DAB 103 – Project 1

CO2 EMISSION FROM CARS IN CANADA

Background / Motivation :

Claiming and addressing themselves as a least polluted countries in the world, Canada has maintained it’s position through all means by offering a clean environment to the people. But here’s a data of the vehicles which do cause harm to the ecosystem by emitting CO2 into the atmosphere. Diving deeper into such data could uncover some interesting and unknown facts about the happening.

Problem statement :

From the dataset that we got, the ultimate purpose would be identify which fuel type best suits the environment with respect to its different vehicle characteristics, by equally considering the corelation of some important variables like Fuel Consumption, Engine size and Fuel Type.

Proposal

This project is important as CO2 emission contributes directly towards the global warming which is a very major threat to humanity. This project helps to find out which category of vehicles and fuel type is making the worst impact to the environment and identify the model and make and let these companies work on their research and development team to reduce this emission in their upcoming models.

The main target audience of this project is the government and the environment monitoring agencies/committee which monitors and regulates the CO2 emission across the nation. The main beneficiaries from this project is the humanity where this project tries to give a better future.

Analysis Questions :

- Determine which vehicle type has highest consumption of fuel and largest release of Co2.
- What's identify the optimized engine size and the no.of.cylinders with minimal release of Co2 ?
- Which fuel type contributes to the maximum emission of Co2 in the atmosphere ?
- What are the most influencing features that affect the CO2 emission the most?
- Which company car model has least emission of Co2 ?

Dataset Description :

- This dataset captures the details of how CO2 emissions by a vehicle can vary with the different features. The dataset has been taken from Canada Government official open data website. This is a compiled version. This contains data over a period of 7 years. There are total 7385 rows and 12 columns. There are few abbreviations that has been used to describe the features.
- Below are few expansions for the variables that are used in this dataset

Model :

- 4WD/4X4 = Four-wheel drive
- AWD = All-wheel drive
- FFV = Flexible-fuel vehicle
- SWB = Short wheelbase
- LWB = Long wheelbase
- EWB = Extended wheelbase

Transmission :

A = Automatic

AM = Automated manual

AS = Automatic with select shift

AV = Continuously variable

M = Manual

3 - 10 = Number of gears

Dataset Description :

Fuel Type :

- X = Four-wheel drive
- Z = All-wheel drive
- D = Flexible-fuel vehicle
- E = Short wheelbase
- N = Long wheelbase

Fuel Consumption :

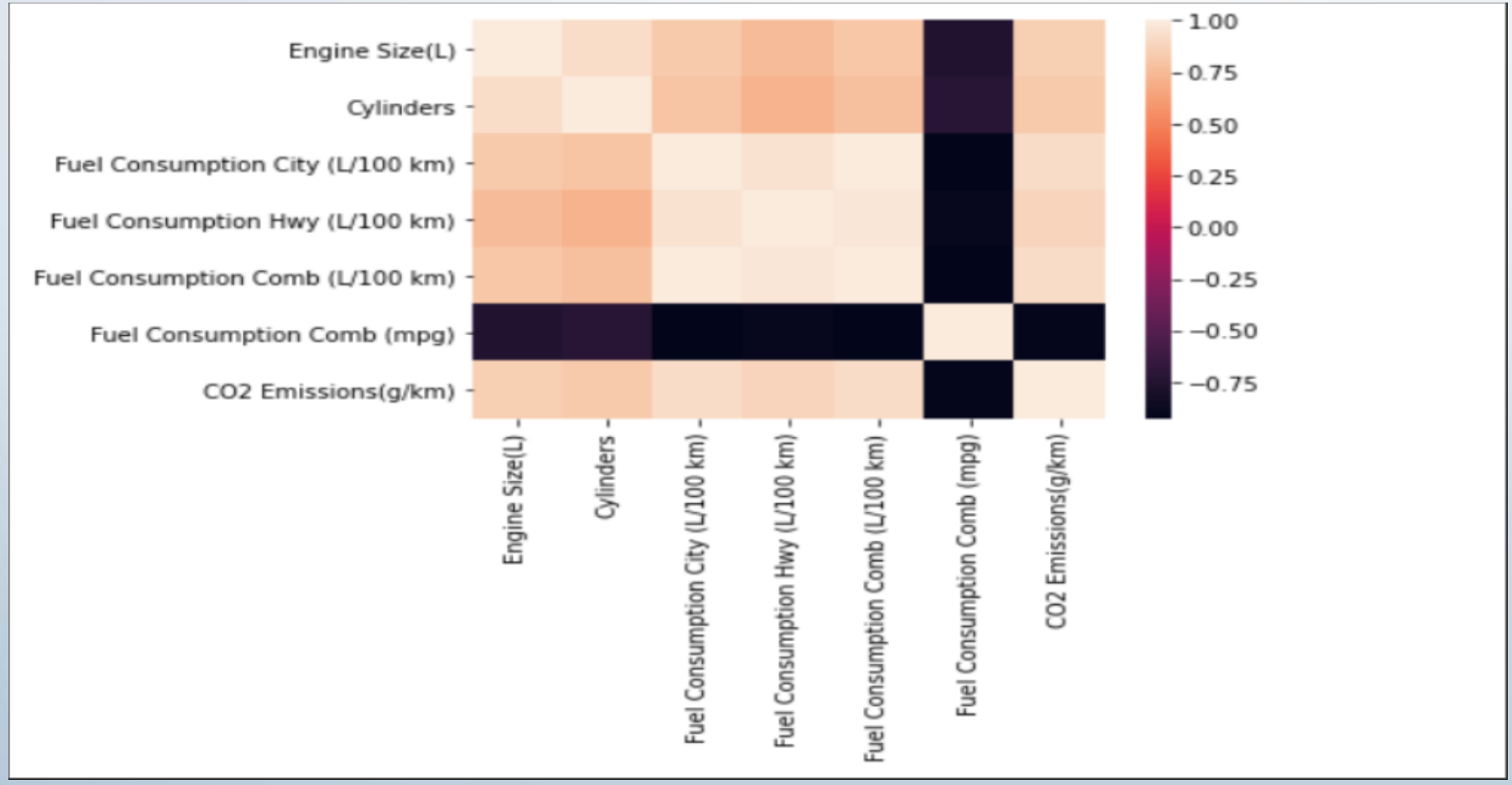
City and highway fuel consumption ratings are shown in litres per 100 kilometres (L/100 km) - the combined rating (55% city, 45% hwy) is shown in L/100 km and in miles per gallon (mpg)

CO2 Emissions :

The tailpipe emissions of carbon dioxide (in grams per kilometre) for combined city and highway driving

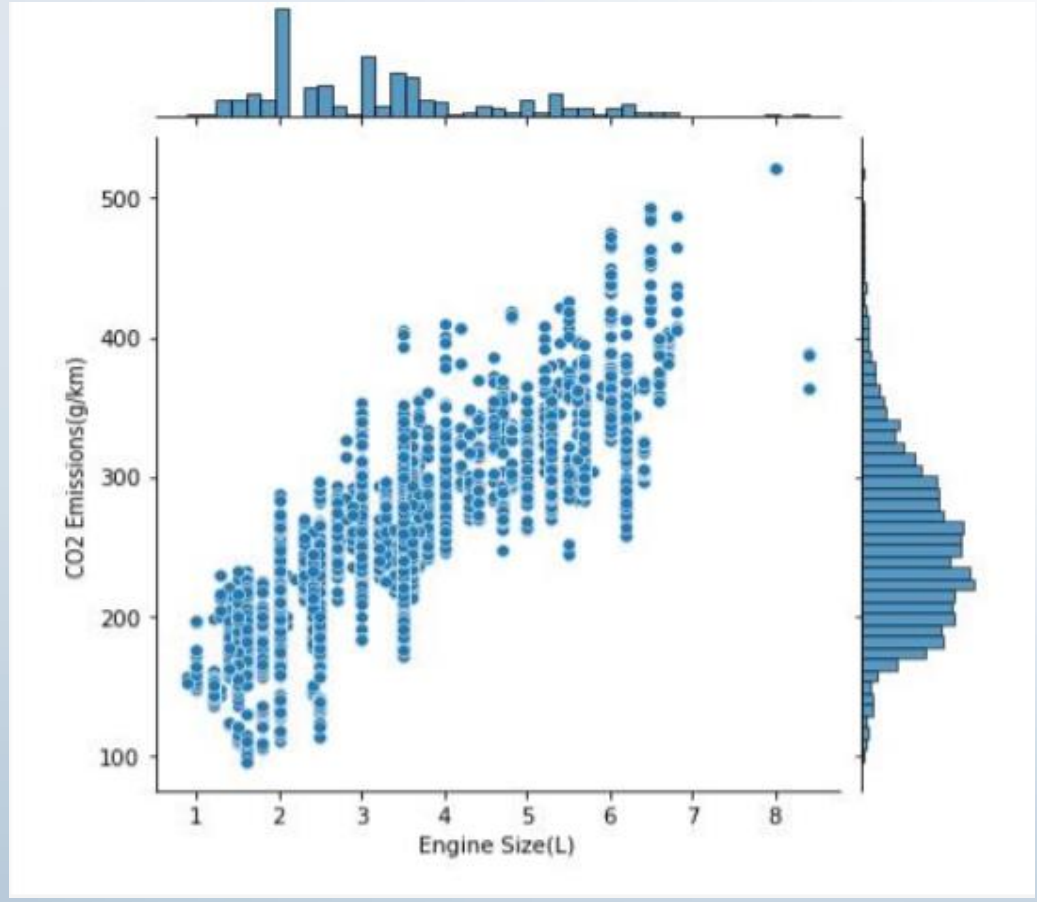
Basic Stats of the variables

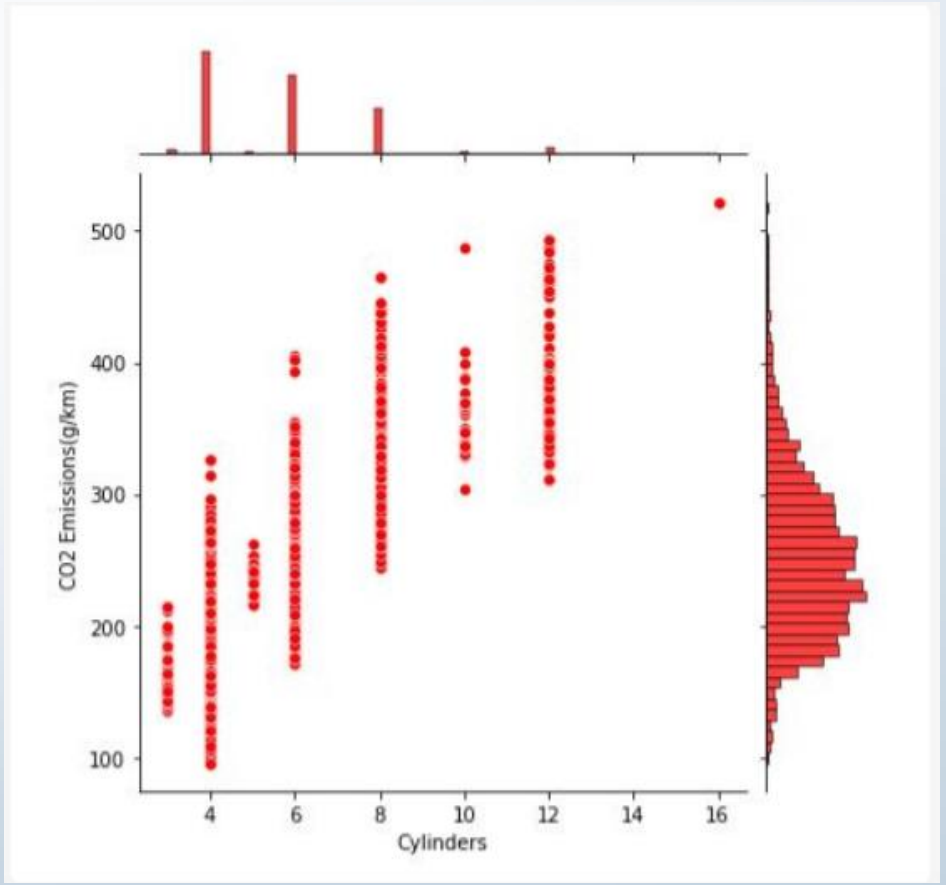
Heat Map (Indicating the Co-relation of all the available variables) :



