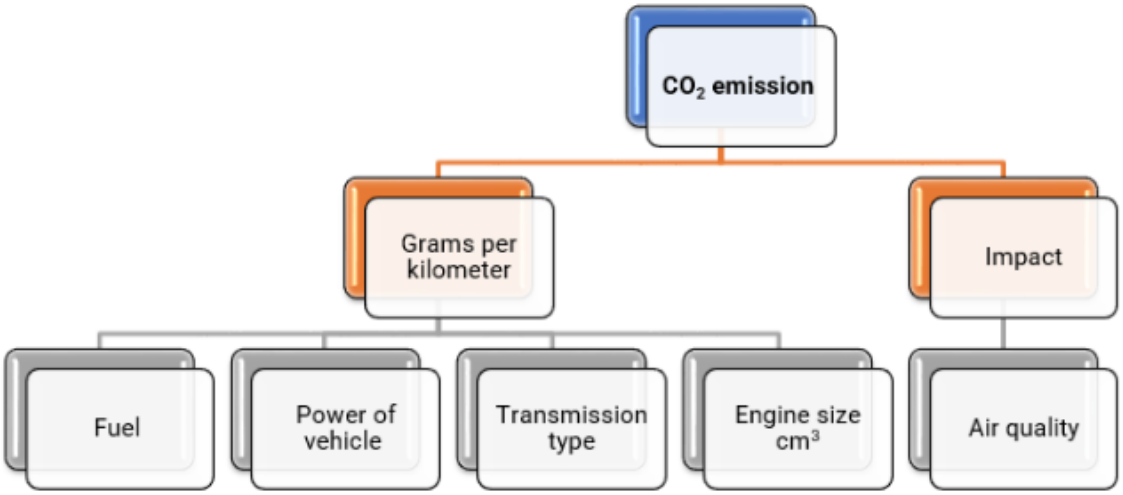
Vehicle CO2 Emissions – EDA

June 2024



Note: CO2 emission characteristics

By William Gray Renton, Isbelis Castro, Gunel Garayeva, and Eugenia Chien – Group 06.
Alexander Booth (Instructor) and Reed Schelitzche (TA).
Project 1 - Rice University Data Analytics and Visualization Boot Camp

1. Introduction

In this project we decided to do some EDA on a Kaggle dataset containing information on vehicle carbon emissions. The dataset also includes other properties of the vehicle along with their respective CO_2 emissions.

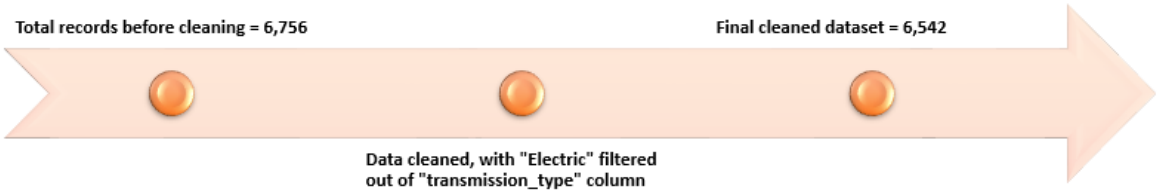
2. Our Dataset

- Source: We acquired our dataset through Kaggle.

Item	Field	Description	Type
1	car_id	A unique identifier for each vehicle.	Float
2	manufacturer	Name of manufacturer e.g. “Tesla”.	Object
3	model	Name of model.	Object
4	description	Brief description of vehicle.	Object
5	transmission	Gearbox identifier if applicable.	Object
6	transmission_type	“Manual”, “Automatic”, or “Electric”	Object
7	engine_size_cm3	Volume of gas displacement in cm3.	Float
8	fuel	Type of fuel e.g. “Diesel”	Object
9	powertrain	Vehicles powertrain	Object
10	power_ps	Power of vehicle in PferdStarke (metric measure of horsepower, equivalent to 98.6% of one HP).	Float
11	co2_emissions_gPERkm	CO_2 emissions (WLTP measurement) in g/km.	Float

3. Data Cleaning

The dataset initially consisted of 6,756 unique vehicles. Upon investigation we noticed that there were many “electric” vehicles in the “transmission_type” column. When removing these vehicles and any other null attributes, we were able to get to a final number of 6,542 rows.



1. Research questions

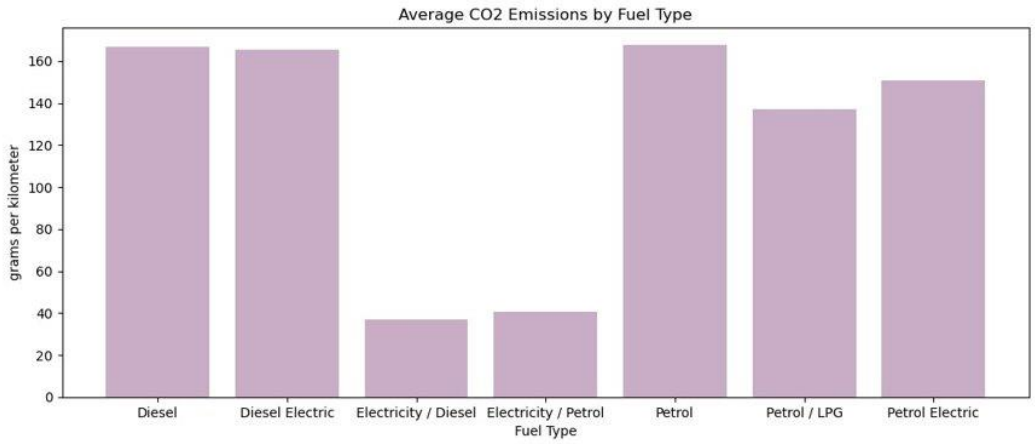
1.1. Research question 1 –

Does the fuel type have an effect on CO2 emissions?

