



GLOBAL JOURNAL OF SCIENCE FRONTIER RESEARCH: H
ENVIRONMENT & EARTH SCIENCE
Volume 23 Issue 6 Version 1.0 Year 2023
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals
Online ISSN: 2249-4626 & Print ISSN: 0975-5896

Analysis and Visualization of Fuel Consumption Against CO₂ Emission

By Shakir Adeyemi Adeyemi Tewogbade

Abstract- CO₂ emission has an adverse effect on the environment and cause greenhouse effect with significant negative climatic changes. This subsequently lead global warming which hurts both human and crops. It is important for us to perform visual analysis with available dataset using Canada as a case study.

GJSFR-H Classification: LCC Code: TD885.5.C6



Strictly as per the compliance and regulations of:



Analysis and Visualization of Fuel Consumption Against CO₂ Emission

Shakir Adeyemi Adeyemi Tewogbade

Abstract- CO₂ emission has an adverse effect on the environment and cause greenhouse effect with significant negative climatic changes. This subsequently lead global warming which hurts both human and crops. It is important for us to perform visual analysis with available dataset using Canada as a case study.

I. INTRODUCTION

The progress in technological development has led to few negative effects on the climate. Carbon Dioxide is the major compound accountable for climatic change (Nataly and Yiu, 2014). Consumption of automobile fuel produces CO₂ emission globally. As claimed by USEAP 2022, a standard vehicle releases about 4.6 metric tonnes of CO₂ annually. A study by Environment Canada in 2015 shown that private automobiles released 82 million tonnes of carbon dioxide in year 2013 alone. As noted by Carvaheira 2018, fuel consumption and CO₂ emission of a particular vehicle is based on operating variables and design characteristics (mass, aerodynamics, tyres, auxiliary systems). Beyond vehicle operating variables and design characteristics, there are other factors mentioned by Fontara, Zacharof and Ciuffo, (2017) such as weather conditions, traffic conditions, road morphology, vehicle maintenance and driving style. The degree of causal effect of CO₂ on the climate has warranted the need to study and analyze root causes based on available data that captured these variables and conditions. In this study, fuel consumption open

data from 2010 to 2014 is used for visual analysis of influencing factors of CO₂ emissions for new light-duty vehicles for retail sale in Canada. The data used is from an open source and it was collected from *Fuel consumption ratings - Open Government Portal (canada.ca)*. Open data has consent for re-use and let researchers build on existing studies (Brandon and Weber, 2022). Our dataset includes the following variables: vehicle make, vehicle model, vehicle model year, make of vehicle, size of vehicle engine, transmission, cylinder, type of fuel, fuel consumed during movement in the city, fuel consumed in the highway, and emission values.

For appropriate visualization, we have selected Python libraries like Num Py (computation of numerical values), pandas (loading and manipulating data), matplotlib (handling plots) and seaborn to carry out our analysis in Jupyter Notebook. Various types of graphs like line graph, bar chart, heat maps were plotted to answer the questions formulated from our dataset.

II. BACKGROUND

It is important to use correct visualization techniques in data analysis (Xi and Xinyu, 2021). Visualization technique selected for data analysis will be good if it is efficient, suitable and expressive (Mackinlay, 1986. Schumann and Muller, 2000). Our visual analysis is being carried out on Jupyter notebook platform. Jupyter notebook allow us to create and share files which include texts, live codes and visualizations.



Figure 1: Jupyter Web Interface

Jupyter is interactive and web-based platform where computational activities can be executed with visualization. With the notebook, users can view their codes outcomes in-line independent of other segment of the project work. Each cell containing lines of codes are seen with their corresponding outputs. To use Jupyter notebook for our study, we divided the process into four stages:

1. Launch the platform
2. Load dataset
3. Clean & process the data
4. Analysis and Visualization

To load, clean, process and analyze data, important libraries are used with Python Jupyter such as

1. Panda-loading, reshaping, merging, slicing, sorting and aggregation of data through its special data structure and operations. With Python, panda perform efficiently with data structures (Rupal and Khushboo, 2022)
2. NumPy-It is used for mathematical and numerical computation on python coding environment with capabilities for quick array processing
3. Matplotlib-a low-level library used for plotting graphs and it is a great alternative to MATLAB
4. Seaborn-was developed by Michael Waskom in 2012 to handle statistical plots. It is a high-level source unlike matplotlib with an improvement in terms of aesthetics and readability. With seaborn, line of codes for making plots will be fewer compare to matplotlib.