Relationship between the Engine Size , Cylinders with CO2 Emission :

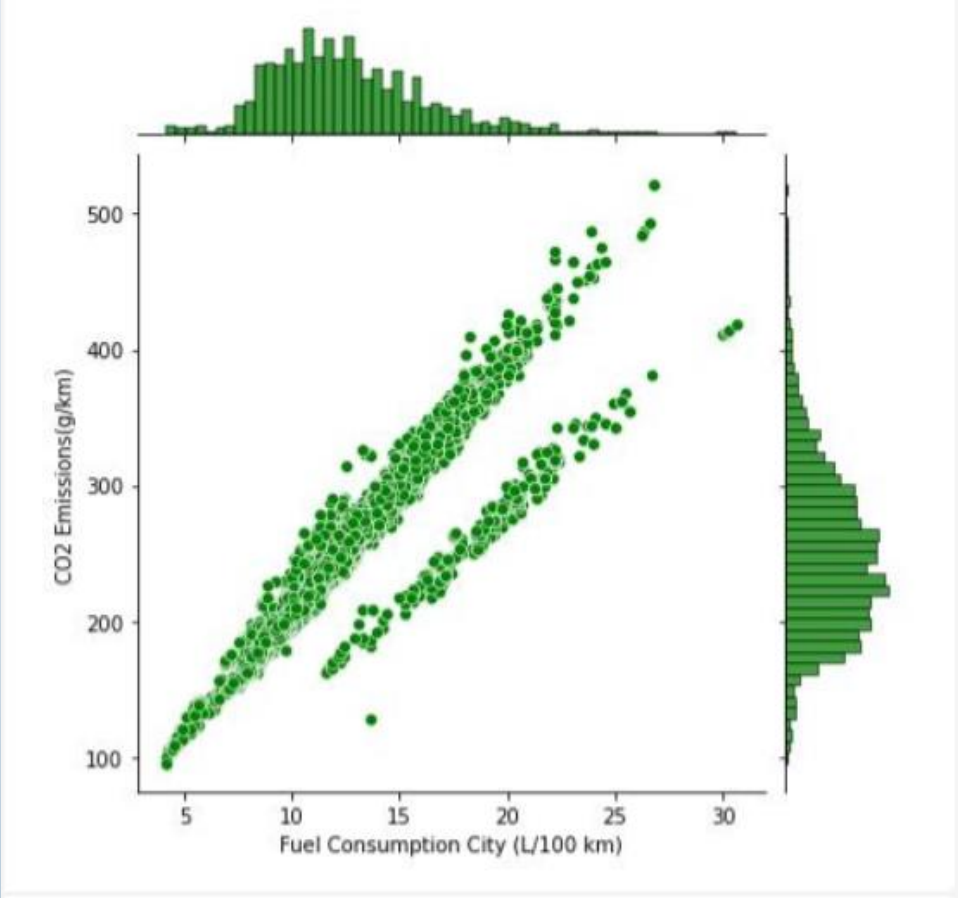
- Positive Co-relation

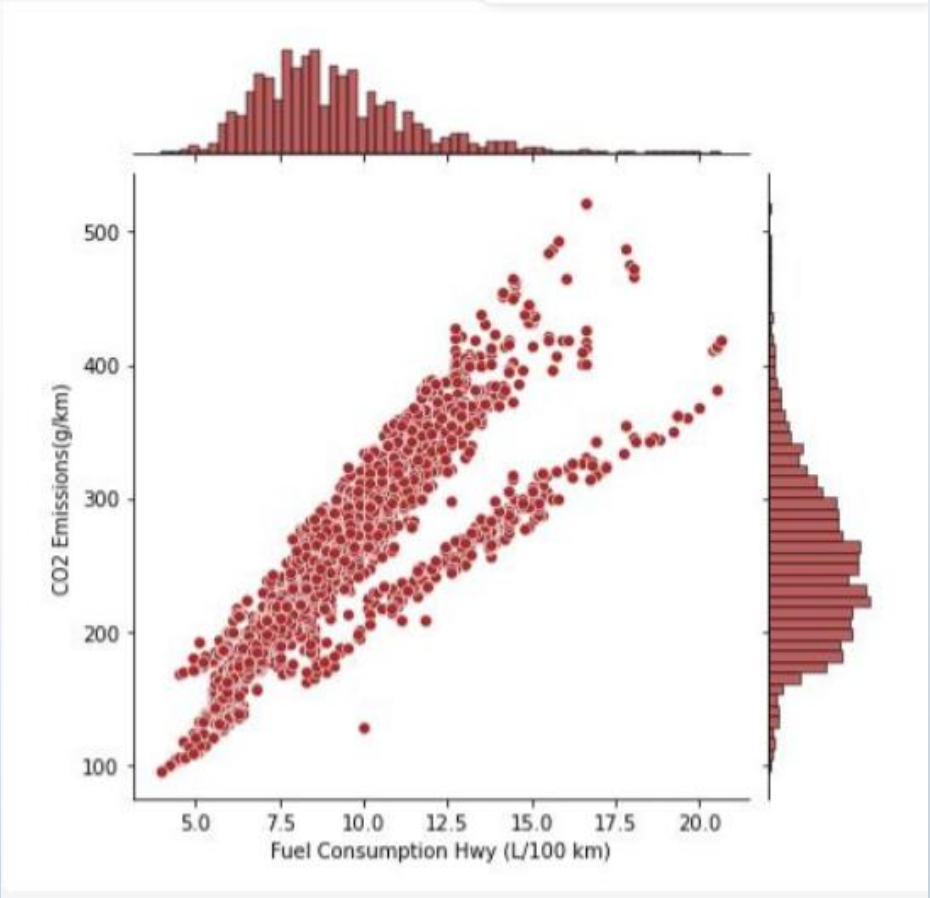




Co-relation of Engine Size , Cylinders with CO2 Emission :

- Positive Corelation





Exploratory Data Analytics (EDA) :

Enclosing the GGCOLAB link , where we have made some Data transformations , preliminary visualizations and correlations that well describes the data.

https://colab.research.google.com/drive/1xxI-w8BU_2agW_st0bK0iV7vmepVv0RC?usp=sharing

References :

- The data has been taken and compiled from the below Canada Government official link https://open.canada.ca/data/en/dataset/98f1a129-f628-4ce4-b24d-6f16bf24dd64#wb-auto-6

Thank You !!

DAB103 - Project

CO2 EMISSION FROM CARS IN CANADA

Submission 2

Addressing the feedback from previous submission

Problem Statement

- To determine which fuel type and vehicle category emits the least CO2.
- To find which vehicle characteristics is less harmful to the environment .

Proposal

Identifying which category of vehicles are making the worst impact to the environment by majorly targeting the manufacturers and the environment monitoring agencies who can regulate this mass emission into our atmosphere.

EDA Visualizations

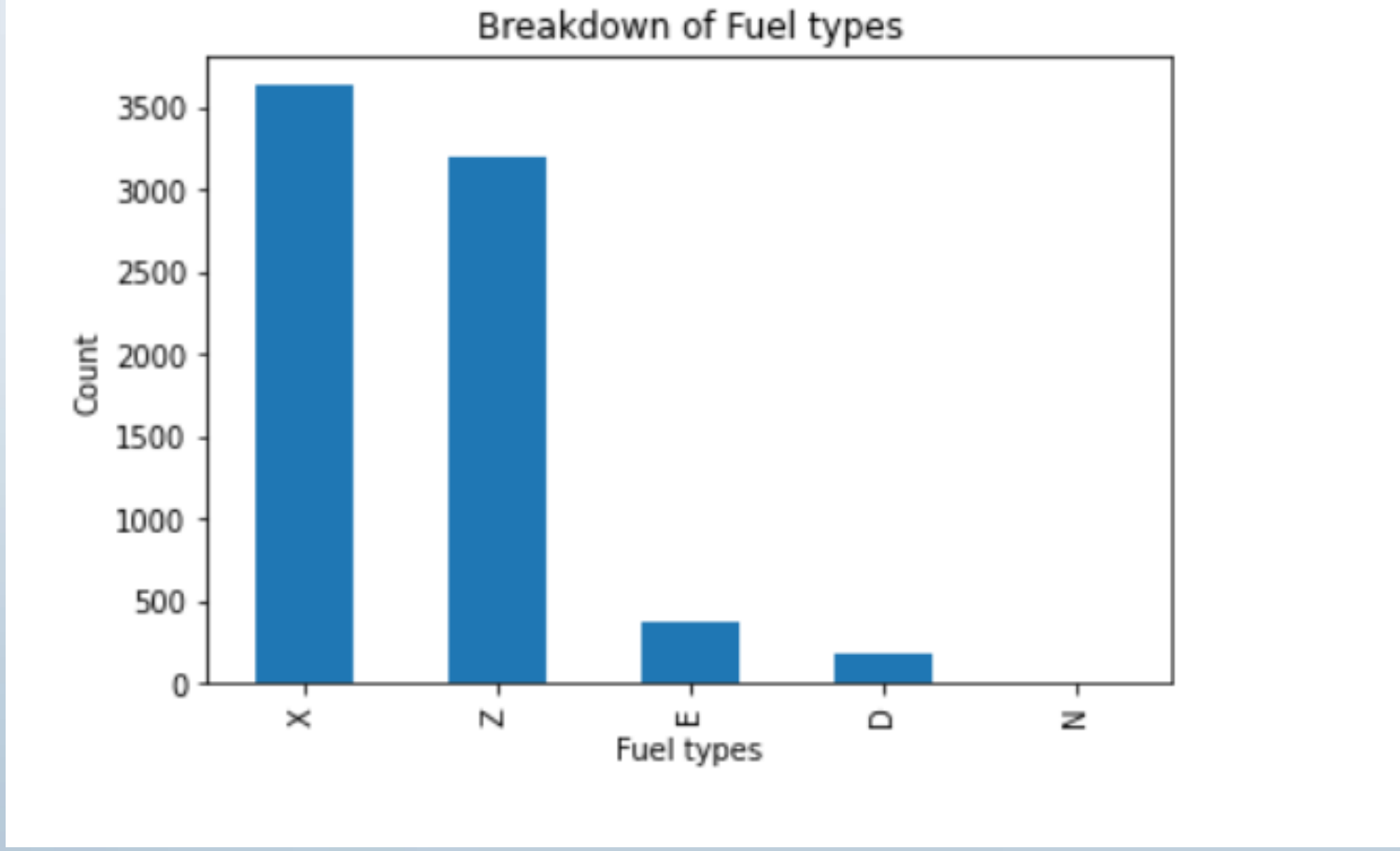
Descriptive Statistics :

Using describe function , the basic stats of the all the variable in the dataset is uncovered.