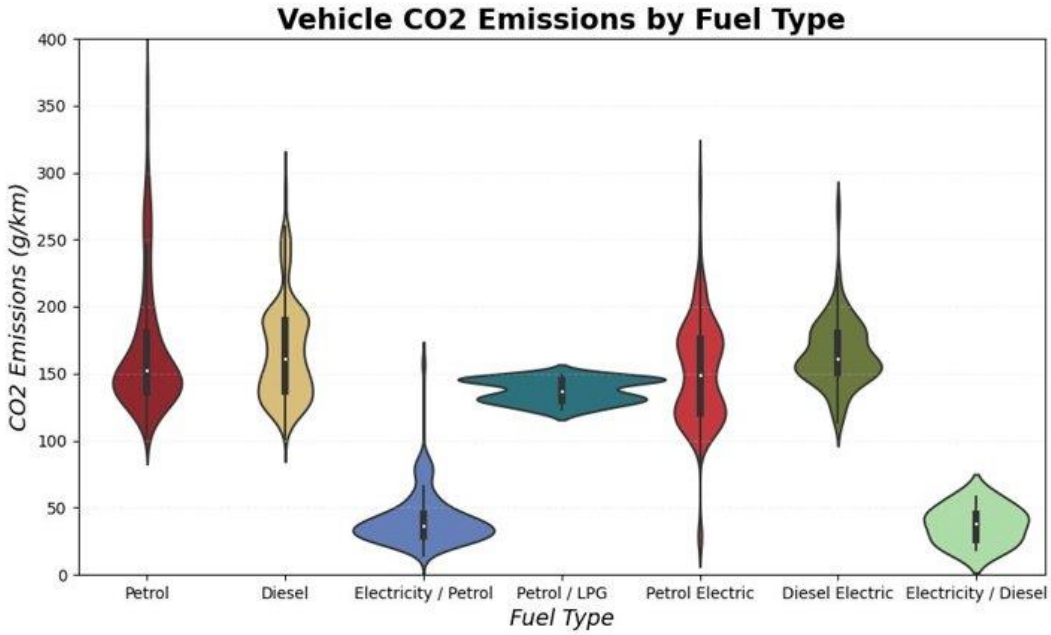
Hypothesis:

H0= CO2 emissions are not affected by fuel type.

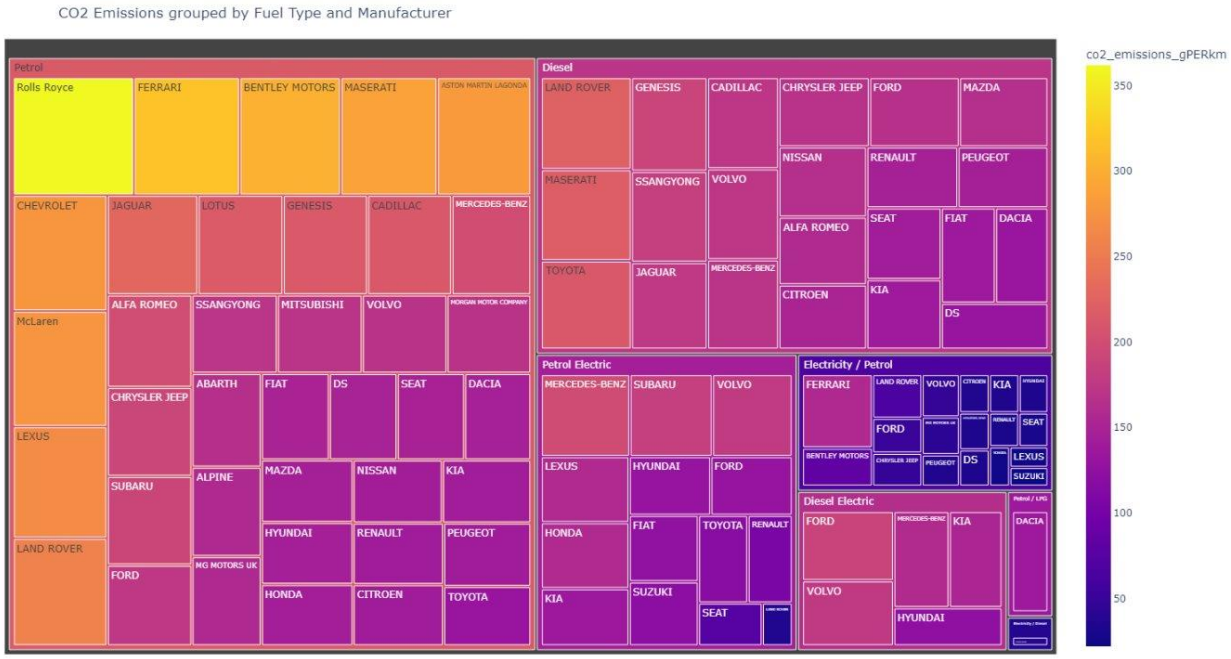
H(a)= CO2 emissions are affected by fuel type.



Petrol, Diesel, and Diesel Electric are the three fuel types that tied for the highest CO2 emissions. This is measured by the industry standard of grams per kilometer. The bottom two are Electricity/Diesel, and Electricity/Petrol.