III. MAIN PART

Jupyter notebook exist in document format with three segments which are cells for marking down, cells for coding and result parts (Park and Sekerinski, 2018). The architectural design of Jupyter notebook is based on JavaScript browser which interacts with HTTP server through WebSocket. The webserver utilize tornado embedded in Python to relate incoming message to the kernel. kernel that provides appropriate outputs after processing of the messages and these are communicated through notebook web interface. The kernel is the core actor in carrying out execution of codes in the Notebook. In this work, our target is to write codes that import fuel consumption dataset in.csv, clean it, prepare it and perform visual analysis. As stated earlier our graphs and plots are achieved with aid of libraries with Python Jupyter with appropriate codes. Important graphs and plots that put answers to the questions posed by our fuel consumption dataset are:

a) Line Graphs

Widely used visualization technique where independent variables and dependent variables are projected on X and Y axis. Various data points are joined to show appropriate line produced by selected

dataset. Line graphs show quick glance of upward movement (direct proportional) and downward movement (inverse proportional).

b) Scatter Plot

This is like line graphs but data points are not joined but rather trendline are fitted. It displays clusters, trends and relationships. Scatter plots allow users to deduce how much data points are in close or spread out (SAS, 2014).

c) Bar Chart

Also known as column chart. It is used to display categorical data either vertically or horizontally where value of each category is represented by corresponding bar. Bar chart is readily modifiable most times with colours to capture significance differences. These are seen in stacked bar chart and clustered bar chart.

d) Box Plot

Box plots show extreme values, median and quartiles. While the plot is gotten from the interquartile range (length of the box) and median, the whisker moves the box to the minimum and maximum values without including the outliers.

e) Heat Map

It is used to display relationship between columns as represented in matrix view mode. Visualization analysis is achieved through selection of appropriate coloring. Heat map is an excellent plot in displaying variance through many variables while patterns are formed.

IV. ANALYSIS

Data are raw fact based on occurrences in human daily life and its environment. One of the means of turning data into comprehensible information and knowledge is through visualization (Narr and Yashaswini, 2020). When there is huge amount of data, there will be difficulty in understanding facts in it. With existence of data visualization techniques, visual illustrations that reveal hidden insights can be readily created. Thus, this study will be answering the questions that do with fuel combustion and CO₂ emission considering different models of cars in Canada.

1. CO₂ trend in the years of study
2. What fuel type caused most emission?
3. What make of car produced most CO₂ emission during the period of study?
4. Which vehicle class considering fuel consumption produced most CO₂ emission ?

a) Data Collection

The first step was to import the needed data in .csv file format to our working environment using Panda library.

```
#Load dataset
df = pd.read_csv("MY2010-2014_Fuel_consumption.csv")
```

Figure 3: Loading of Fuel Consumption Dataset

b) Data Cleaning

The dataset imported was viewed using syntax `df.head()` to display the first five rows for our perusal.

MODEL	MAKE	MODEL.1	VEHICLE CLASS	ENGINE SIZE	CYLINDERS	TRANSMISSION	FUEL	FUEL CONSUMPTION*	Unnamed: 9	Unnamed: 10	Unnamed: 11	CO2 EMISSIONS	
0	YEAR	NaN	# = high output engine	NaN	(L)	NaN	NaN	TYPE	CITY (L/100 km)	HWY (L/100 km)	COMB (L/100 km)	COMB (mpg)	(g/km)
1	2010	ACURA	CSX	COMPACT	2	4.0	AS5	X	10.9	7.8	9.5	30	219
2	2010	ACURA	CSX	COMPACT	2	4.0	M5	X	10	7.6	8.9	32	205
3	2010	ACURA	CSX	COMPACT	2	4.0	M6	Z	11.6	8.1	10	28	230
4	2010	ACURA	MDX AWD	SUV	3.7	6.0	AS6	Z	14.8	11.3	13.2	21	304

Figure 4: Checking the Dataset

The unnamed columns (9, 10 and 11) were given right title.

```
# renaming the columns
df.rename(columns={'MODEL':'MODEL_YEAR','MODEL.1':'MODEL','FUEL':'FUEL_TYPE',
                  'FUEL CONSUMPTION*':'FUELCONSUMPTION_CITY',
                  'Unnamed: 9':'FUELCONSUMPTION_HWY','Unnamed: 10':'FUELCONSUMPTION_COMB',
                  'Unnamed: 11':'FUELCONSUMPTION_COMB1','CO2 EMISSIONS ':'CO2_EMISSIONS' }, inplace = True)
```