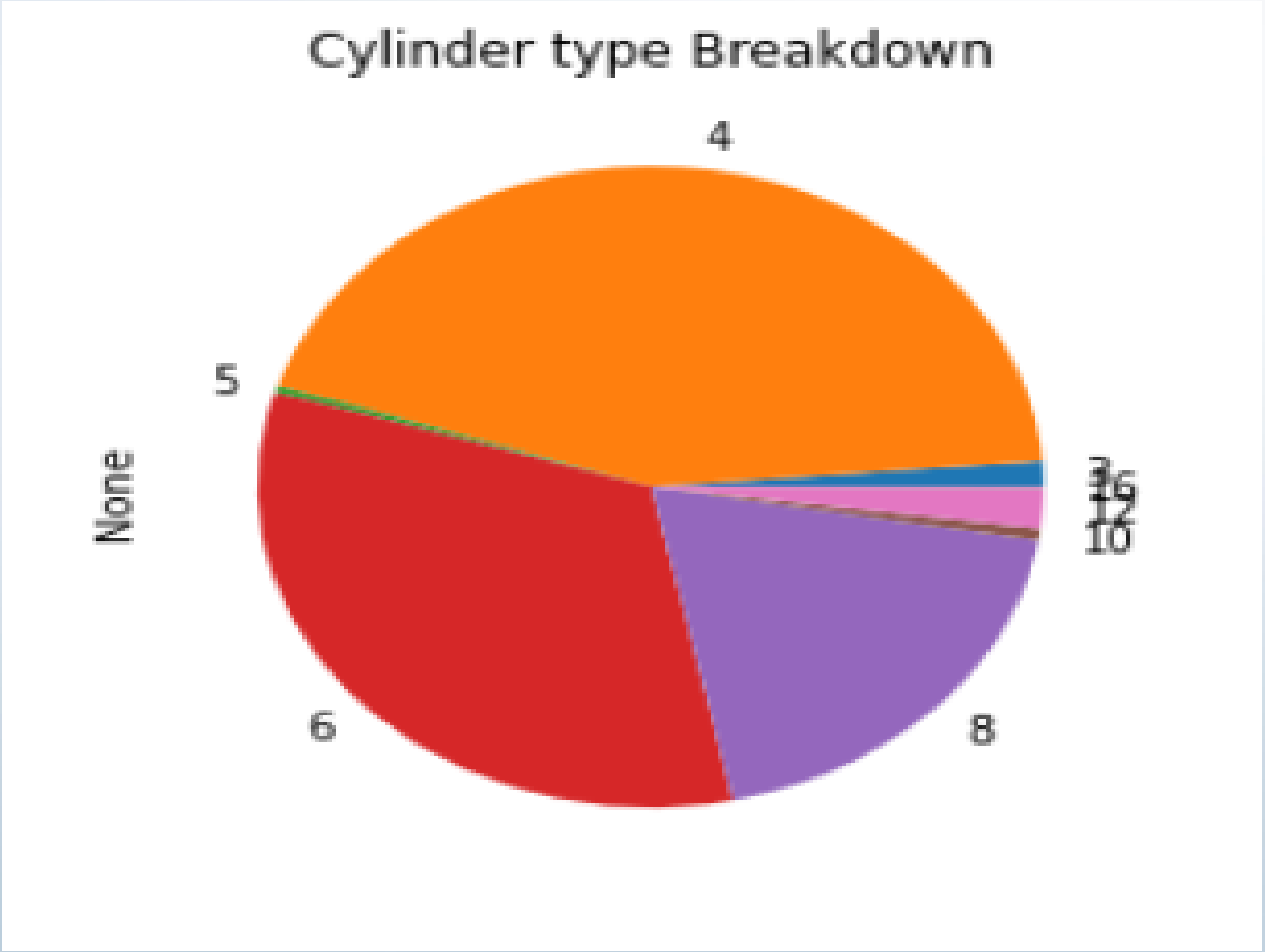
	Engine Size(L)	Cylinders	Fuel Consumption City (L/100 km)	Fuel Consumption Hwy (L/100 km)	Fuel Consumption Comb (L/100 km)	Fuel Consumption Comb (mpg)	CO2 Emissions(g/km)
count	7385.000000	7385.000000	7385.000000	7385.000000	7385.000000	7385.000000	7385.000000
mean	3.160068	5.615030	12.556534	9.041706	10.975071	27.481652	250.584699
std	1.354170	1.828307	3.500274	2.224456	2.892506	7.231879	58.512679
min	0.900000	3.000000	4.200000	4.000000	4.100000	11.000000	96.000000
25%	2.000000	4.000000	10.100000	7.500000	8.900000	22.000000	208.000000
50%	3.000000	6.000000	12.100000	8.700000	10.600000	27.000000	246.000000
75%	3.700000	6.000000	14.600000	10.200000	12.600000	32.000000	288.000000
max	8.400000	16.000000	30.600000	20.600000	26.100000	69.000000	522.000000

- To understand the graph of each variable below are the Visualizations :

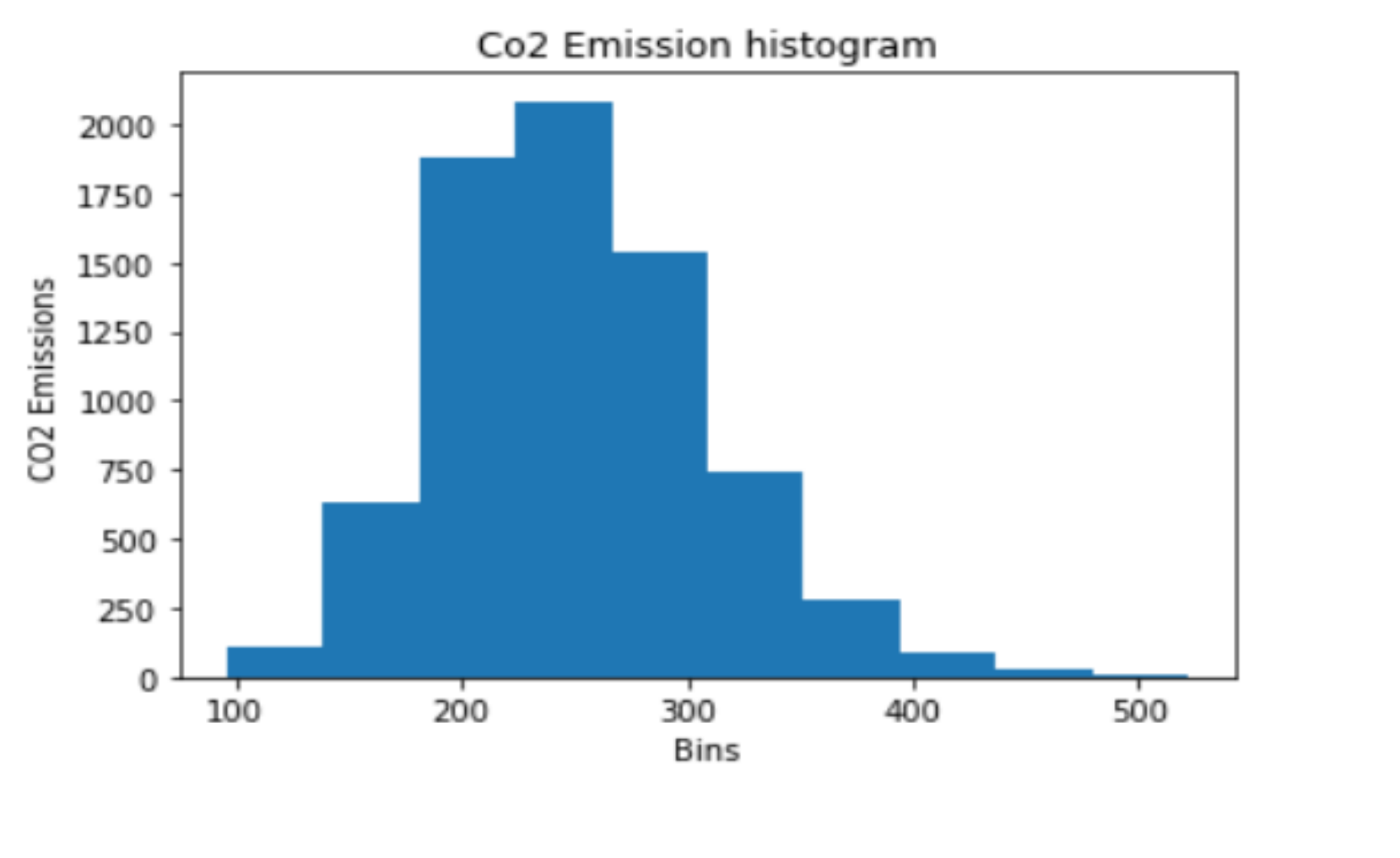
  Fuel Types :



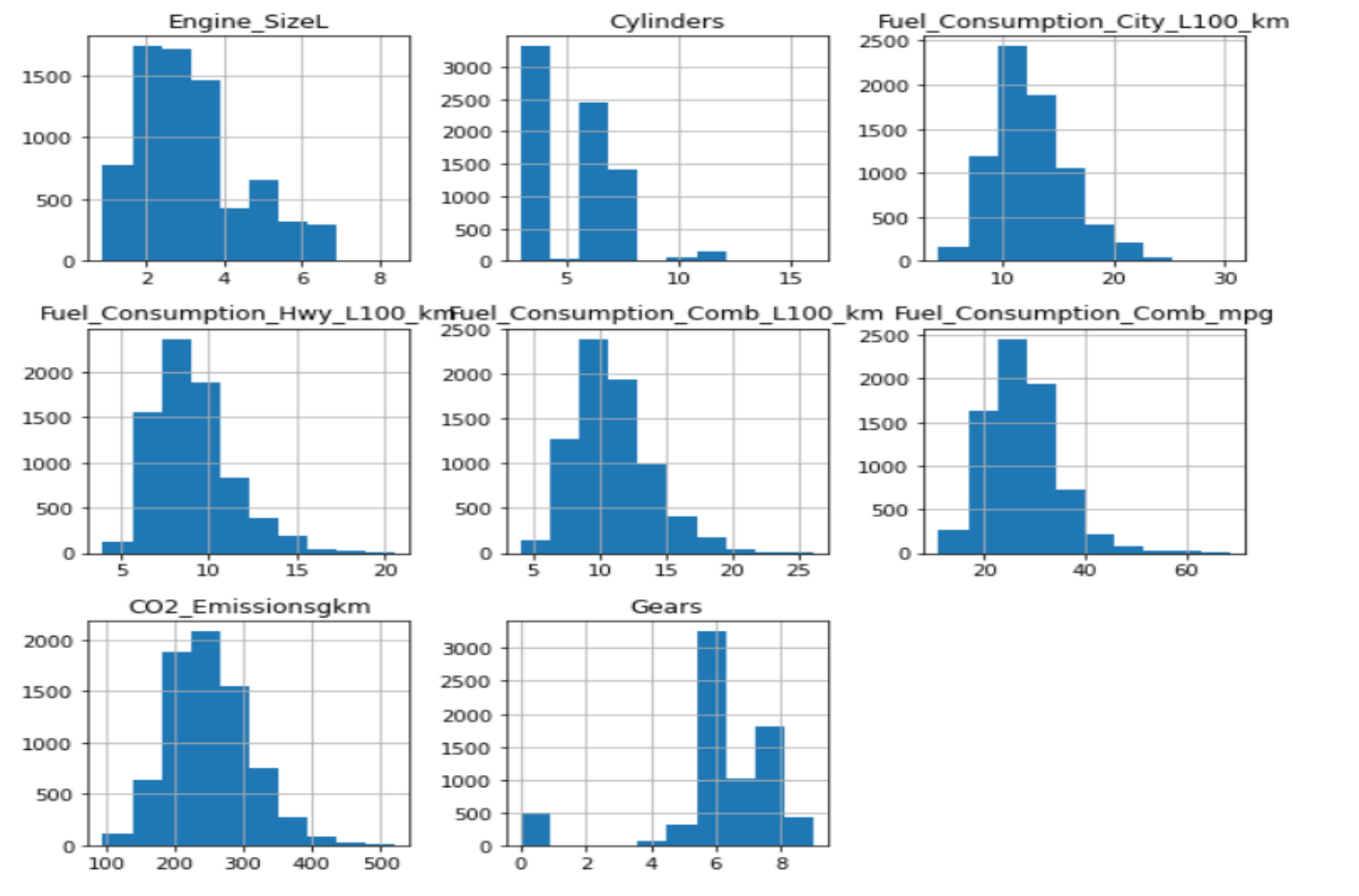
Number of Cylinders :



