

CrossRef DOI of original article:

# 1 Analysis and Visualization of Fuel Consumption Against CO 2 2 Emission

3 Shakir Adeyemi Adeyemi Tewogbade<sup>1</sup>

4 <sup>1</sup> Bradford, West Yorkshire, United Kingdom

5 *Received: 1 January 1970 Accepted: 1 January 1970 Published: 1 January 1970*

6

---

## 7 **Abstract**

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

12

---

13 *Index terms—*

## 14 **1 Analysis and Visualization of Fuel Consumption**

15 Against CO 2 Emission Shakir Adeyemi Adeyemi Tewogbade an open source and it was collected from Fuel  
16 consumption ratings -Open Government Portal (canada.ca). Open data has consent for re-use and let researchers  
17 build on existing studies (Brandon and Weber, 2022). Our dataset includes the following variables: vehicle make,  
18 vehicle model, vehicle model year, make of vehicle, size of vehicle engine, transmission, cylinder, type of fuel, fuel  
19 consumed during movement in the city, fuel consumed in the highway, and emission values.

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

24 It is important to use correct visualization techniques in data analysis (Xi and Xinyu, 2021). Visualization  
25 technique selected for data analysis will be good if it is efficient, suitable and expressive (Mackinlay, 1986.  
26 Schumann and Muller, 2000). Our visual analysis is being carried out on Jupyter notebook platform. Jupyter  
27 notebook allow us to create and share files which include texts, live codes and visualizations. Abstract-CO2  
28 emission has an adverse effect on the environment and cause greenhouse effect with significant negative climatic  
29 changes. This subsequently lead global warming which hurts both human and crops. It is important for us to  
30 perform visual analysis with available dataset using Canada as a case study.

## 31 **2 II. Background**

32 Author: e-mail: pingcommercial@gmail.com

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

## 37 **3 c) Bar Chart**

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

41 **4 d) Box Plot**

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

45 **5 e) Heat Map**

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

49 **6 IV. Analysis**

50 Data are raw fact based on occurrences in human daily life and its environment. One of the means of turning  
51 data into comprehensible information and knowledge is through visualization (Narra and Yashaswini, 2020).  
52 When there is huge amount of data, there will be difficulty in understanding facts in it. With existence of data  
53 visualization techniques, visual illustrations that reveal hidden insights can be readily created. Thus, this study  
54 will be answering the questions that do with fuel combustion and C02 emission considering different models of  
55 cars in Canada.

56 1. C02 trend in the years of study 2. What fuel type caused most emission? 3. What make of car produced  
57 most C02 emission during the period of study? 4. Which vehicle class considering fuel consumption produced  
58 most C02 emission ?

59 **7 a) Data Collection**

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

61 **8 Analysis and Visualization of Fuel Consumption Against Co2  
62 Emission**

63 **9 Global Journal of Science Frontier Research ( H ) XXIII Issue  
64 VI Version I Year 2023**

65 **10 © 2023 Global Journals**

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

75 **11 a) Line Graphs**

76 Widely used visualization technique where independent variables and dependent variables are projected on X  
77 and Y axis. Various data points are joined to show appropriate line produced by selected The dataset imported  
78 was viewed using syntax df.head () to display the first five rows for our perusal. We checked our dataset after all  
79 the cleaning to be sure that it is reading for visualization.

---

80 **12 III. Main Part**  
81 **13 Analysis and Visualization of Fuel Consumption Against Co2**  
82 **Emission**  
83 **14 Global Journal of Science Frontier Research ( H ) XXIII**  
84 **Issue VI Version I Year 2023**  
85 **15 Analysis and Visualization of Fuel Consumption Against Co2**  
86 **Emission**  
87 **16 Global Journal of Science Frontier Research ( H ) XXIII**  
88 **Issue VI Version I Year 2023 d) Data Analysis**

89 The first set of plots with our dataset is to show trend in C02 emission from 2010 to 2014. The plot shows  
90 downward trend which support various governmental policies in reducing emission and greenhouse effect. The  
91 next visual analysis is setting up heat map to show interaction between our dataset attributes.