Figure 5: Renaming the Unnamed Columns in the Dataset

Also, unwanted row 0 was dropped and the new data head was called to confirm if row 0 has been dropped.

```
# droping the row
df.drop(0, inplace = True)
```

Figure 6: Dropping Unnecessary Row

```
# checking to make sure the row has been dropped
df.head()
```

MODEL_YEAR	MAKE	MODEL	VEHICLE CLASS	ENGINE SIZE	CYLINDERS	TRANSMISSION	FUEL_TYPE	FUELCONSUMPTION_CITY	FUELCONSUMPTION_HWY	
1	2010	ACURA	CSX	COMPACT	2	4.0	AS5	X	10.9	7.8
2	2010	ACURA	CSX	COMPACT	2	4.0	M5	X	10	7.6
3	2010	ACURA	CSX	COMPACT	2	4.0	M6	Z	11.6	8.1
4	2010	ACURA	MDX AWD	SUV	3.7	6.0	AS6	Z	14.8	11.3
5	2010	ACURA	RDX AWD TURBO	SUV	2.3	4.0	AS5	Z	13.2	10.3

Figure 7: Rechecking Our Dataset to Confirm the Row Has Been Updated

Codes were run to confirm the data type represented by each column and existence of null in the whole dataset.

```
df.info()
<class 'pandas.core.frame.DataFrame'>
Int64Index: 5359 entries, 1 to 5359
Data columns (total 12 columns):
 #   Column           Non-Null Count  Dtype  
--- 
 0   MODEL_YEAR        5359 non-null   int32  
 1   MAKE              5359 non-null   object  
 2   MODEL             5359 non-null   object  
 3   VEHICLE_CLASS    5359 non-null   object  
 4   ENGINE_SIZE       5359 non-null   float64 
 5   CYLINDERS         5359 non-null   float64 
 6   TRANSMISSION      5359 non-null   object  
 7   FUEL_TYPE         5359 non-null   object  
 8   FUELCONSUMPTION_CITY 5359 non-null   float64 
 9   FUELCONSUMPTION_HWY   5359 non-null   float64 
 10  FUELCONSUMPTION_COMB 5359 non-null   float64 
 11  CO2_EMISSIONS     5359 non-null   int64  
dtypes: float64(5), int32(1), int64(1), object(5)
memory usage: 523.3+ KB
```

Figure 8: Checking Out the Data Type

The last action in our data cleaning was to confirm any missing values and none was observed in the dataset.

```
# checking for missing values
df.isnull().sum()

MODEL_YEAR      0
MAKE            0
MODEL           0
VEHICLE_CLASS   0
ENGINE_SIZE     0
CYLINDERS       0
TRANSMISSION    0
FUEL_TYPE       0
FUELCONSUMPTION_CITY 0
FUELCONSUMPTION_HWY   0
FUELCONSUMPTION_COMB 0
CO2_EMISSIONS   0
dtype: int64
```

Figure 9: Checking Out Missing Values

c) Data Preparation

We checked our dataset after all the cleaning to be sure that it is reading for visualization.

```
df.head()

  MODEL_YEAR  MAKE  MODEL  VEHICLE_CLASS  ENGINE_SIZE  CYLINDERS  TRANSMISSION  FUEL_TYPE  FUELCONSUMPTION_CITY  FUELCONSUMPTION_HWY
1  2010  ACURA  CSX  COMPACT          2        4.0       AS5      X        10.9          7.8
2  2010  ACURA  CSX  COMPACT          2        4.0       M5      X        10          7.6
3  2010  ACURA  CSX  COMPACT          2        4.0       M6      Z        11.6          8.1
4  2010  ACURA  MDX AWD          SUV          3.7        6.0       AS6      Z        14.8          11.3
5  2010  ACURA  RDX AWD TURBO      SUV          2.3        4.0       AS5      Z        13.2          10.3
```

Figure 10: Confirming That the Dataset is Properly Cleaned

Columns are confirmed for uniqueness.

```
len(data['Make'].unique())
```

45

Figure 11: Confirming Uniqueness of Value in "Make" Column

d) Data Analysis

The first set of plots with our dataset is to show trend in CO2 emission from 2010 to 2014.