Here is another look at CO2 emissions by fuel type but this time as a violin chart. This shows us the range and approximate frequency of the data in terms of fuel type. If we go back to the previous slide, using Petrol/LPG as an example, on the bar chart the average CO2 emissions is about 140. Here on the violin chart, we can see that yes, the data ranges from just over 150 to about 120, so the 140 average makes sense. Another insight that we can glean from this is the consistent results that Petrol/LPG and Electricity/Diesel produce. The short and fat shape means there isn't a wide range of data. Compare that with Petrol or Diesel and we can see that the shape is a lot longer and narrower. So based off the consistent results from two graphs, the answer to our question is the alternative hypothesis: "CO2 emissions are affected by fuel type."



This is kind of a bonus visualization. I was interested in seeing which manufacturers were producing the vehicles with the most CO2 emissions. This treemap is grouped by Fuel type, Petrol, Diesel, Petrol Electric etc. The color bar on the right tells us the average CO2 emissions with yellow being the highest at 350 and dark blue being the lowest. We can see that most of the Petrol manufacturers are in the higher range with the warmer colors. The majority of the highest CO2 emission manufacturers are luxury car brands.

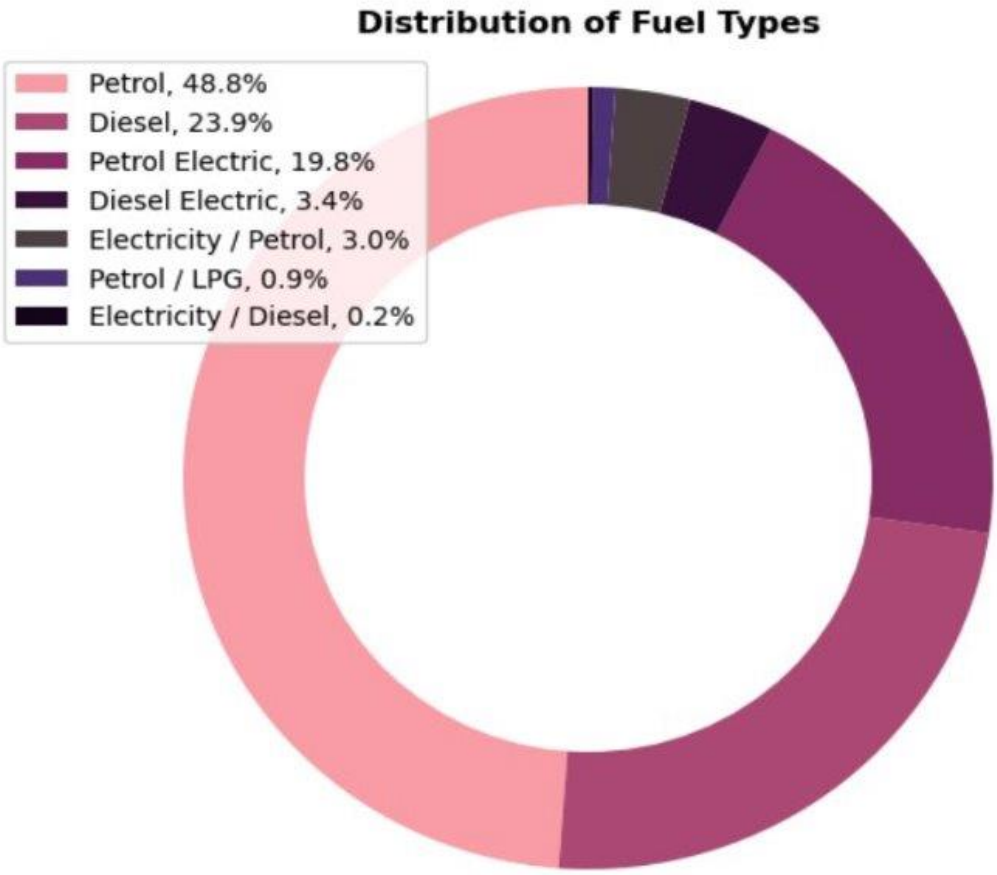
1.2. Research question 2 –

What is the effect of engine power (PS) on CO2 emissions (g/km)?

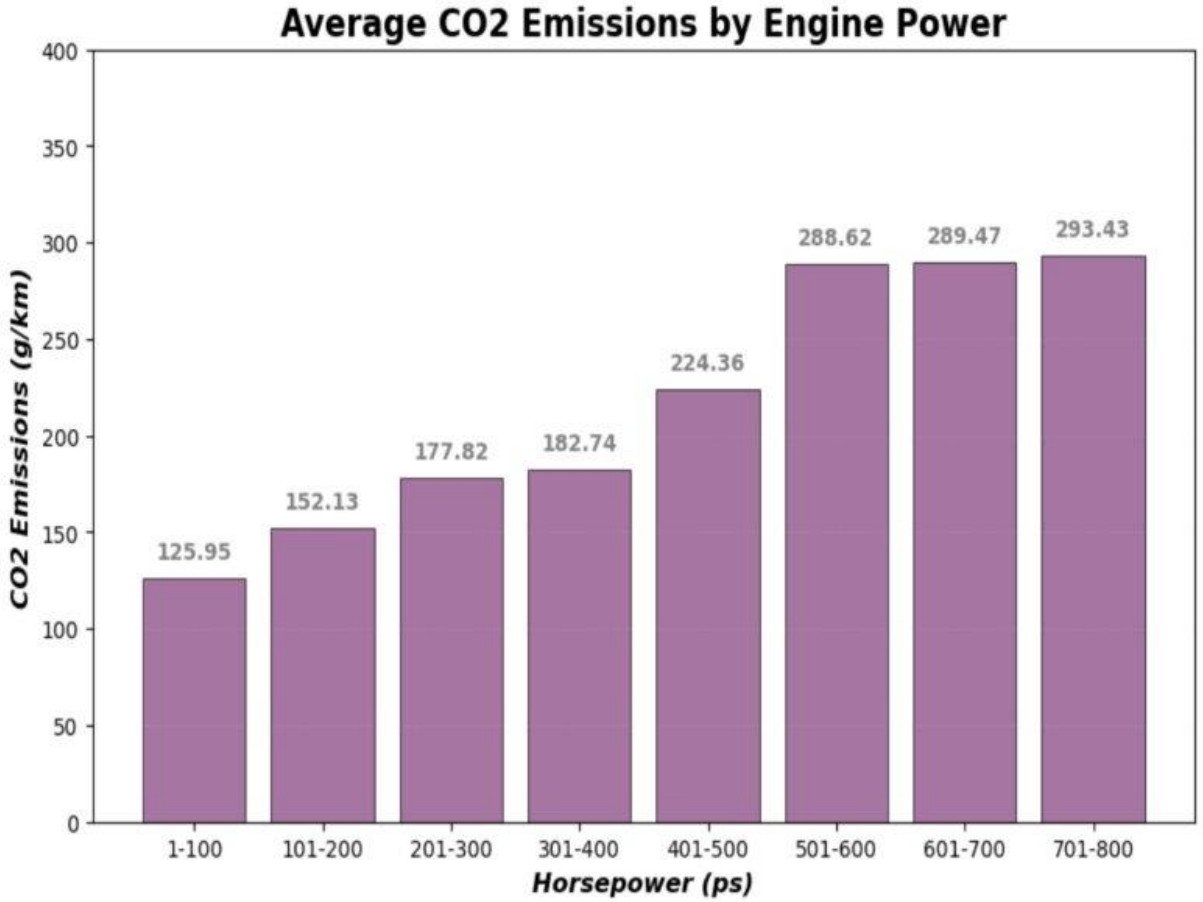
Hypothesis:

