Springboard Data Science Career track  
Capstone Project Final Report  
Future trend in Fuel Economy, CO2 Tailpipe Emission

Prepared by: Dalbir Singh

Mentor: Raghunandan Patthar

**Introduction**

This report consists of a unique dataset that helps to get current updates on fuel economy and carbon dioxide (CO2) emissions trends for new personal vehicles in the United States. This data supports the reports since 1984 – 2018. The U.S Environmental Protection Agency (EPA) supports the U.S. Department of Transportation’s National Highway Traffic Safety Administration (NHTSA), Fuel Emission Laboratory, and Corporate Average Fuel Economy (CAFE) programs.

Demand for oil has steadily increased globally; the most significant reason behind this demand is ample because of rapid car consumption taking place in developing countries with rapidly growing economies, such as India, China, and others. Increasing demand in oil is primarily because of the population growth of vehicle consumers, mainly private vehicles and the distance these vehicles travel. Over the years, it has become critical to control the green energy demand toward our personal vehicles according to greenhouse gas (GHG) emissions. Limiting the vehicle population growth by reducing the travel demand by providing efficiency fuel vehicles must the top three critical elements to reduce the overall oil demand. There are many varieties of methods which needs to address because these three areas have introduced in different parts of the world. Since the 1970s, the United States has been the first county to create a fuel economy standard for passenger vehicles since the oil crisis.

Nevertheless, standards since have been remained unchanged for a quarter century from the 80s to late 2000s. According to the EPA, the state of California since have moved forward to established and tightened the emission/fuel economy standards. Meanwhile, recognizing the threat of possible oil shortage and climate change the efforts to strengthen the eco-friendly vehicle standards have been encouraging globally by the United States. The goal of this paper is to provide an updated analysis of the current trends in the United States by identifying best practices of tailpipe emission, vehicle fuel economy, and make recommendations for future policymaking.

**What is Tailpipe Emission?**

“Automobiles typically run on gasoline or diesel fuel, both of which are a composite of different chemicals collectively known as hydrocarbons. Mainly carbon and hydrogen, hydrocarbons are Volatile Organic Compounds (VOCs), meaning that they readily evaporate at normal temperatures. According to the Environmental Protection Agency, they are a precursor to ground-level ozone and smog. Hydrocarbons result in dangerous emissions because our vehicles do not operate in a vacuum. In a “perfect” combustion engine, all the hydrogen in the fuel would be converted to water by the oxygen in the air. Meanwhile the carbon in the fuel would be converted to carbon dioxide, and the nitrogen already in the air would remain unaltered. However, as is clearly the case, we do not operate in a vacuum, and none of our vehicles are equipped with a perfect engine. Instead, incomplete combustion, combined with high pressure and temperature, results in several toxic exhaust pollutants” (2017),:

• Carbon Monoxide

• Nitrogen Oxides

• Sulfur Oxides

• Carbon Dioxide

“In addition to exhaust pollutants, petroleum-based fuels also emit dangerous compounds due to evaporation. Typically, these emissions stem from storage tanks and fueling lines. As has been stated, hydrocarbons are VOCs which means that, especially on hot days, they contribute significantly to ground level ozone as they are released from vehicles (both parked and driving), gas stations, and anywhere gas or diesel is stored. A key component of smog, ground-level ozone is formed by reactions involving hydrocarbons and nitrogen oxides in the presence of sunlight. Particulate matter as well, though not a chemical compound, is nevertheless a serious health concern as an auto emission” (2017).

**What is Fuel Economy?**

The fuel economy of a vehicle can be calculated by comparing distance traveled by a vehicle with the amount of [fuel](https://en.wikipedia.org/wiki/Fuel) consumed. Consumption can be expressed in terms of volume of fuel to travel a distance, or the distance travelled per unit volume of fuel consumed. For gasoline vehicles, the label shows City, Highway, and Combined MPG (miles per gallon) values. The Combined MPG value is the most prominent for the purpose of quick and easy comparison across vehicles. Since 1977, there has been label on vehicles stating miles per gallon. Combined fuel economy is a weighted average of City and Highway MPG values that is calculated by weighting the City value by 55% and the Highway value by 45%.