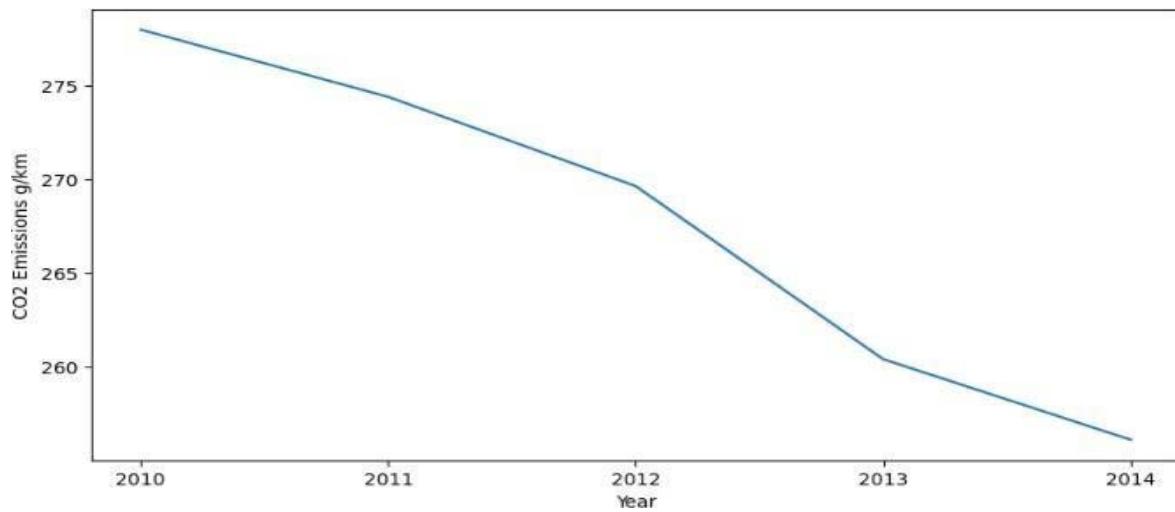


Figure 12: Line Graph to Show Trend Between Year 2010 and 2014

The plot shows downward trend which support various governmental policies in reducing emission and greenhouse effect.

The next visual analysis is setting up heat map to show interaction between our dataset attributes.

```
plt.figure(figsize=(12,8))
sns.heatmap(data.corr());
```

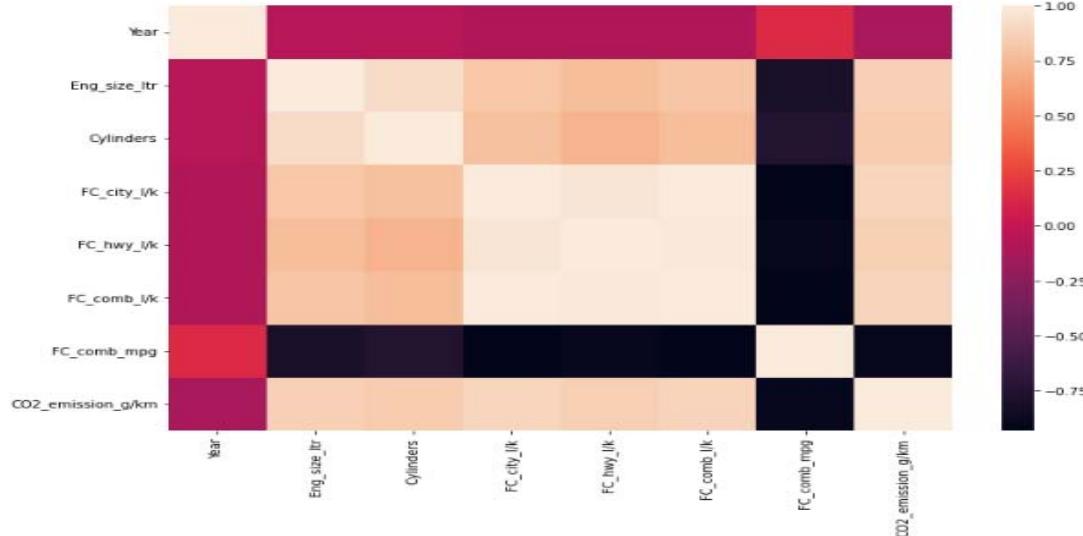


Figure 13: Heatmap Showing Correlation Among Variables

It was observed from the heat map that there is high positive correlation between fuel consumptions, engine size and cylinders with CO2. Thus, we extend our visualization to plotting of fuel types with CO2.