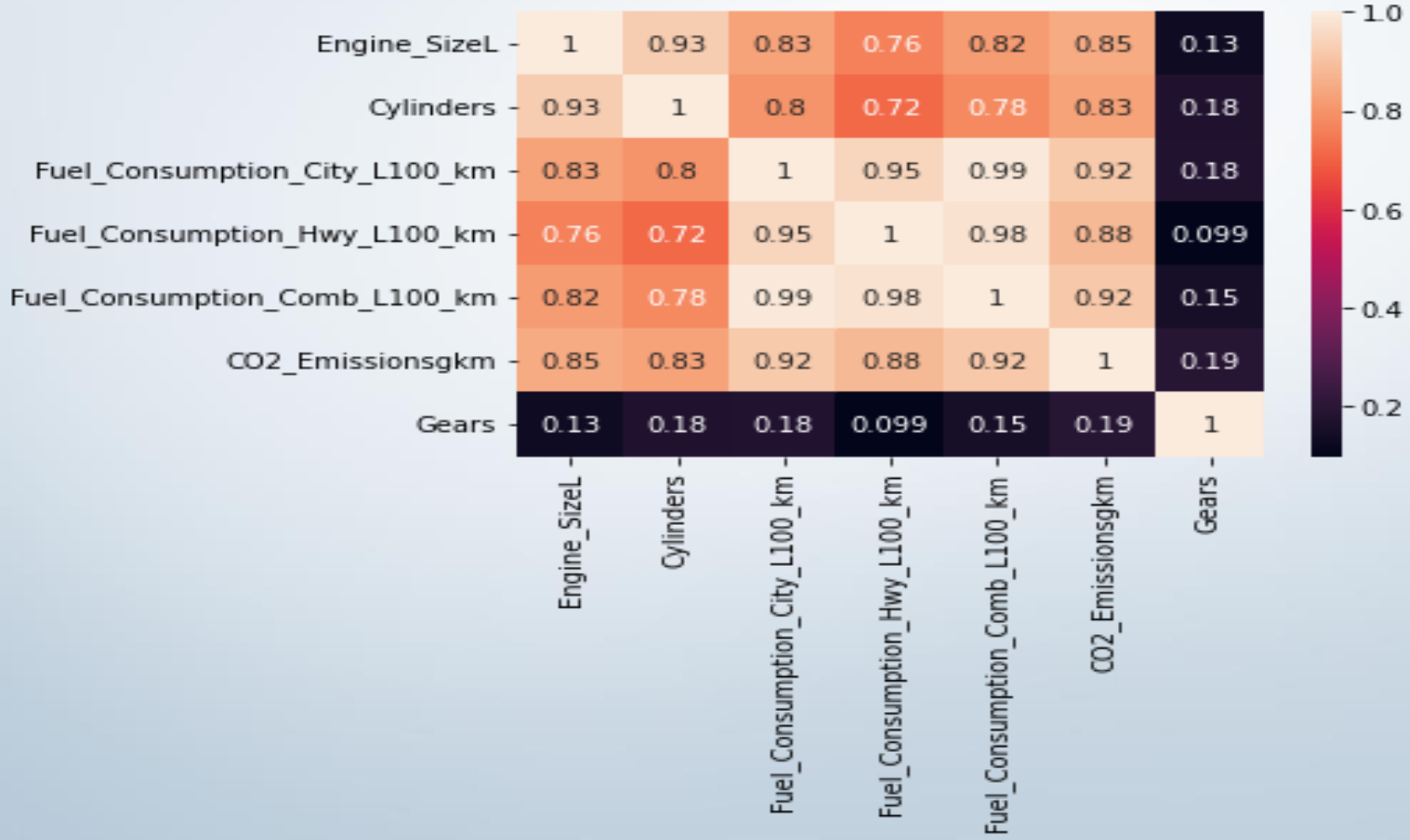
CO2 Emission:



Numerical Variables visualized through Histograms [ref: 1]



Heatmap Showing Co-relation of the dependent variables



Data Cleaning & Transformation

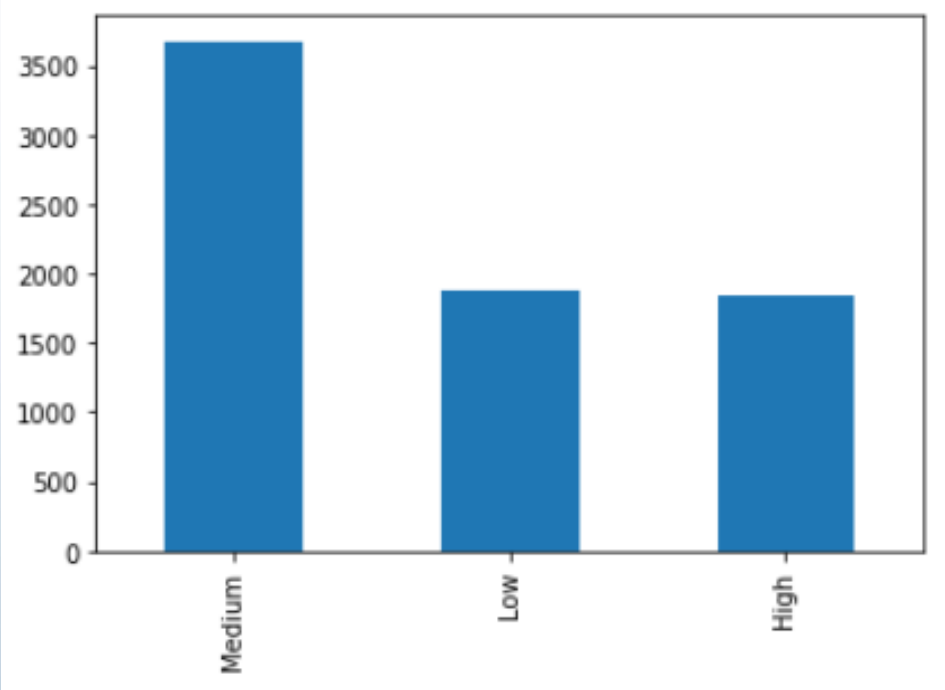
The data we have has only one observation for the fuel type "N". It is tough to conclude the analysis with a single observation, and also like a potential outliner, hence removing this from the data set.

Chevrolet Mid-Size has the one and only Natural Gas (N) Fuel Type.

	Make	Model	Vehicle_Class	Engine_SizeL	Cylinders	Transmission	Fuel_Type	Fuel_Consumption_City_L100_km	Fuel_Consumption_Hwy_L100_km	Fuel_Consumpt
2439	CHEVROLET	IMPALA DUAL FUEL	MID-SIZE	3.6	6	AS6	N	15.2	9.5	

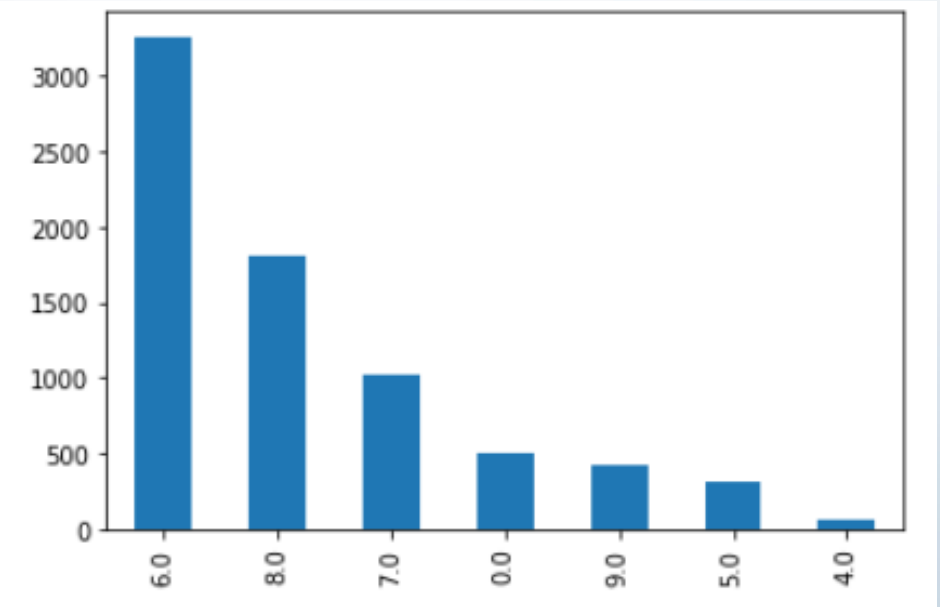
Segregating the Co2 emissions by labeling them with various levels High, medium, low based on their quartile range, which results in adding a new column with the name “CO2 Label” having values with respect to CO2 emission.

```
Medium    3674
Low       1870
High      1840
Name: Co2_label, dtype: int64
```



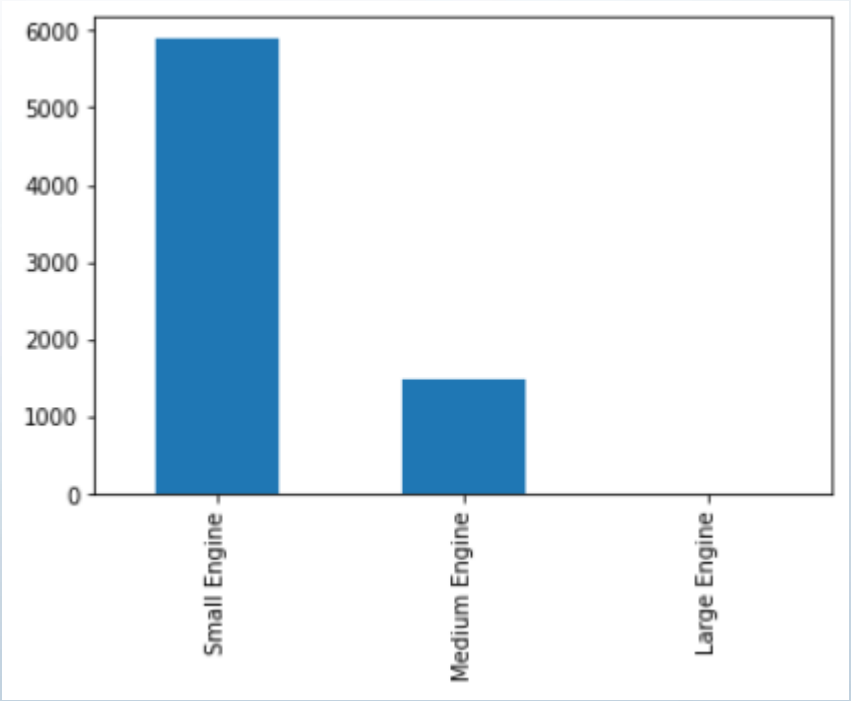
Transforming the “Transmission” variable to give the actual Gear count for each vehicles and replacing ‘V’ (Automatic) with ‘0’.

6	3258
8	1802
7	1026
0	505
9	419
5	307
4	67



From the data set separating the vehicle into three types of engine category namely small, medium and large based on the EngineSizeL variable which denotes the engine size in litres capacity.