H0= CO2 emissions are not affected by engine power.

H(a)= CO2 emissions are affected by engine power.

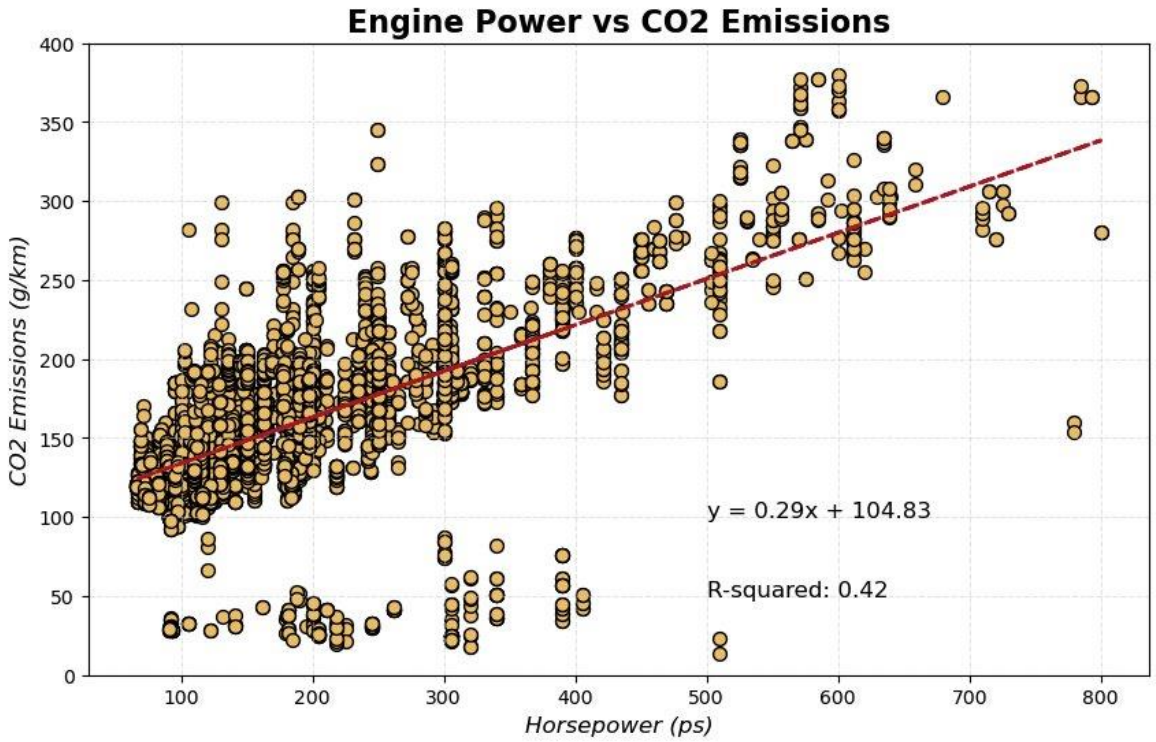


Note: The chart highlights the diverse range of fuel types and combinations used in vehicles, with petrol being the most prevalent fuel type followed by diesel and various hybrid options.



Note:

The data illustrates a correlation between horsepower and CO2 emissions, with higher horsepower vehicles generally exhibiting increased CO2 emissions. Understanding these trends can inform decisions related to vehicle efficiency and environmental impact.
The highest average CO2 emission of 293.43 g/km is recorded for vehicles with horsepower in the 701-800 ps range.



There is a correlation between power output and CO2 emissions in vehicles, but it is not a direct one. Other factors also play a crucial role in determining the overall emissions of a vehicle. Manufacturers are continuously working on improving engine efficiency, reducing vehicle weight, and developing alternative powertrains to minimize the environmental impact of their vehicles.

1.3. Research question 3

Does the type of vehicle transmission have any effect on CO2 emissions?