**The Goal:**

Through this paper I will like to explain the following:

1. Change in fuel economy over the period

2. Compare fuel economy before and after the 2009 Recession.

3. Compare the CO2 Emission from tailpipe with cars engine size.

**The Dataset:**

This paper uses a unique data set of Fuel economy by detailed model from 1984-2018, and the results of vehicle testing done at the Environmental Protection Agency's National (EPA) Vehicle and Fuel Emissions Laboratory in Ann Arbor, Michigan, and by vehicle manufacturers with oversight by EPA. This Dataset contains 37936 rows and 49 columns. The following are the main features that will be used in the data set:

* Combined, City and Highway Fuel Mileage of different Models of cars
* Number of Cylinders in the cars
* Tailpipe Emission for different models
* Engine size of different models of cars
* Change in cars over the years
* EV and hybrid cars over the years

**Data Wrangling and Cleaning**

The data sets were relatively comprehensive and well put together and hence did not require any major transformations. However, there were quite a few features that would not be required for the analysis, along with features that were made up primarily of NAN values. Formatting was required on several columns in the project such as cylinder size, displacement, tailpipe emission etc. as some cars were electric so they don’t all those things. The following shows the steps undertaken to transform the data into a desirable format suitable for the analysis.

**Deleting features**

The first task was to delete any obvious columns that would not be of any benefit for the analysis. Once this was done, further investigation into each feature was carried out to determine the non-NAN values. We use replacement method to replace the values to NAN values using following method.

print (gas.replace(r'', np.nan, regex=True))

Once we replaced the Non-NAN values, we will convert all the other values to Zero so that it won’t affect our analysis. We used following technique to solve this problem.

gas['cylinders'].fillna(0, inplace=True)

gas['drive'].fillna(0, inplace=True)

gas['displ'].fillna(0, inplace=True)

gas['eng\_dscr'].fillna(0, inplace=True)

gas['trany'].fillna(0, inplace=True)

gas['guzzler'].fillna(0, inplace=True)

gas['trans\_dscr'].fillna(0, inplace=True)

gas['tCharger'].fillna(0, inplace=True)

gas['sCharger'].fillna(0, inplace=True)

gas['atvType'].fillna(0, inplace=True)

gas['fuelType2'].fillna(0, inplace=True)

gas['rangeA'].fillna(0, inplace=True)

gas['evMotor'].fillna(0, inplace=True)

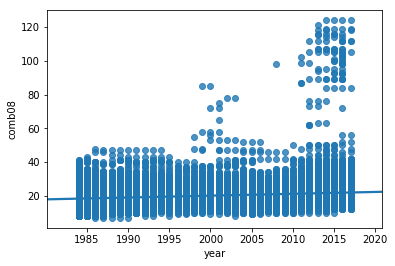
gas['startStop'].fillna(0, inplace=True)

Zero was used for all the values which were not provided.

**Data Exploration**

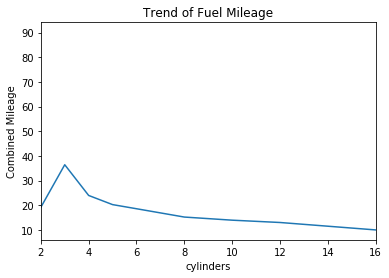
The purpose of the initial exploration was to gauge the relationship between the fuel mileage over the years and compare fuel economy before and after the recession of 2009. We will also compare the relationship between Tailpipe Co2 Emission and the Cylinder size of the engine. The objective of this was to take a selection of the more influential variables in the data set to crudely determine the variables that could show us the better picture of data. Once determined, the focus of the further data visualization and exploration will be to deep dive into these variables and determine how they interact with one another. The following are the variables to be investigated in this section: comb08, city08, highway08, range08, Co2TailpipeGpm, cylinders, displ etc. This will be done by plotting year against each of the variables of interest.