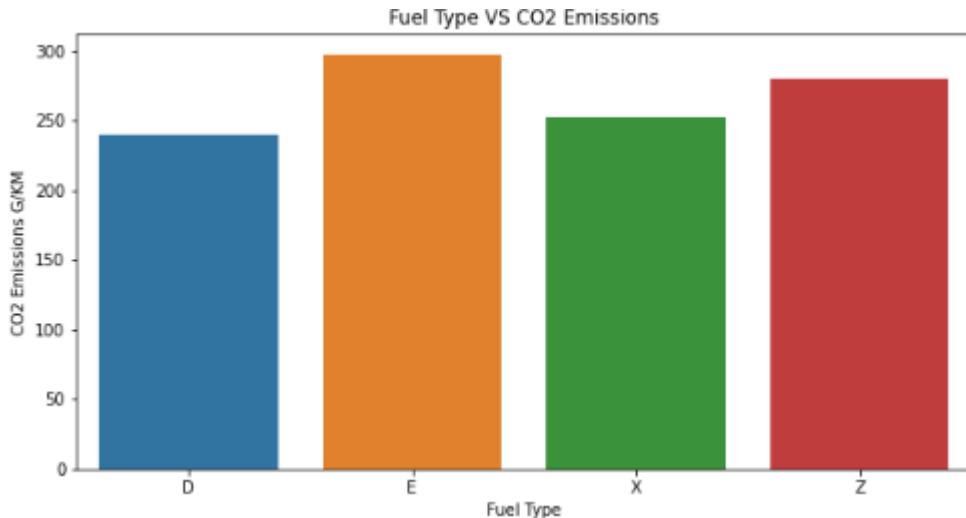


Figure 14: Bar Plot of CO2 Emission Against Fuel Type

The plot showed that fuel type E and Z yields more CO2 emission during the period under study. In a similar manner we plotted vehicle class based on fuel consumption against CO2 emission using seaborn library. The graph showed that Van passenger and Van cargo yielded highest number of CO2 emission between 2010 to 2014. We can also infer that weight of the vehicle has determining effect on CO2 emission. The bigger vehicles are seen towards the right with high value of emission. This supports proposal by Pagerit et al, 2006, Wohlecker et al 2007 and Bishop et al 2014 as mentioned in our introduction.

