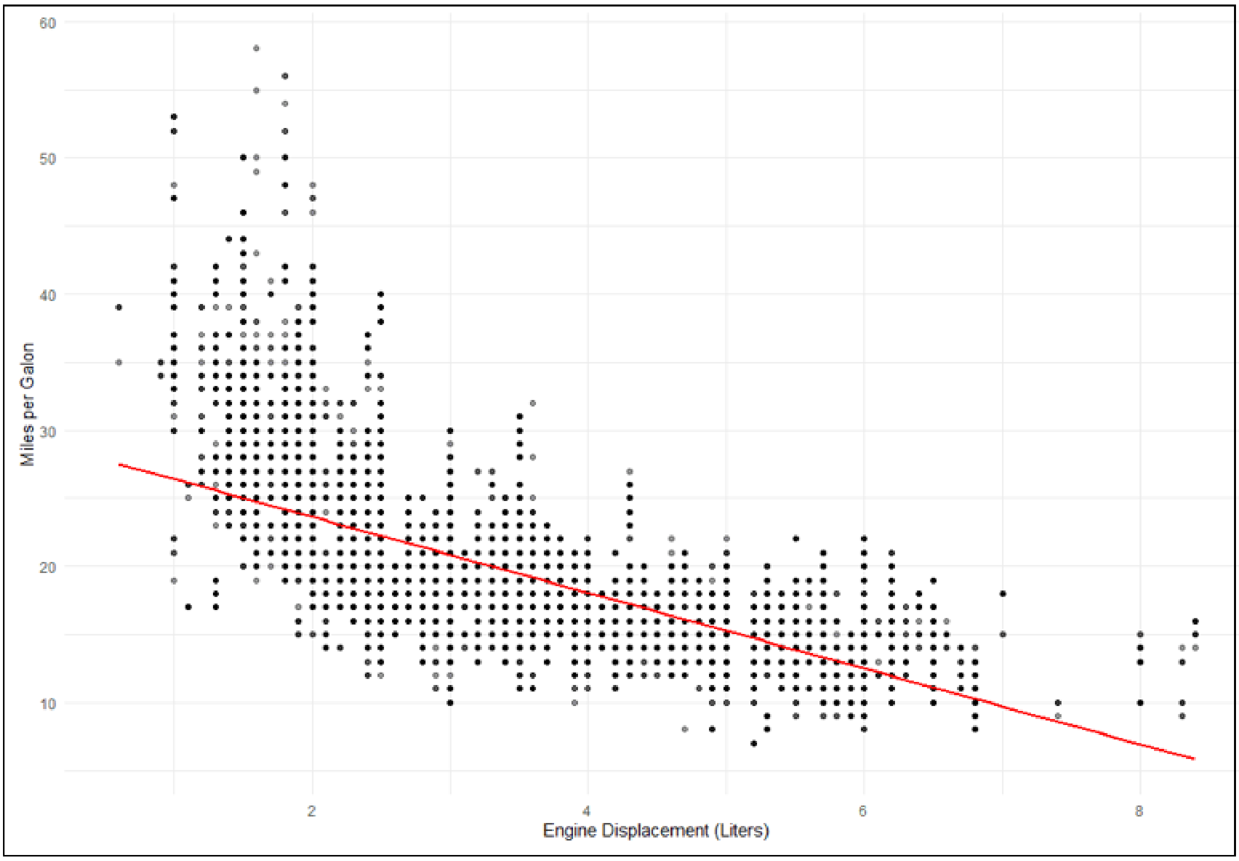

Figure 2: Scatter Plot of Miles Per Gallon against Engine Displacement in Litres.

To further make predictions on this data, two machine learning algorithms were used; these algorithms had 70% of the total dataset to be used as training data, with the 30% to be used as a testing dataset. This split ensures the algorithm has enough data to be well-trained and leaves the testing part of the dataset to be sufficient to confirm whether the algorithm is working properly or not. The two machine learning algorithms we used were Linear Regression and Random Forest and the results, along with a comparison on how the two models perform, are presented in Section 5.

5. Results

We used predictive modelling using Linear Regression and Random Forest, to evaluate how well the vehicle characteristics we discussed in Section 4 can predict fuel economy. Two predictive models were developed: a linear regression model and a random forest model. The data were randomly split into 70% training data and 30% testing data. Both models were trained on the same set of variables which are engine displacement, number of cylinders, and vehicle category. Their performance was evaluated using the test dataset.

Linear regression was used as a baseline model given its simplicity and ease of interpretation. The linear regression approach assumes a straight-line relationship (the line of best fit) between the predictors and fuel economy. The linear regression model achieved a prediction error (RMSE) of 3.15 MPG, indicating that its predictions (performed using the testing data) were, on average, about 3.15 MPG away from the actual values.

The random forest model, which is a machine learning method that combines many decision trees and can capture more complex and non-linear relationships, performed better. It achieved a lower RMSE of 2.64 MPG, showing improved predictive accuracy.