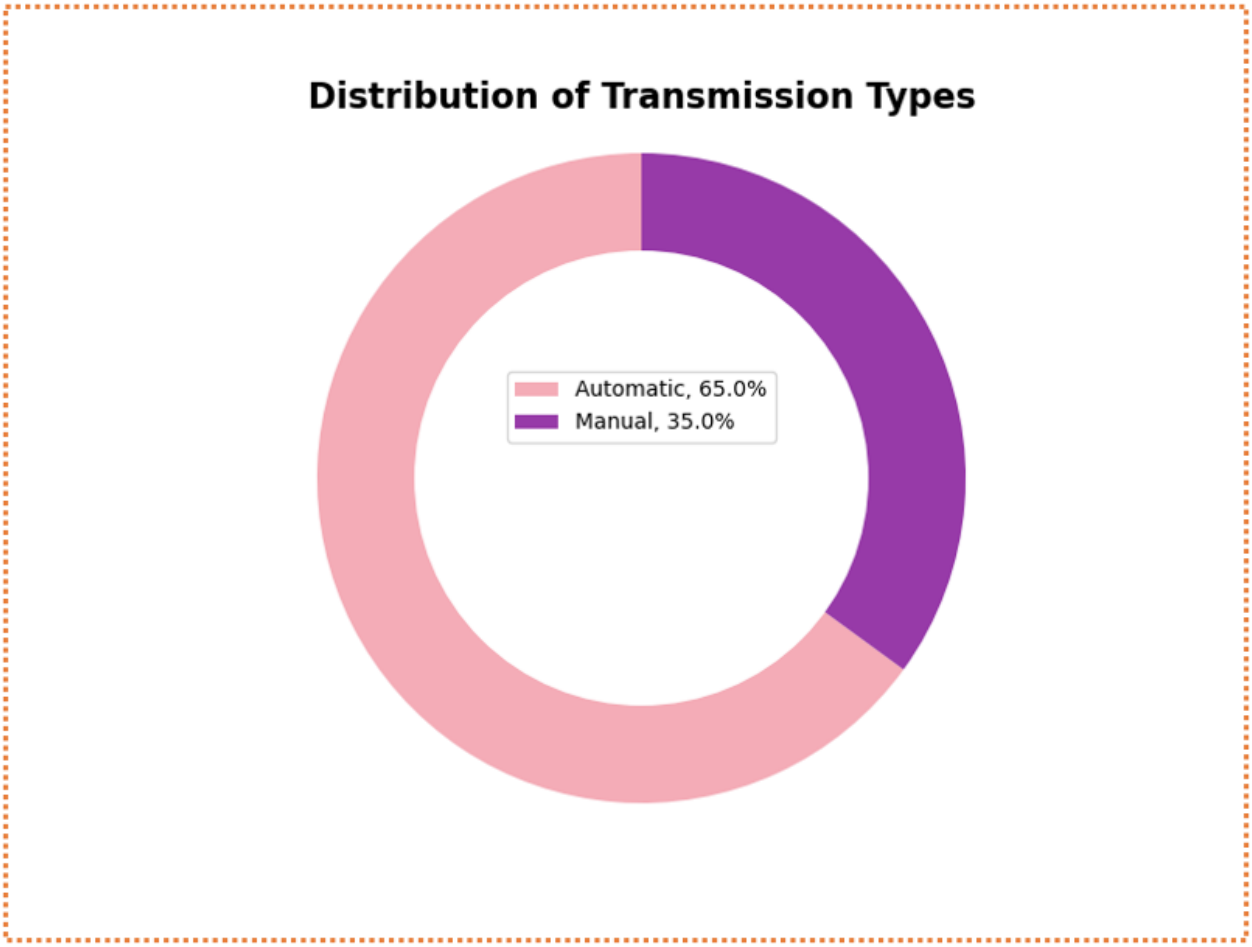
Hypothesis:

H0= CO2 emissions are not different depending on the type of vehicle transmission.

H(a)= CO2 emissions are different depending on the type of vehicle transmission.

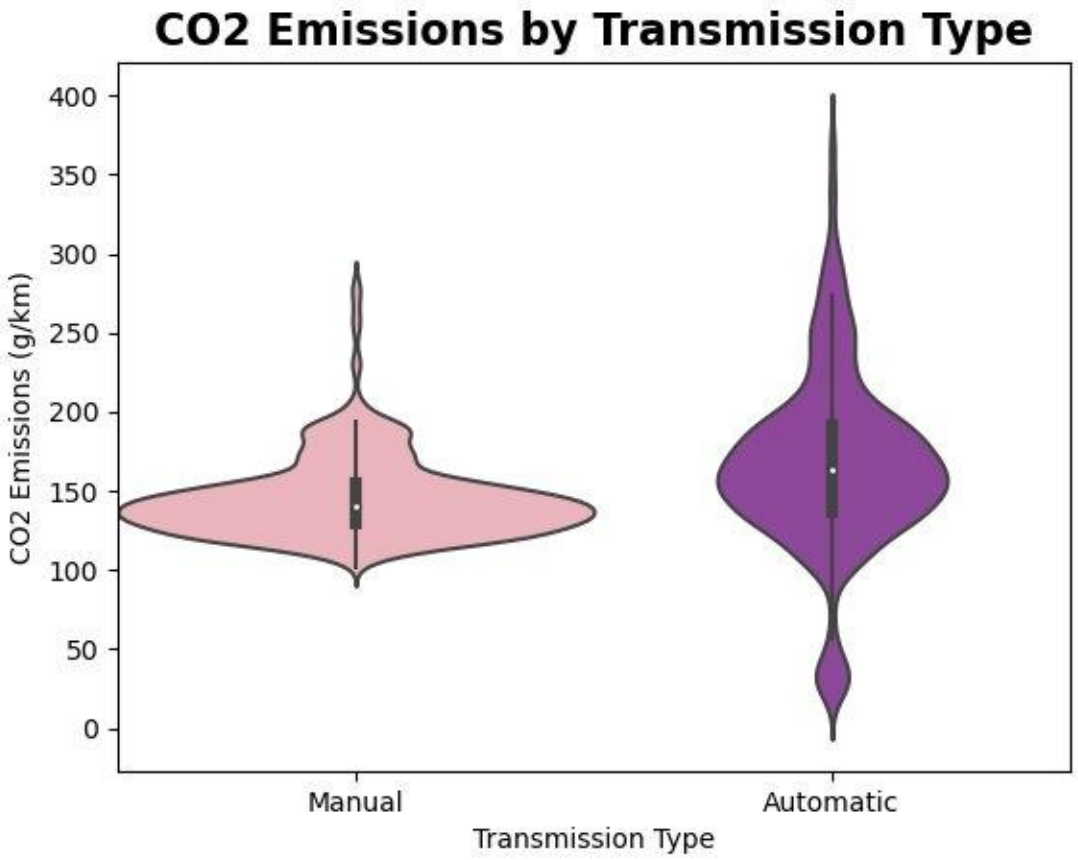
For this analysis, we are using two variables: CO2 emissions as our dependent variable and transmission type as our independent variable.