**Fuel Mileage trend:**



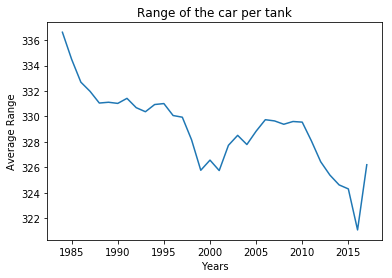
As expected, there is a change in gas mileage trend since 1985. The graph above shows us that the fuel economy is getting better with time. As we can see from the graphs there are some points close to 100 mpg that is because with time mankind has made hybrid cars that use gas and electricity as fuels. These hybrid cars have great fuel mileage which is getting better with time.

**Effect of Cylinder with Fuel Mileage**



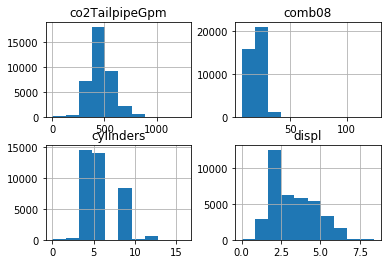
The graph above shows us the Fuel Mileage with comparison to Cylinders in the cars. It shows us that number of cylinders has negative effect on mileage.

**Range of car per tank**



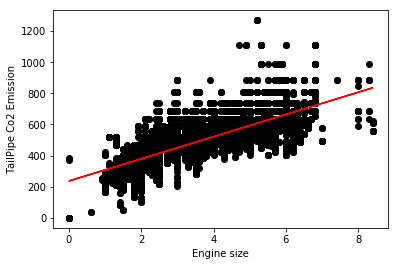
The graph shows us the range a car can travel with a full tank of gas. It also tells us that with time gas tanks are becoming smaller as cars are getting smaller.

**Data**

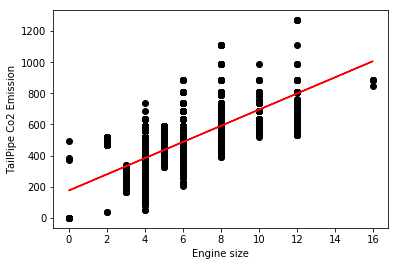


Here we plot some graphs that shows us more details into our data. We have graphed Tailpipe Co2 emission, Combination gas mileage, number of cylinders, displacement of engine. It shows us more incite of our data.

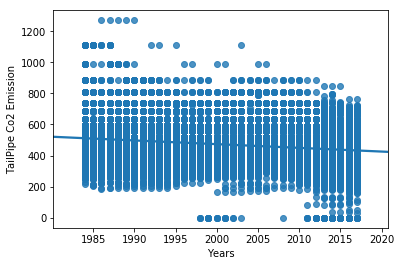
**Tailpipe Emission – Engine Displacement Graph:**



In this Graph, we are comparing Tailpipe Emission with Engine displacement. It shows us that the size of engine displacement is positively corelated with Tailpipe emission, bigger the size of engine displacement more the tailpipe emission.

**Cylinder size – Tailpipe Emission Graph:**

This graph shows us that number of cylinders in a car is directly related to emission from the tailpipe. We also plotted a linear model that shows us the trend.

**Tailpipe Emission trend:**

The figure above shows us the trend of Tailpipe emission over time. As we can see that with time the tailpipe emission is getting lower.

**Machine Learning technique:**

We used R-squared (R2) and Ordinary Least Squares (OLS) techniques.

**R-Square: -**

“R-square is a statistical measure that represents the proportion of the variance for a dependent variable that's explained by an independent variable or variables in a regression model” (Hayes, 2019). The score of the relationship explains the correlation between a dependent variable and an independent variable. The score explains the variance of one variable to the second variable. So, If R square value is 0.50, that means that the model can explain half of the variance. The Formula For R-Squared Is:

R-square = 1 – {Explained Variation / Total Variation}

R-squared value is calculated by a line that best fits the data points of dependent and independent variables. Then we calculated the predicted values, Mean Squared Error, and Mean Absolute Error. We estimate total variance by subtracting the average actual value from the predict values and then sum of the square of results. To calculate the R-square, we will divide the explained variation with total variation and subtract it from one.