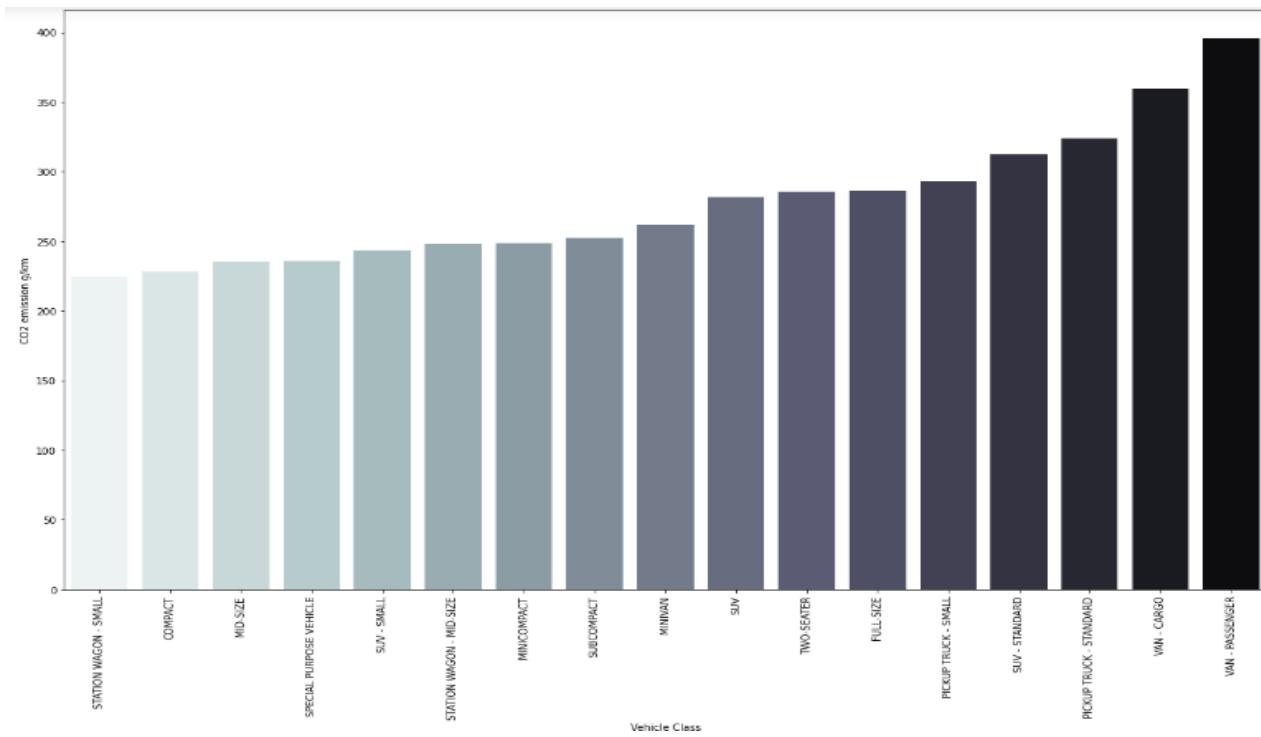


Figure 15: Bar Plot of Vehicle Class With Fuel Consumption Against CO2 Emission

Moving forward, we produced another visual that display which of the vehicle make produced most emission.

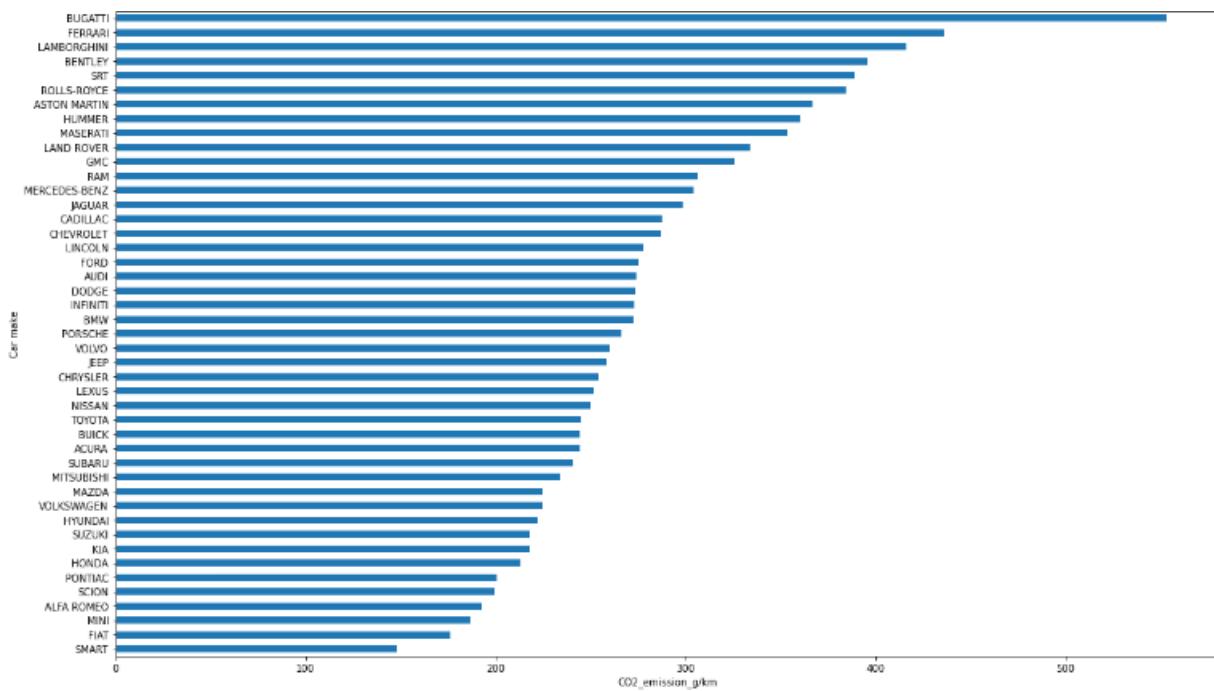


Figure 16: Bar Plot of Vehicle Make Against CO2 Emission

The graph (Fig. 16) shows that Bugatti lead in term of amount of CO2 emission produced into the environment. Bugatti uses fuel type Z with more controlling impact CO2 emission as shown in Fig. 14 above. As identified by our heat map, interaction between engine size CO2 emission is plotted using scatter plot through seaborn.