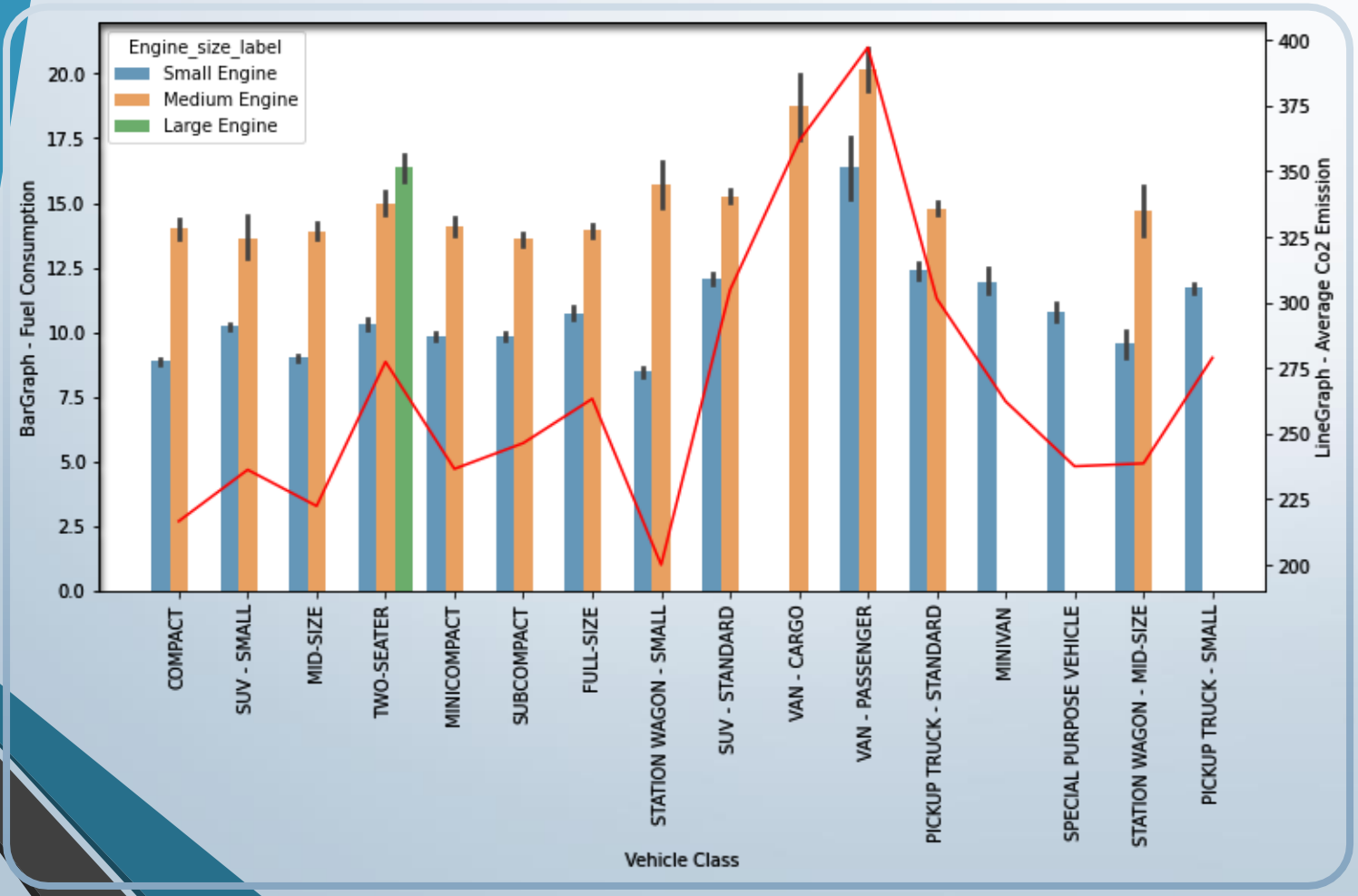
```
Small Engine     5889
Medium Engine    1490
Large Engine        5
Name: Engine_size_label, dtype: int64
```



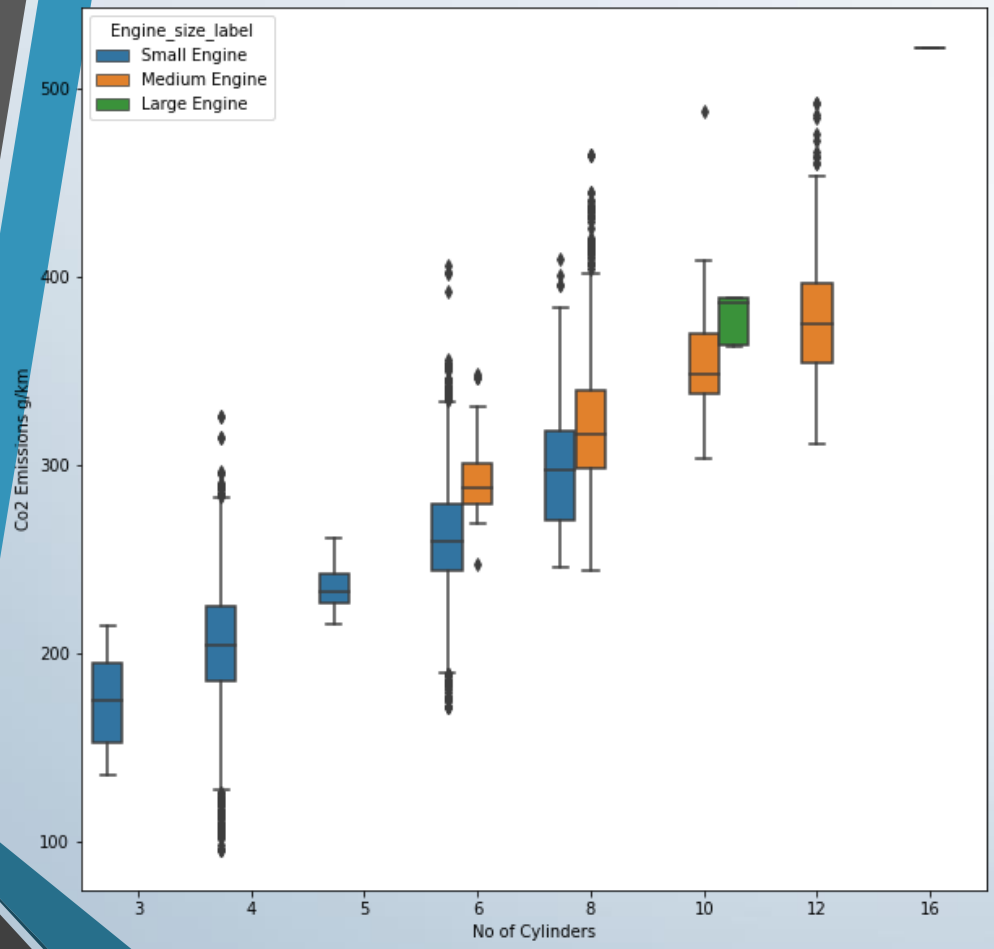
Data Analysis and Visualization

Data Analysis

1. Determining the vehicle type that has highest consumption of fuel and largest release of Co2. [ref: 2]



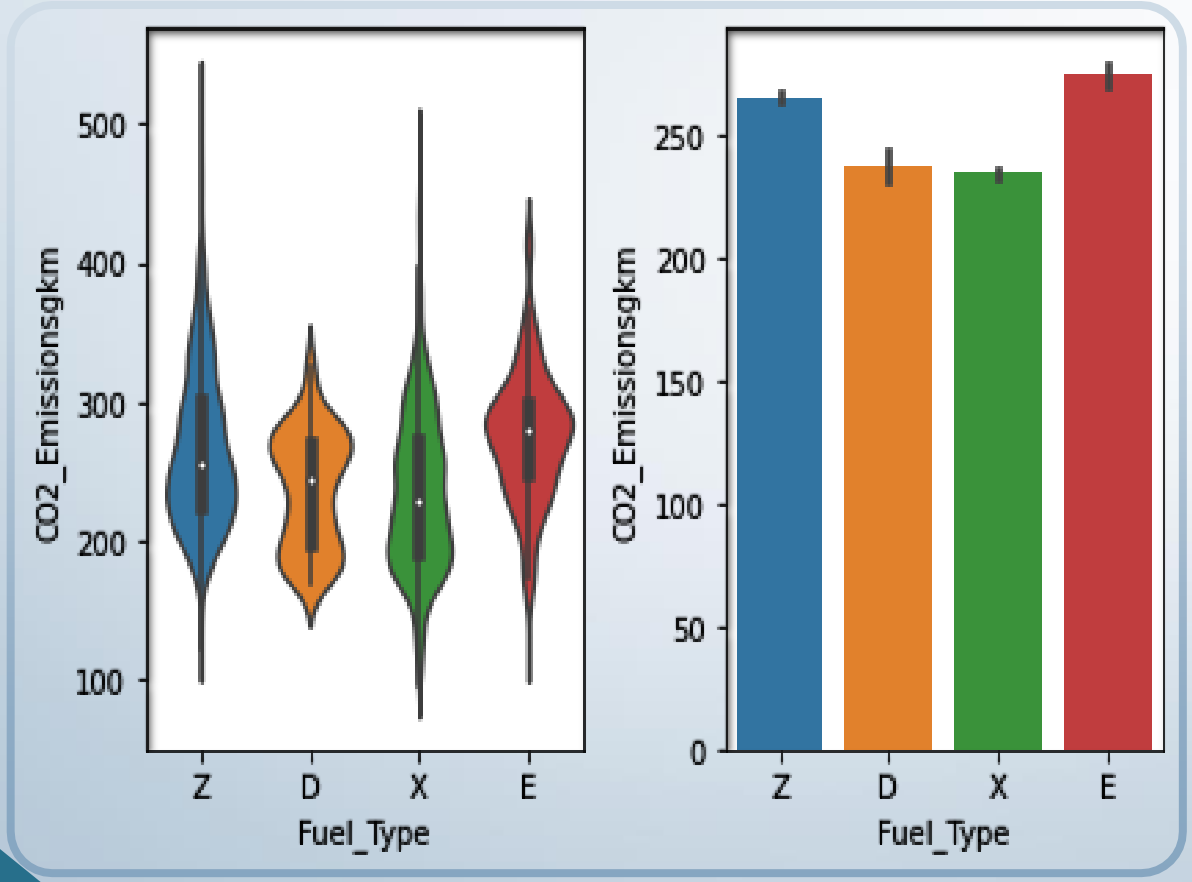
From the graph we could see that the largest CO2 emission was produced by Passenger Vans with engine size of medium.



2. Identifying the optimized engine size and the number of cylinders with minimal release of Co2.

Engine Size seems to have strong positive Co-relation with the emission of C02. So small engine vehicles tend to be more eco-friendly than the larger and medium ones.