The following graph shows that description of the transmission type is remarkable for the concentration on automatic transmission, with more than 60%, the sample size (n) is 4363. It is higher than manual.

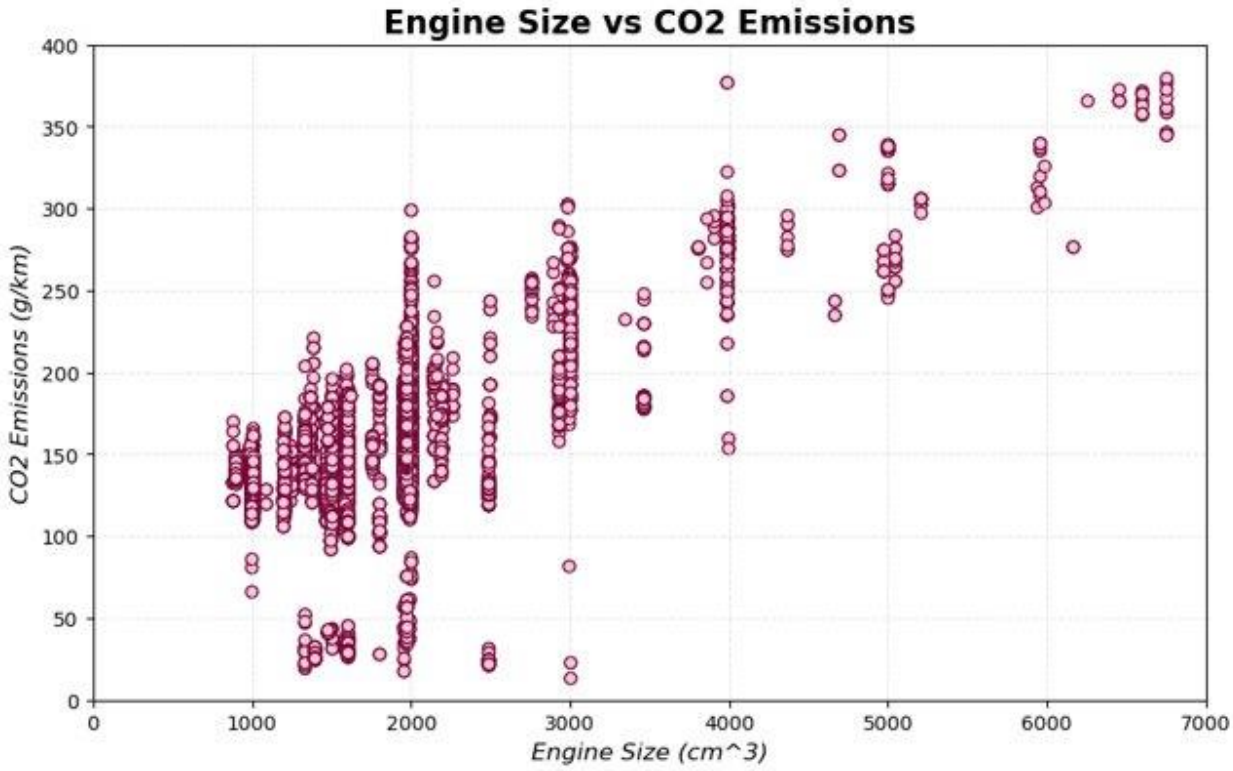


After the analysis of distribution of data, we can say the CO2 emissions for manual and automatic transmission type are not normally distributed. According to the above, analyzing the null hypothesis, to remember "the distribution of CO2 emissions is the same for manual and automatic transmissions". The U statistic of 759 represents the rank sum of the smaller sample in the two-sample comparison. And the difference is significant with a p-value less than 0.005 (0.0064). That is to say, the result is a highly significant difference in CO2 emissions between manual and automatic transmissions because the manual transmission presents lower emissions than automatic transmission, how to shows this violin plot. On this graph can look at variance.