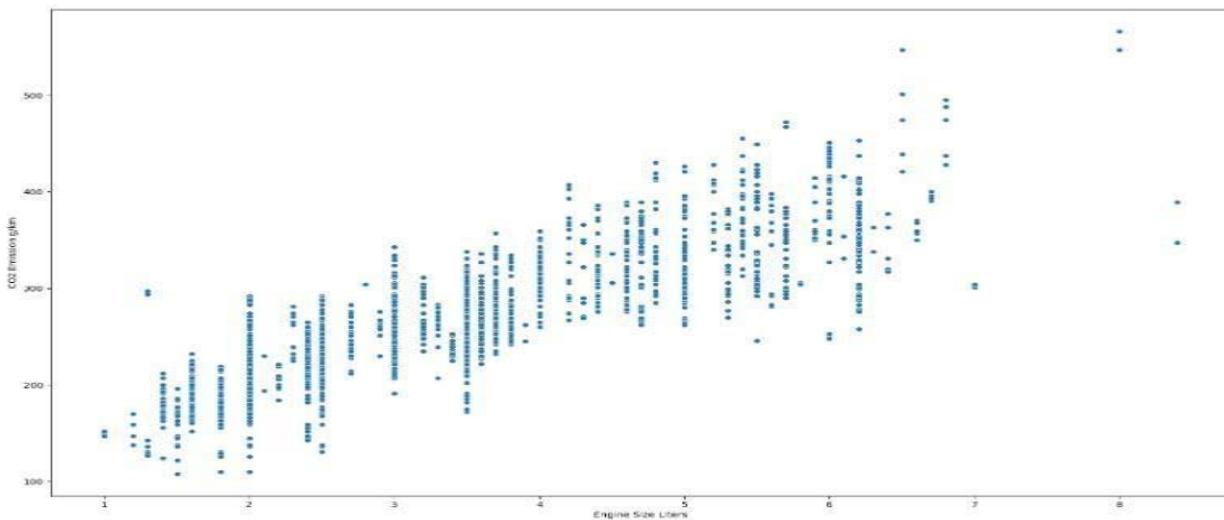


Figure 17: Scatter Plot CO2 emission against Engine size

From the display, there is a strong direct proportional relationship between the two variables. The bigger the engine size the higher the CO2 emission.

V. DISCUSSION

Analysis performed on the dataset will validate emission models as generated by the various attributes. The dataset contains 5359 records with 12 attributes and as such it is hard to see information the raw figures

is speaking to. Representing the data on various visual plots allow us to see hidden information at ease. Ben and Rachel 2015, used applicable visualization techniques like column chart, pie chart, line plot to analyze fuel consumption data. Similarly, Bielaczyc, Szcztka and Woodburn 2019 used column chart to represent fuel type plot against CO2 emission and contour map to display emission in certain locations with vehicle load and speed.

The application of bar charts (Fig.15 & 16) in our analysis has easily been achieved because of its robustness in representing categorical data, perhaps Cleveland's dot plot would have taken fewer spaces with improved aesthetic for plot like fuel type vs CO2 emission. For simplicity each dot will be represented as 40 g/km of emission. Dot plot uses minimum ink to optimum effect and still deliver excellent design (Tufte, 1983. Dave, Jaap and Ian, 2005).

Heat map has been widely used in many visualizations analysis due to its intuitive approach of colouring and ability to present interaction among variables in a single diagram. In work such as ours, we could have introduced our heat map after plotting table lens graphs. As asserted by Sinar 2015, table lens has a very high efficiency in yielding many interactions in a single plot while serving as starter in dataset visualization. Also, in addition to our heatmap, facet (Trellis) plots can be used to create additional interactions (sub-plots) for variables showing strong correlation from the map.

VI. CONCLUSION

Data visualization has gained great popularity with advancement of software technology and variety of platforms. One of the popular platforms to create visualization is Jupyter notebook where cells for codes and visual displays are available on the interface. We have used data visualization to investigate controlling effects of fuel consumption on CO2 emission. Variety of techniques such as line chart, bar chart, heat maps and scatter plot were used to analyze the field data in order to create informative patterns on level of influence of various variables on CO2 emission. Our visual analysis revealed resultant effects of important variables that need to be curtailed to minimize CO2 emission in the environment. This kind of study will assist policy makers to find effective solutions to climatic changes caused by vehicle movements.