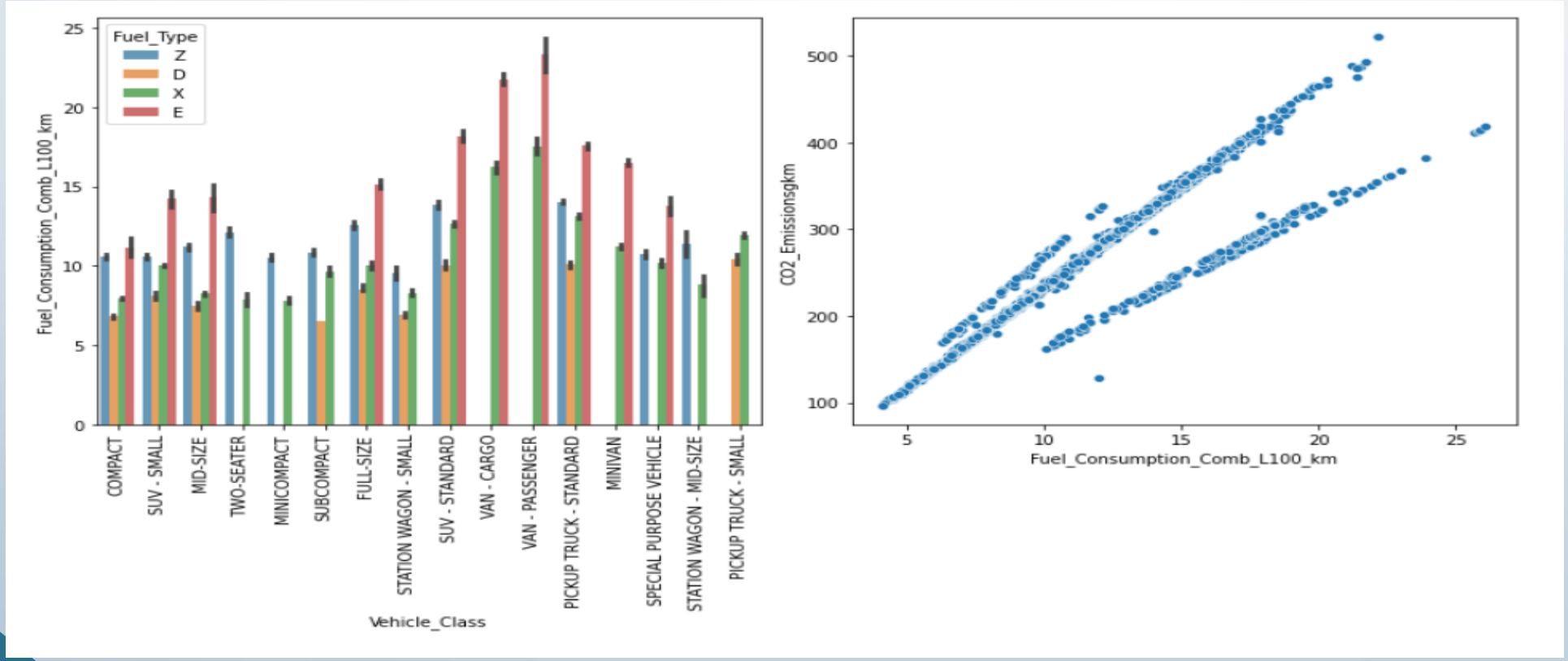
3. Which fuel type contributes to the maximum emission of Co2 in the atmosphere ? [ref: 4]



- Almost all the fuel type emits approximately same amount of emission of more than 200, however on comparison we can see that among the all, fuel type Ethanol (E) and Premium Gasoline (Z) has most emission.

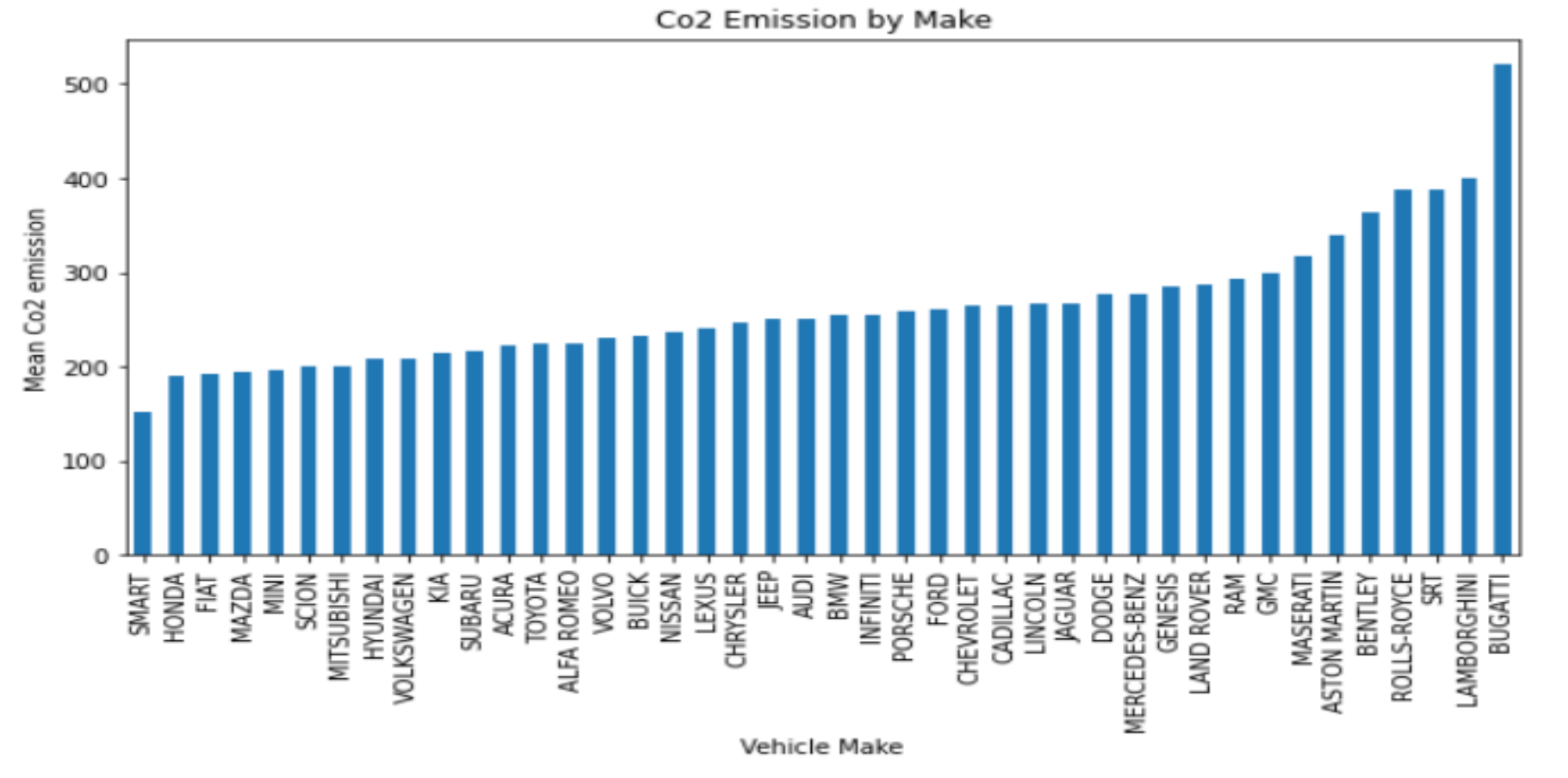
4. What are the most influencing features that affect the CO2 emission the most ? [ref: 5]

From the graph all the variables Mileage , Fuel Consumption and Engine Size are highly influential variables that causes the large emission of CO2 .