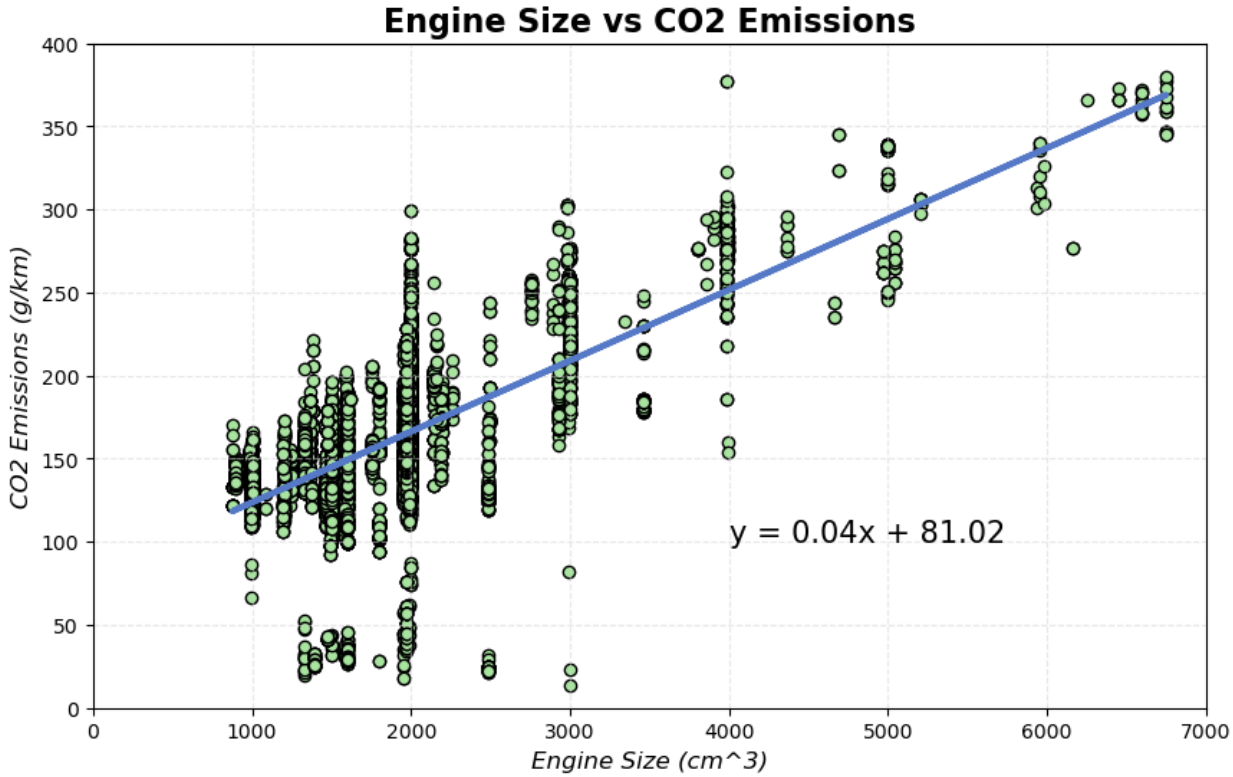
Finally, manual transmission presents lower emissions than automatic transmission, and the difference is significant with a p-value less than 0.005. The result is a highly significant difference in CO2 emissions between manual and automatic transmissions.



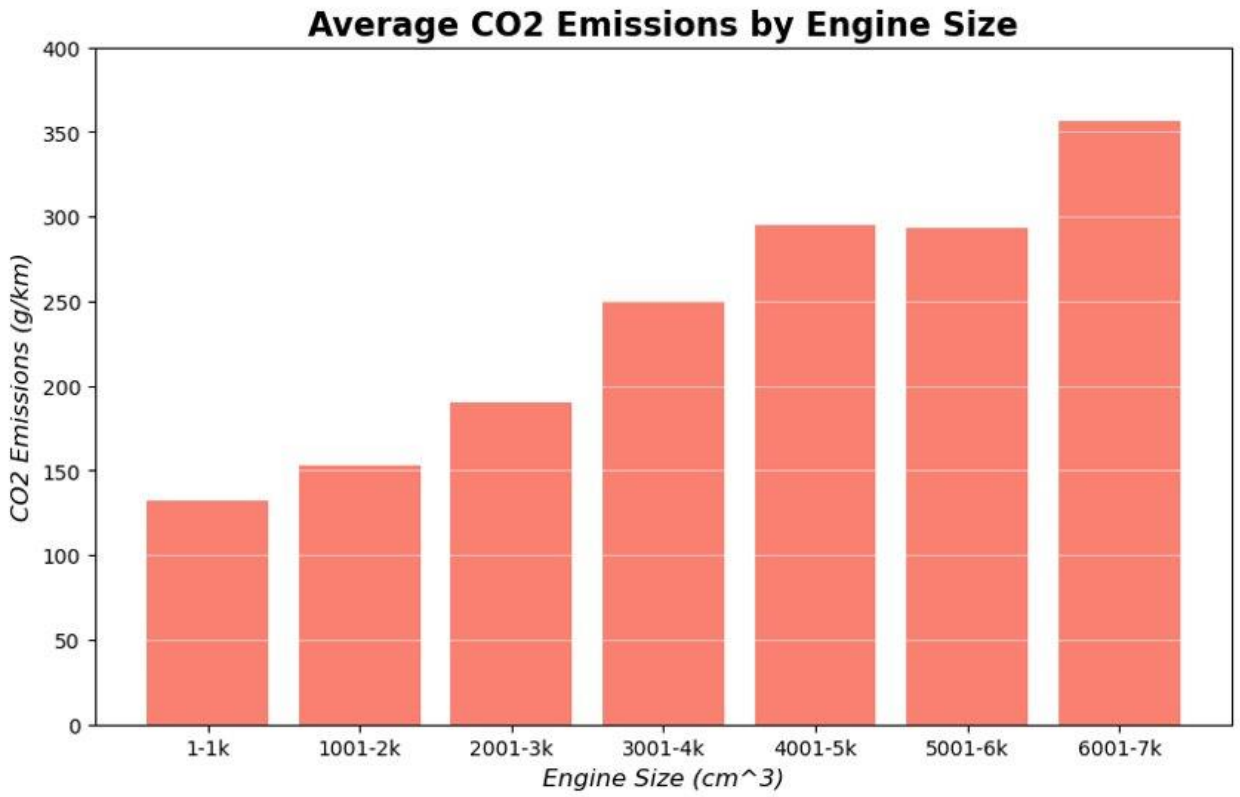
1.4. Additional Analysis–



Much like our engine power vs CO^2 emissions graphic we can make a scatter plot using engine size and see a strong and clear correlation (aside from outliers which consist of hybrid mostly models). With that being said, from this we can do some work to get our best line fit for useful predictions on CO^2.