We used the R square method to calculate the variance between Cylinder, Combined fuel mileage, and Tailpipe emission, which came out to be 0.76. The score tells us that there is a relation between the variables. We separated training and test data, and then we used a linear regression model to predict the relationship. It shows us that there is 76% of chance that the variables are related. We also calculated the variance score between Displacement, Combined fuel mileage, and Tailpipe emission, which came out to be 0.79. It shows us that there is an almost 80% chance that our variables are related to each other.

**Ordinary Least Squares (OLS):**

OLS is a method to measure the approximate value of unknown parameters in a linear regression model. We used OLS to calculate the R square value and F-Statistic. With this method we can analysis the relationship between one or more independent variables and a dependent variable. The results show us that our variables are related to each other. It shows us that our dependent variables like combination fuel mileage, cylinder, tailpipe emission, etc. are related to independent variable years. The R-square value is 0.47, F-statistic is 280.

**OLS Regression Results**

==============================================================================

Dep. Variable: year R-squared: 0.414

Model: OLS Adj. R-squared: 0.414

Method: Least Squares F-statistic: 6700.

Date: Tue, 25 Jun 2019 Prob (F-statistic): 0.00

Time: 16:36:17 Log-Likelihood: -1.3262e+05

No. Observations: 37936 AIC: 2.653e+05

Df Residuals: 37931 BIC: 2.653e+05

Df Model: 4

Covariance Type: nonrobust

==================================================================================

Coef. std err t P>|t| [0.025-0.975]

----------------------------------------------------------------------------------

Intercept 2001.8295 0.202 9933.020 0.000 2001.434 2002.224

co2TailpipeGpm -0.0338 0.001 -54.897 0.000 -0.035 -0.033

ghgScore 2.2387 0.019 119.484 0.000 2.202 2.275

displ 1.1480 0.078 14.776 0.000 0.996 1.300

cylinders 1.8443 0.054 33.904 0.00 1.738-1.951

==============================================================================

Omnibus: 5256.038 Durbin-Watson: 0.709

Prob(Omnibus): 0.000 Jarque-Bera (JB): 1362.486

Skew: -0.117 Prob (JB): 1.38e-296

Kurtosis: 2.102 Cond. No. 2.47e+03

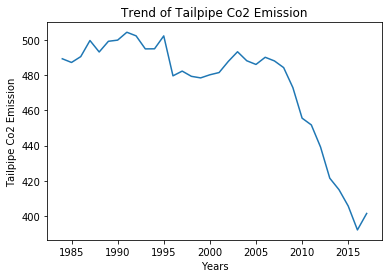
=========================================================================

Above OLS Regression shows us the relationship between years and number of cylinders, tailpipe Co2 emission, engine displacement, ghg score. It tells us that R square value is .414, F-statistic value is 6700. It tells us that our Null hypothesis is wrong. There is a relationship between all our variables.

**The Trends**

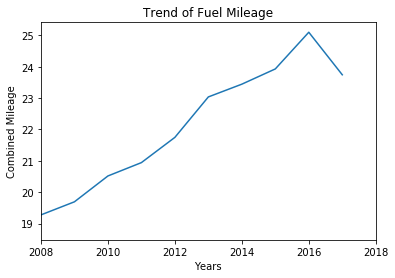
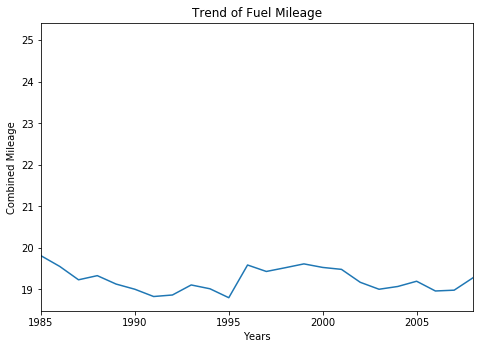
**Tailpipe Co2 Emission Trend:**

The average 2016 real world CO2 Tailpipe emissions rate for all new personal vehicles is 392.24 g/mi, which is a 13.55 g/mi decrease from 2015. The 2016 Tailpipe Co2 emission was the lowest ever recorded. Since 2006, Co2 Tailpipe emission has decreased by 97.84 which is 19.96% till 2016 but there is an increase in Tailpipe Co2 emission for 2017. In 2017, the emission has increased by 9.32 g/mi. The trend still shows that the Tailpipe Co2 emissions are going in downward direction. The line graph below shows us how Tailpipe Co2 emission has changed over time.