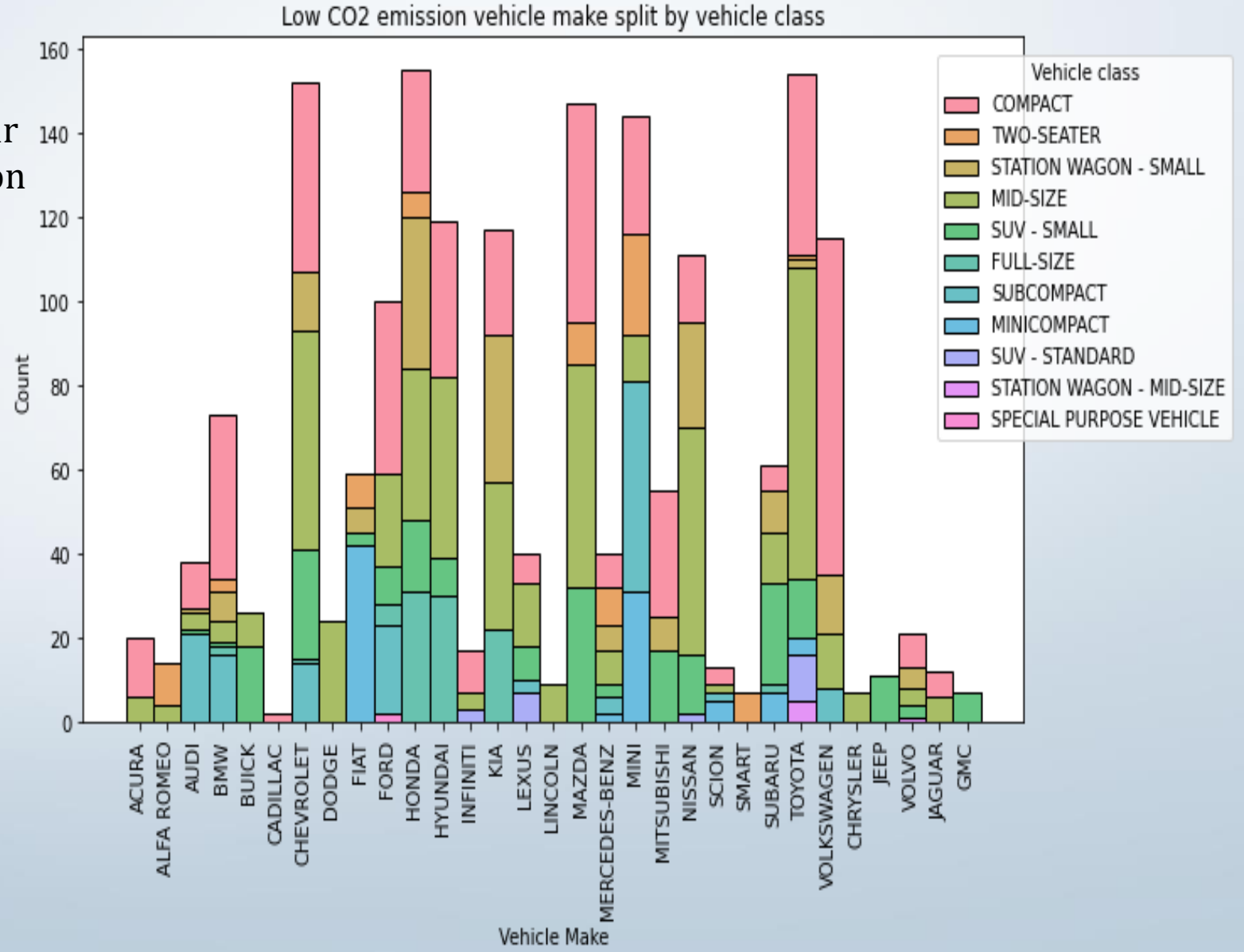


5. Which company car model has least emission of Co2 ? [ref: 6]

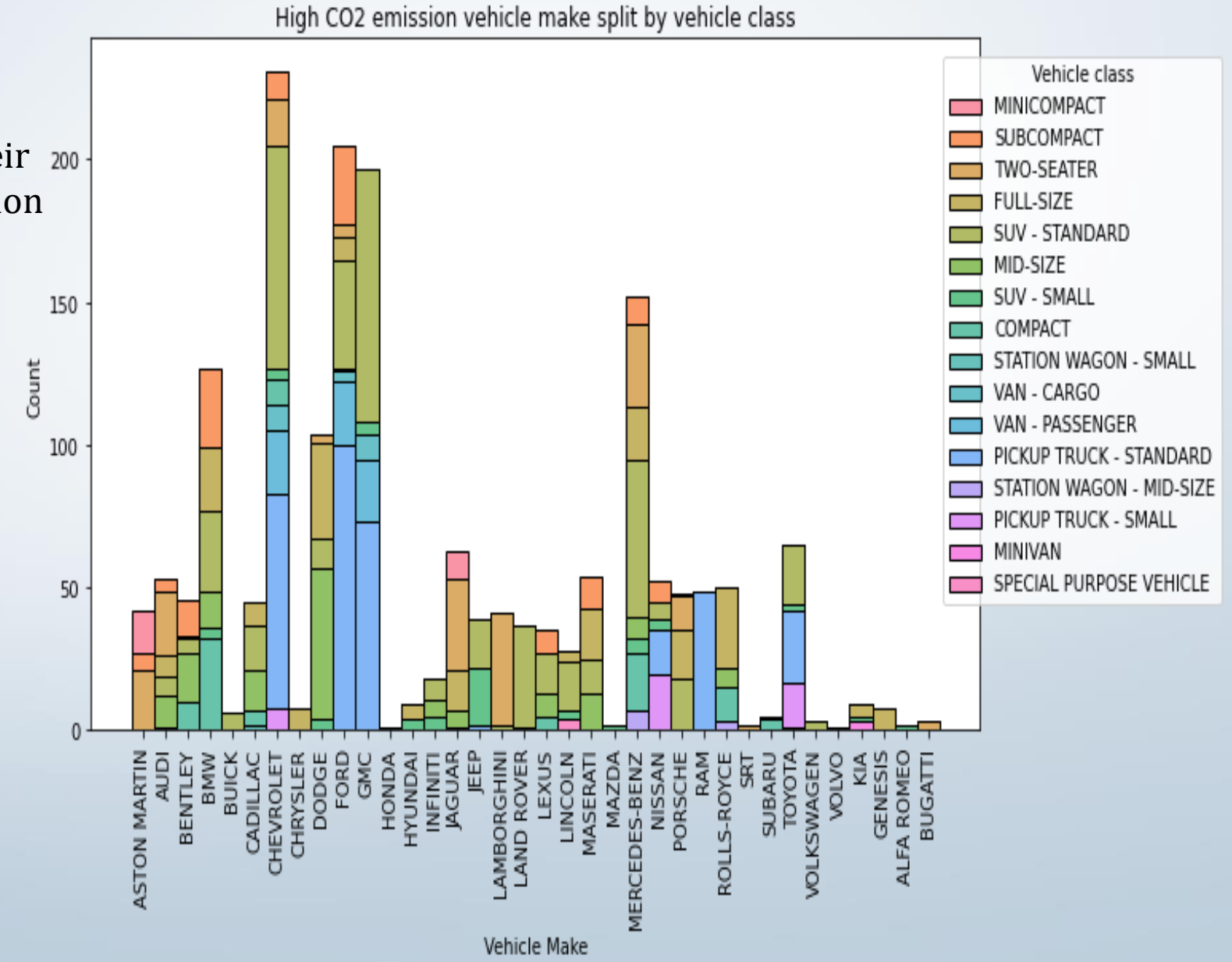
Although the manufacturing rate of ‘Smart’ model is comparatively lower than their competitors, their importance towards the environment is evident , showing very minimal release of CO2 in their vehicles.



Split up of Vehicle make with their vehicle class for Low Co2 Emission category. [ref: 3]



Split up of Vehicle make with their vehicle class for High Co2 Emission category.



Split up of Vehicle make with their vehicle class for Medium Co2 Emission category.

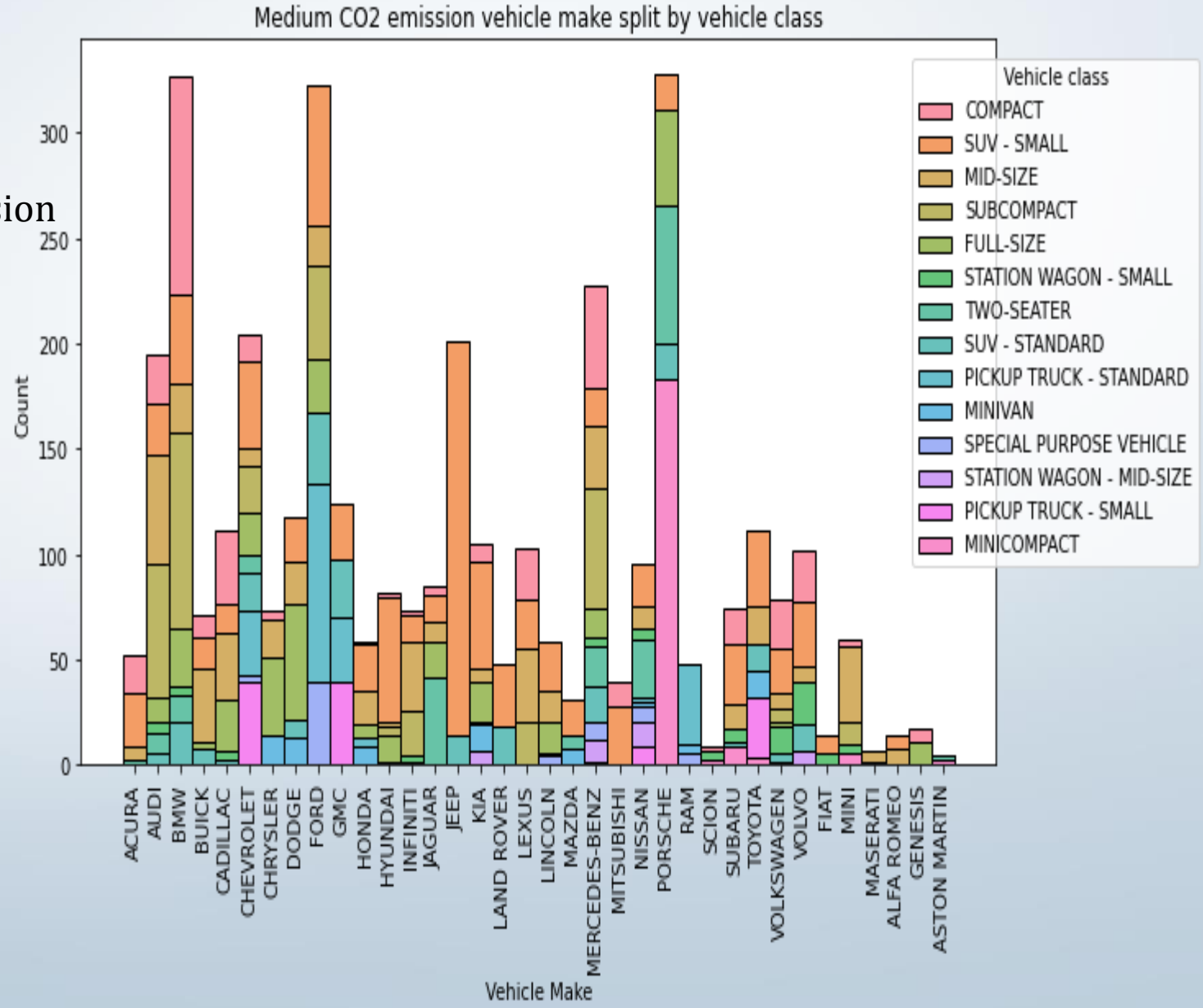
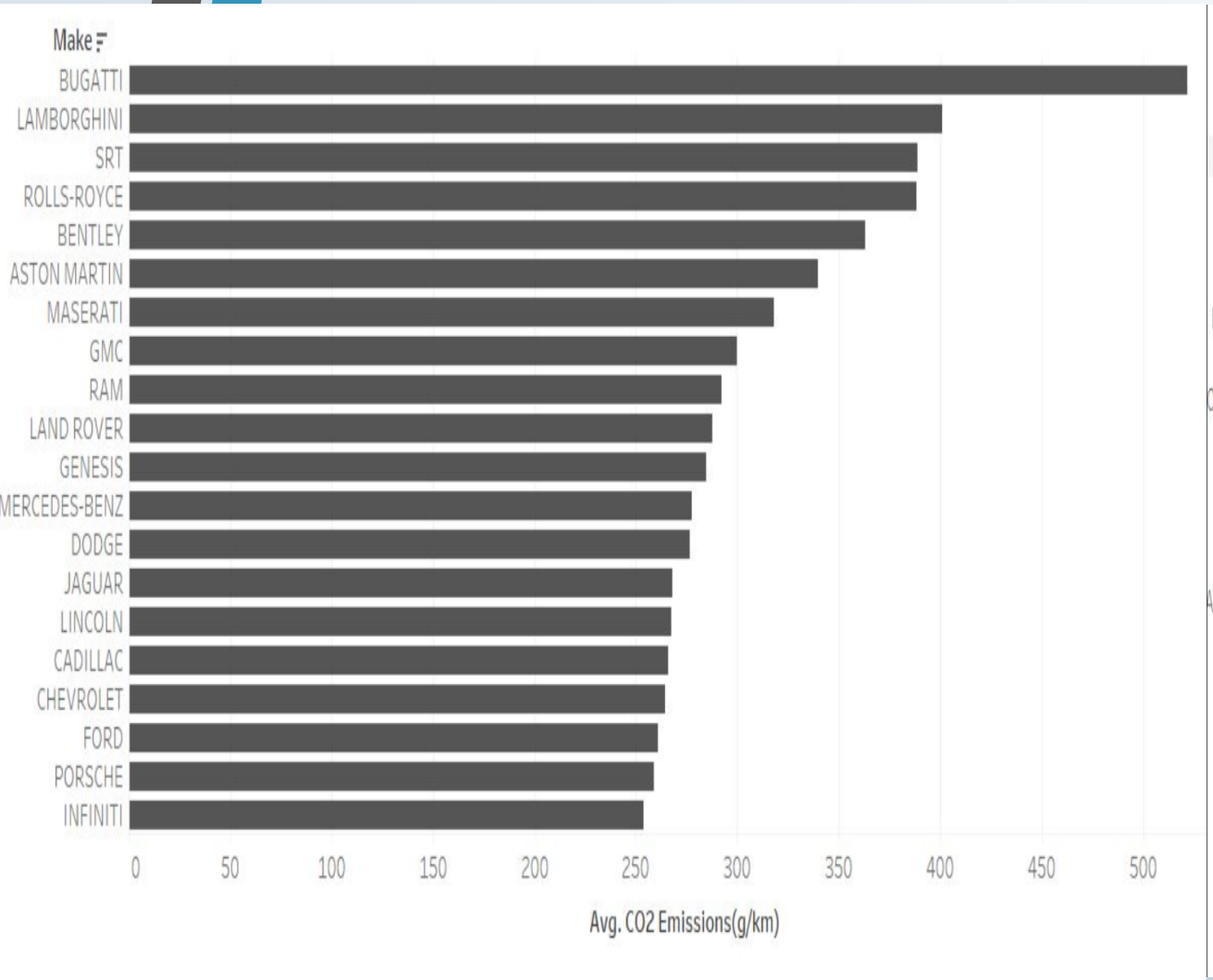