This is the same graphic as the previous, but with a beautiful line that allows us to make predictions with an R^2 of 0.472. With this If we had an engine with a size of 3000cm^3 then we can expect co2 emissions of approximately 201.02 g/km.



Note: Engine sizes are grouped in ranges of 1,000 cm3

In this graphic I binned each size group in ranges of 1000 cm3. I used the ranges to display the mean/average of co2 emissions within those groups. This Histogram tells a clear story, as our engine size increases our average co2 emissions will also increase.

2. Conclusion

The following factors have a large impact on CO2 emissions:

- ✓ Fuel Type
- ✓ Engine Power
- ✓ Transmission Type
- ✓ Engine Size

3. Recommendation

According to the results presented, the following recommendations are important:

To the Government:

- Increase efforts and set targets to inform collaborators and the community about the impact of CO2 emissions.
- Provide tax incentives for vehicles with lower CO2 emissions.
- Offer more options for public transportation.

To the Community:

- Should receive more education about factors that impact on CO2.

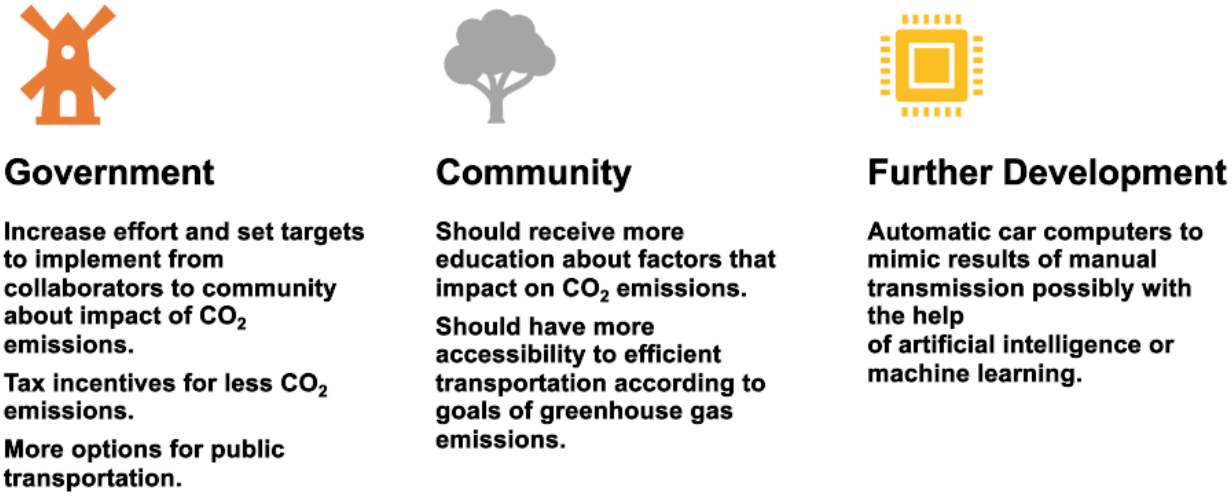
Enhance education about factors that impact CO2 emissions.

- Improve accessibility to efficient transportation in line with greenhouse gas emission goals.

For Further Development:

- Develop automatic car computers that can mimic the results of manual transmissions, possibly with the help of artificial intelligence or machine learning.

Finally, these recommendations are approached for environmental considerations, consumer decision-making, and automative manufacturing.



4. References

- Raman209. (2023, October 23). *Vehicle CO2 emissions prediction*. Kaggle. https://www.kaggle.com/code/raman209/vehicle-co2-emissions-prediction

- Drfrank. (2021, January 24). *CO2 Emission Eda & Visualization & Machine Learnin*.Kaggle. https://www.kaggle.com/code/drfrank/co2-emission-eda-visualization-machine-learnin

- Karankrishna. (2021, February 20). *India Air Quality Analysis*. Kaggle. https://www.kaggle.com/code/karankrishna/india-air-quality-analysis

- Ashikchandradas. (2024, April 20). *CO2 emission Regression Analysis and Visualization*. Kaggle. https://www.kaggle.com/code/ashikchandradas/co2-emission-regression-analysis-and-visualization

- Accelerating Decarbonization of the U.S. Energy System. (2021). United States: National Academies Press.

- U.S. Department of Energy. (2023, January). *The U.S. national blueprint for transportation decarbonization*. https://www.energy.gov/sites/default/files/2023-01/the-us-national-blueprint-for-transportation-decarbonization.pdf

- U.S. Department of Energy. (2023, January). *The U.S. national blueprint for transportation decarbonization: Fact sheet*. https://www.energy.gov/sites/default/files/2023-01/EERE_TranspoDecarb_factsheet-508_0.pdf

- EPA. (2024, June 6). https://www.epa.gov/newsreleases/biden-harris-administration-finalizes-strongest-ever-pollution-standards-cars-position