Variable importance results from the random forest model indicate that vehicle category, number of cylinders, and engine displacement all play important roles in predicting fuel economy. This means that whether a vehicle is a compact car, SUV, truck, or van has a stronger impact on fuel efficiency than engine size or cylinder count alone. The number of cylinders and engine displacement are also important, but their influence is smaller compared to vehicle type. This result is based on evaluating the percentage (%) increase in MSE. In general, the higher the % increase in MSE, the more important the variable is. If removing a variable causes the prediction error to increase, then it is a clear sign that the variable contributes effectively to predicting the response of interest. In this paper we discussed different factors of fuel economy and according to results vehicle type is the most important factor in predicting fuel economy, as the % increase in MSE rises by over 64.1% with Vehicle Category included, following 41.1% and 37.2% for the number of cylinders and engine displacement, respectively; see Table 1 below.

Table 1: Percentage (%) in MSE for the three factors used for predicting MPG

Variable	% Increase in MSE
Vehicle category	64.1
Number of Cylinders	41.1
Engine Displacement	37.2

Overall, the results suggest that while linear regression provides useful insights, the random forest model is better suited for accurately predicting fuel economy due to its ability to capture complex interactions between vehicle characteristics.

6. Conclusion

This paper discussed the topic of fuel economy and its different factors. We analysed how these factors truly affect fuel economy in their independent ways. For the analysis, we used the US Environmental agency data focusing on the factors, Vehicle Type, Number of Cylinders and Engine Displacement. However, we can use the dataset to differentiate between weight and aerodynamics as well due to there being clear differences in both of these between different vehicle types.

The results we got from applying linear regression and random forest suggest (with RMSE of 3.15MPG and 2.64MPG respectively) that vehicle category is the most significant factor affecting the fuel economy as there is a 64% increase in MSE when vehicle category is included. Next is the number of cylinders with a 41.1% increase in MSE, and next is engine displacement with a 37.2% increase in MSE. MPG is highest for the Other Vehicle category, then Compact Cars, Midsize/Large Cars, SUVs next and lastly Pickup Trucks and Vans with the same efficiency.

A limitation of this work is that we did not have access to the weight information of the vehicles we studied. Having this in mind, a future objective is to update the analysis taking into account more variables that are affecting MPG and fuel economy, but also, consider more ML approaches like XGboost.

References

[1] Ahmed, R., & Stater, M. (2023). Automobile technology and tradeoffs: How does automobile footprint affect its fuel economy? Transportation Research Interdisciplinary Perspectives, 21, 100897
[2] Atabani, A. E., Badruddin, I. A., Mekhilef, S., & Silitonga, A. S. (2011). A review on global fuel economy standards, labels and technologies in the transportation sector. Renewable and Sustainable Energy Reviews, 15(9), 4586-4610.
[3] Bento, A., Gillingham, K., & Roth, K. (2017). The effect of fuel economy standards on vehicle weight dispersion and accident fatalities (No. w23340). National Bureau of Economic Research.
[4] Cheah, L. W. (2010). Cars on a diet: the material and energy impacts of passenger vehicle weight reduction in the US (Doctoral dissertation, Massachusetts Institute of Technology).
[5] Cooper, K. R. (2003). Truck aerodynamics reborn-lessons from the past. SAE transactions, 132-142.
[6] Essenhigh, R. H., Shull, H. E., Blackadar, T., & McKinstry, H. (1979). Effect of vehicle size and engine displacement on automobile fuel consumption. Transportation Research Part A: General, 13(3), 175-177.
[7] Johnson, T., & Joshi, A. (2018). Review of vehicle engine efficiency and emissions. SAE International Journal of Engines, 11(6), 1307-1330.
[8] Lam, T. N. (1985). Estimating fuel consumption from engine size. Journal of Transportation Engineering, 111(4), 339-357.
[9] Mirmahdi, E., Karimi, M. H., Khoubrou, A., & Sajed, S. A. (2021). The effect of aerodynamic forces on automotive design and reducing fuel consumption. International Journal of Robotics and Automation, 7(1), 36-41p.
[10] UNITED STATES ENVIRONMENTAL PROTECTION AGENCY. (1976). Factors Affecting Vehicle Fuel Economy. UNITED STATES ENVIRONMENTAL PROTECTION AGENCY.
[11] Vehicle Technologies Office, Energy.gov. (2013b, November 22). Internal Combustion Engine Basics.
[12] Xu, L. (2025). A Comparative Study of Engine Configurations: Fuel Consumption, Efficiency, and Application-Specific Advantages. Science and Technology of Engineering, Chemistry and Environmental Protection, 1(5).
[13] Yuan, Z., & Wang, Y. (2017). Effect of underbody structure on aerodynamic drag and optimization. Journal of Measurements in Engineering, 5(3), 194–204.