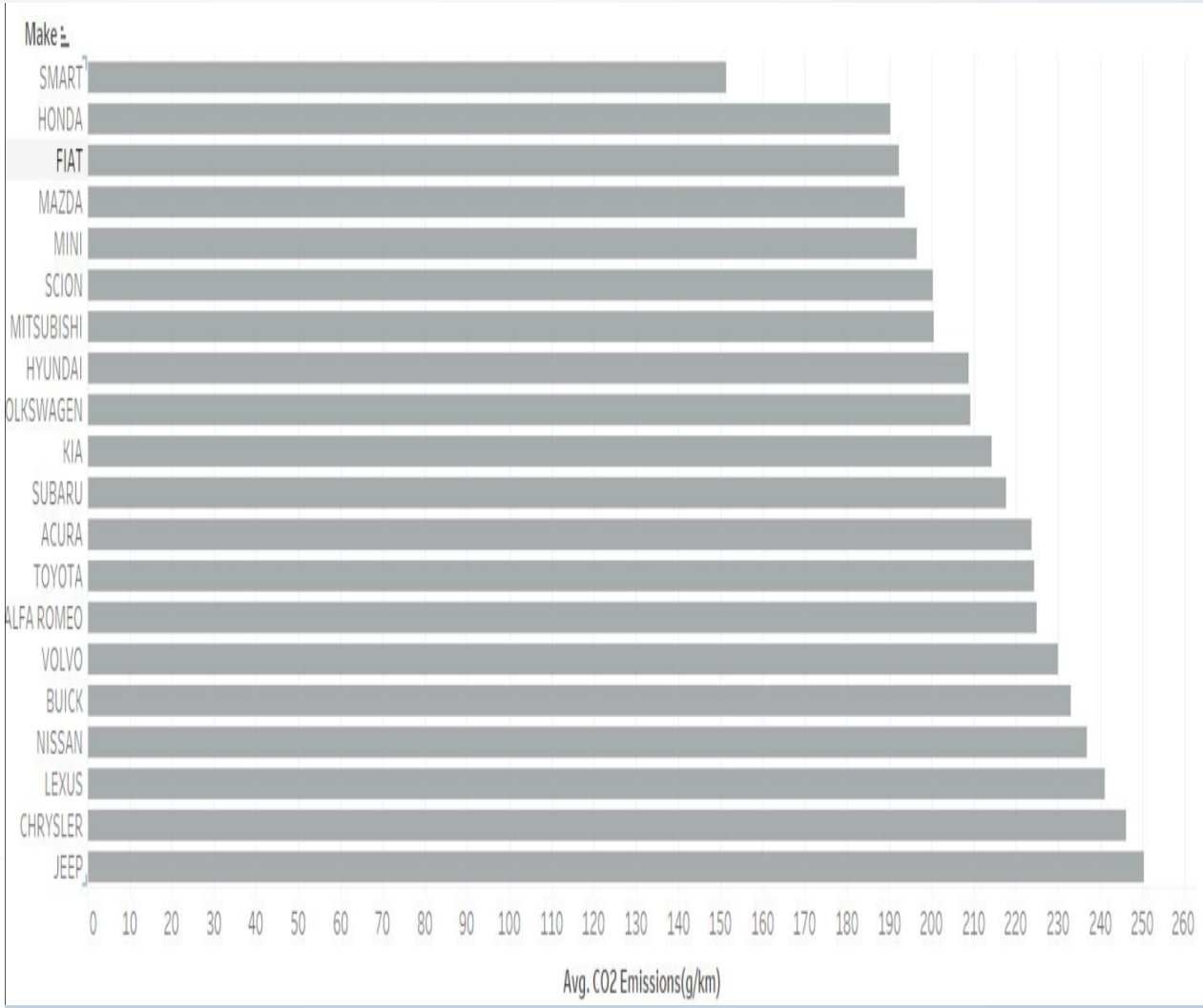
Tableau Dashboard

Avg. CO2 Emission by Vehicle Make

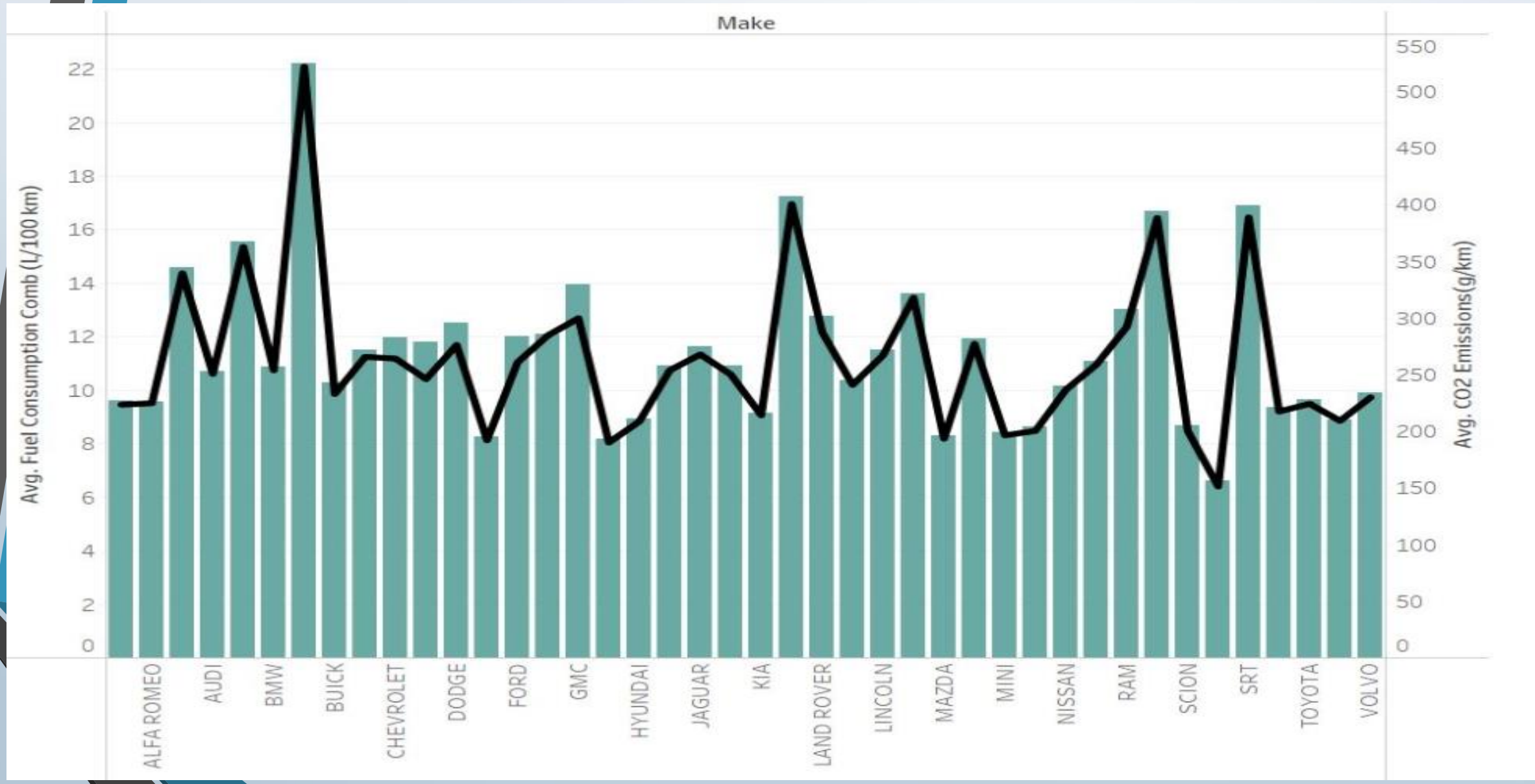
Maximum Emission



Minimum Emission

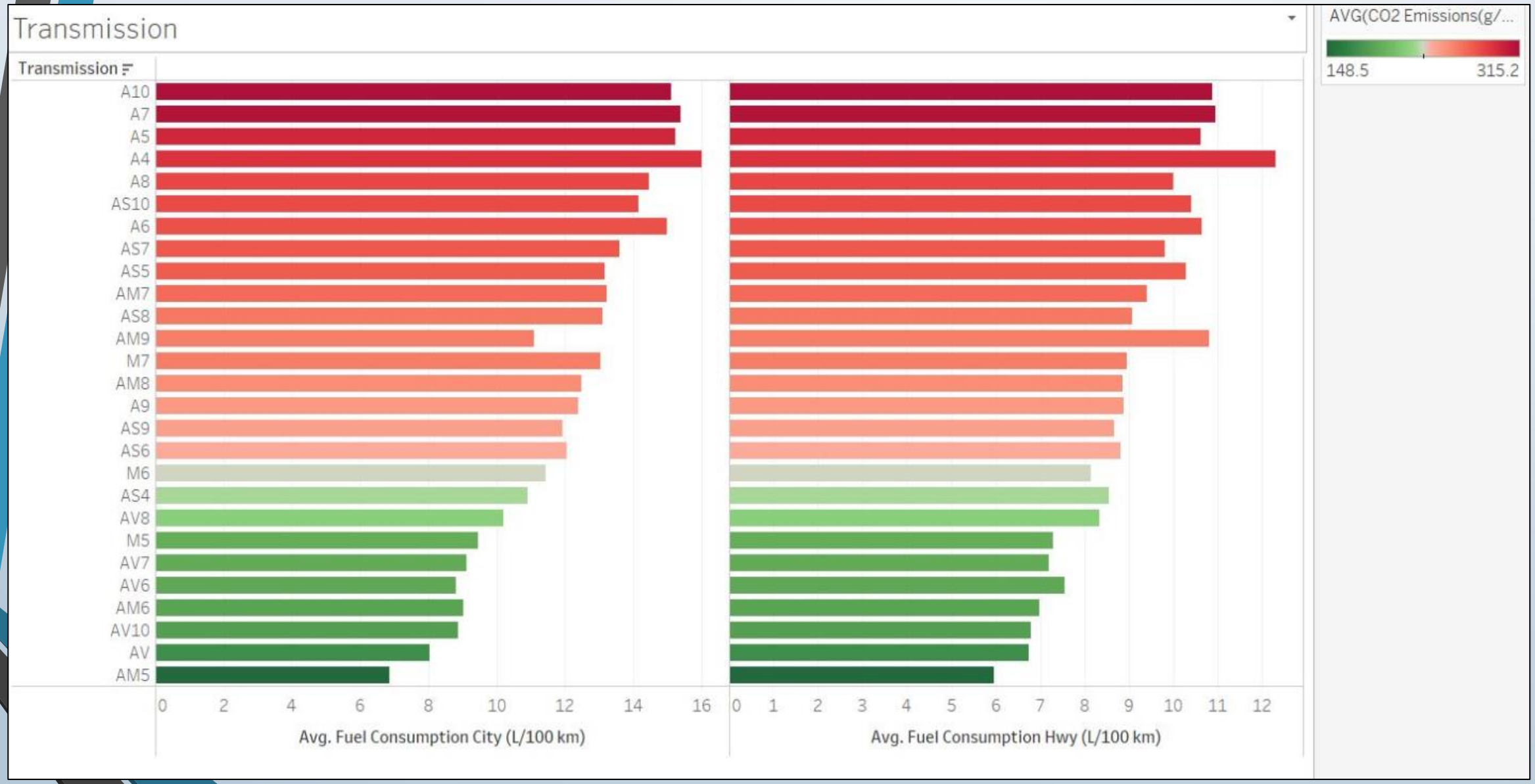


Relationship of Vehicle Make between Avg. Fuel Consumption and Avg.CO2 Emission



Vehicle class by Avg.CO2 Emission by Fuel Type

Avg.Fuel Consumption in City and Highway by Avg.Co2 Emission with respect to Transmission.



Conclusion

From all the graphs and calculations, we could arrive at the below conclusion.

- The fuel consumption in City roads are higher when compared with highway and thus contributes to major Co2 emission correlation.
- Engine size and the number of cylinders has the second highest impact on CO2 emission as these variables are directly correlated to each other.
- The company SMART produces vehicles that has least impact on environment.
- Fuel type D and X are the best fuels than other with respect to CO2 emission.
- Standard Two Seaters and SUV class vehicles generally come less influential factors, thus resulting in minimal emission of C02.

Recommendations

Overall, smaller engine size with lower number of cylinders has the least impact the emission. Hence manufacturers must take this into consideration to bring in more on technology that can create a vehicles having an optimized engine size and cylinders wherever possible and still better serve its purpose.

In addition to that, heavy vehicle manufacturers can spend their research on creating engines which runs on D and X type so that their Co2 emission can be reduced to a certain level.

Code repository

- Data Product Git hub link: https://github.com/kailash1070/dab103/blob/main/DAB103_project_1.ipynb
- Video Presentation – OneDrive Link:
- Tableau Public link:

References

1. https://www.geeksforgeeks.org/python-pandas-dataframe-select_dtypes/
2. https://stackoverflow.com/questions/37930693/how-can-i-make-a-barplot-and-a-lineplot-in-the-same-seaborn-plot-with-different
3. https://stackoverflow.com/questions/27019079/move-seaborn-plot-legend-to-a-different-position
4. https://stackoverflow.com/questions/41384040/subplot-for-seaborn-boxplot https://www.kite.com/python/answers/how-to-set-the-spacing-between-subplots-in-matplotlib-in-python
5. https://stackoverflow.com/questions/43943034/seaborn-python-xtick-labels-wont-rotate
6. https://pandas.pydata.org/docs/reference/api/pandas.Series.plot.html