In as much as we have efficient visuals which produced graphical display of raw data., there are few exceptions. The exceptions were critically reviewed to create room for improvement. The improvement will yield visual that create more robust outcomes where concentrated interactions are revealed in our visuals and nicer aesthetic.

REFERENCES RÉFÉRENCES REFERENCIAS

1. Ben, S., and Rachel, M. (2015). Literature review: Real-world fuel consumption of heavy-duty vehicles in the United States, China and European Union. The International Council on clean transportation. White paper.
2. Belaczyc, P., Szczotka, A., and Woodburn, J. (2019). Carbon dioxide emissions and fuel consumption from passenger cars tested over the NEDC and WLTC—an overview and experimental results from market representative vehicles. IOP conference series: Earth and Environmental Science 214 012136. doi:10.1088/1755-1315/214/1/012136
3. Bishop, J., Martin W., and Boies, A. (2014). Cost effectiveness of alternative power-trains for reduced energy use and CO2 emissions in passenger vehicles. Appl Energy 2014: 124, 14 -61.
4. Brandon, L., and Nick, W. (2022). Ethics of open data. arXiv: 2205. <https://doi.org/10.48550/arXiv.2205.10402>: [accessed on 19 November 2022].
5. Carvalheira, P. (2018). A Model for the calculations of CO2 emissions and fuel consumption of a diesel engine driven car in the WEDC. Proceedings of the 1st Iberic Conference on Theoretical and Experimental Mechanics and Materials. 11th National Congress on Experimental Mechanics. ISBN: 978-989-20-8771-9
6. Dave, K., Jaap, J., and Ian, W. (2005). Designing Science graphs for data analysis and presentation. The bad, the good and the better. Science & Technical Publishing Department of Conservation Wellington, New Zealand.
7. Fontaras, G., Zacharof, N., and Ciuffo, B. (2017). Fuel consumption and CO2 emissions from passenger cars in Europe - Laboratory versus real-world emissions. Progress in Energy and Combustion Science.60, 97-131.
8. Mackinlay, J. (1986). Automating the design of graphical presentations of relational information. ACM Trans. Graph, 2(5), 110 -141.
9. Nataly, E., and Yiu, J. (2014). Energy Related CO2 emissions and the progress on CCS projects: a review. Renewable Sustainable Energy Review 31: 368 – 385
10. Nejat, P., et al, (2015). A global review of energy consumption, CO2 emissions and policy in the residential sector (with an overview of the top ten CO2 emitting countries). Renewable and Sustainable Energy Reviews 43: 843 – 862. <https://doi.org/10.1016/j.rser.2014.11.066>.
11. Ombach, G., and Junak, J, (2010). Weight and efficiency optimization of auxiliary drivers used in automobile. IEEE: 2010 p 1-6. doi:10.1109/ICELMACH.2010.5608280.
12. Pagerit, S., Shaver, P., and Rousseau, A. (2006). Fuel economy sensitivity to vehicle mass for advanced vehicle power trains. SAE Technical Paper 2006-01-0665. <https://doi.org/10.4271/2006-01-0665>.
13. Rupal, S., and Khushboo, V. (2022). Data Analysis Using Pandas Library of Python. Acts Scientific Computer Sciences, 4(3)
14. SAS (2014). Data Visualization Techniques from Basics to Big Data with SAS Visual Analytics.

15. Schumann, H., and Muller, W. (2000). Visualisierung-Grundlagen and Allgemeine Methoden. Springer- Verlag, Berlin.
16. Spencer, P., and Emil, S. (2018). A notebook format for the holistic design of embedded systems (Tool Paper). Formal Integrated Development Environment EPTC 284, 2018, 85 -94.
17. Tufte, E. (1983). The visual display of quantitative information. Graphics Press, Cheshire, Connecticut.
18. USEAP. (2022). Greenhouse Gas Emissions from a typical passenger vehicle. Greenhouse Gas Emissions from a Typical Passenger Vehicle | US EPA. Accessed 21 November 2022.
19. Wohlecker, R., Johannaber, M., and Espig M. (2007). Determination of weight elasticity of fuel economy for ICE hybrid and fuel cell vehicles. SAE Technical Paper 2007-01-0343. <https://doi.org/10.4271/2007-01-0343>.
20. Xi, C., and Xinyu, C. (2021). Data Visualization in Smart Grid and Low-Carbon energy systems: A review. International Transactions on Electrical Energy Systems, 13(7). <https://doi.org/10.1002/2050-7038.12889>.