**Fuel Economy Trends:**

The 2016 combined fuel economy is 25.099 mpg, which is 1.16 mpg higher than 2015, and is a highest ever recorded. The fuel economy has increased by 6.14 mpg which is 24.46%, with an average annual improvement of about 0.5 mpg per year. In 2017, the fuel economy has decreased to 23.74 mpg. The trend shows us that in future it will be going even higher. Fuel mileage will rise even further. The graph below shows us change in fuel mileage before and after year 2009. From the graph till 2008 the fuel mileage was quite steady but since 2008 the fuel economy has improved drastically. This shows us how the market has changed. Customer are more interested in buying cars with better fuel mileage, even the vehicle manufactures have started making cars with better fuel mileage. Most vehicle giants have started making smaller cars that give birth to a new segment of smaller SUV’s. The segment is growing at a brisk rate.



**Client Recommendation**

* According to EPA the United States have recently releases its own of “targets in a range of 46 to 60 MPG by 2025”. As DiCicco suggested that enhancing the engines on the vehicles and the use of non-grid enabled hybrid technology, and the fuel economy of the vehicles could be tripled by 2035. This should give self-assurance to any policy maker who are determined to target for fuel efficiency are not only desirable but in fact achievable.
* Due to globalization, it is difficult for most manufacturers to meet the standards of different countries around the world because each country has its regulation of environmental standards. As the world becomes globalized, it is in the interest of the nations around the world to have similar ecological laws that will help the auto industry to make global products. With similar laws around the world, it will help developing countries have access to the latest technology.
* With the development of new electrical and semi-electrical engines that have better fuel economy and lower tail pipe, carbon emission can solve the issues of climate change.
* As the paradigm shift from gasoline vehicles to electric vehicles, the standards for regulation of electric cars should also be changed as electric cars don’t have the same emission standards as gasoline cars. To regulate electric vehicles, we can use the same methods that are used in electricals power sectors.
* With the ever-increasing number of vehicles sold around the world, many governments around the globe do not know how to measure, report, and regulate the emissions from these gasoline vehicles. The best practice for policymakers around the world will be to implement global standards and share technology globally, that will promote a greener and healthier world.
* To address the ongoing challenges of the new technology vehicle which cheat the emission regulations by shifting the emissions from vehicle tailpipe to a different source. To fight this, we suggest that future vehicle emission regulations should focus on a complete lifecycle based GHG emission assessment.

Reference:

1. DECICCO, J. M. (2019). . Retrieved from http:// [https://energy.umich.edu/energy- aculty/faculty/john-m-decicco/](https://energy.umich.edu/energy-%20aculty/faculty/john-m-decicco/)

2. Light Duty Automotive Fuel Economy....Trends Through 1979

J. Murrell - SAE Technical Paper Series – 1979

1. Greenhouse Gas Emissions from a Typical Passenger Vehicle (2018). Retrieved from <https://www.epa.gov/greenvehicles/greenhouse-gas-emissions-typical-passenger-vehicle>
2. Gasoline Vehicles: Learn More About the Label (2017). Retrieved from <https://www.fueleconomy.gov/feg/label/learn-more-gasoline-label.shtml>
3. Hayes, A. (2019). R-Squared Definition. Retrieved from https://www.investopedia.com/terms/r/r-squared.asp

5. Shared MPG Estimates (2019). Retrieved from 5. <https://www.fueleconomy.gov/mpg/MPG.do?action=browseList>

6. Vehicle Fuel Economy Data (2017). Retrieved from [https://datasource.kapsarc.org/explore/dataset/us- ehicle-fuel-economy-data-1984-2017/information/?disjunctive.make&disjunctive.model](https://datasource.kapsarc.org/explore/dataset/us-%20ehicle-fuel-economy-data-1984-2017/information/?disjunctive.make&disjunctive.model)