92 **17 Figure 13: Heatmap Showing Correlation Among Variables**

93 It was observed from the heat map that there is high positive correlation between fuel consumptions, engine size  
94 and cylinders with C02. Thus, we extend our visualization to plotting of fuel types with C02. Moving forward,  
95 we produced another visual that display which of the vehicle make produced most emission. The graph (Fig. 16)  
96 shows that Bugatti lead in term of amount of C02 emission produced into the environment. Bugatti uses fuel  
97 type Z with more controlling impact CO2 emission as shown in Fig. 14 above. As identified by our heat map,  
98 interaction between engine size C02 emission is plotted using scatter plot through seaborn. From the display,  
99 there is a strong direct proportional relationship between the two variables. The bigger the engine size the higher  
100 the C02 emission.

101 **18 Analysis and Visualization of Fuel Consumption Against Co2**  
102 **Emission**

103 **19 Global Journal of Science Frontier Research ( H ) XXIII**  
104 **Issue VI Version I Year 2023**

105 **20 Analysis and Visualization of Fuel Consumption Against Co2**  
106 **Emission**

107 **21 Global Journal of Science Frontier Research ( H ) XXIII**  
108 **Issue VI Version I**

109 Year 2023

110 Analysis performed on the dataset will validate emission models as generated by the various attributes. The  
111 dataset contains 5359 records with 12 attributes and as such it is hard to see information the raw figures is  
112 speaking to. Representing the data on various visual plots allow us to see hidden information at ease. Ben  
113 and Rachel 2015, used applicable visualization techniques like column chart, pie chart, line plot to analyze fuel  
114 consumption data. Similarly, Bielaczyc, Szcztka and Woodburn 2019 used column chart to represent fuel type  
115 plot against C02 emission and contour map to display emission in certain locations with vehicle load and speed.

116 **22 Analysis and Visualization of Fuel Consumption Against Co2**  
117 **Emission**

118 **23 Global Journal of Science Research ( H ) XXIII Issue VI**  
119 **Version I Year 2023**

120 **24 59**

121 © 2023 Global Journals

122 **25 V. Discussion**

123 The application of bar charts (Fig. 15 &16) in our analysis has easily been achieved because of its robustness in  
124 representing categorical data, perhaps Cleveland's dot plot would have taken fewer spaces with improved aesthetic  
125 for plot like fuel type vs C02 emission. For simplicity each dot will be represented as 40 g/km of emission. Dot  
126 plot uses minimum ink to optimum effect and still deliver excellent design (Tufte, 1983. Dave, Jaap and Ian,  
127 2005).

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

134 **26 VI. Conclusion**

135 Data visualization has gained great popularity with advancement of software technology and variety of platforms.  
136 One of the popular platforms to create visualization is Jupyter notebook where cells for codes and visual displays  
137 are available on the interface. We have used data visualization to investigate controlling effects of fuel consumption  
138 on C02 emission. Variety of techniques such as line chart, bar chart, heat maps and scatter plot were used to  
139 analyze the field data in order to create informative patterns on level of influence of various variables on C02  
140 emission. Our visual analysis revealed resultant effects of important variables that need to be curtailed to  
141 minimize C02 emission in the environment. This kind of study will assist policy makers to find effective solutions  
142 to climatic changes caused by vehicle movements.

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

146 **27 References Références Referencias**



Figure 1:



Figure 2: Figure 1 :

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

Figure 3: Figure 4 :

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)
1	2010	ACURA	CSX	COMPACT	2	4.0	AS5	X	10.9	7.8	9.5	30
2	2010	ACURA	CSX	COMPACT	2	4.0	M5	X	10	7.6	8.9	32
3	2010	ACURA	CSX	COMPACT	2	4.0	M6	Z	11.6	8.1	10	28
4	2010	ACURA	MDX AWD	SUV	3.7	6.0	AS6	Z	14.8	11.3	13.2	21
5												304

Figure 4: Figure 5 :

```
6 # 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: Figure 6 :

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

Figure 6: Figure 7 :

	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: Figure 3 :

## 27 REFERENCES RÉFÉRENCES REFERENCIAS

---

```
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
```

8

Figure 8: Figure 8 :

```
# 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
9  dtype: int64
```

9

Figure 9: Figure 9 :

```
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      10.0        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
10
```

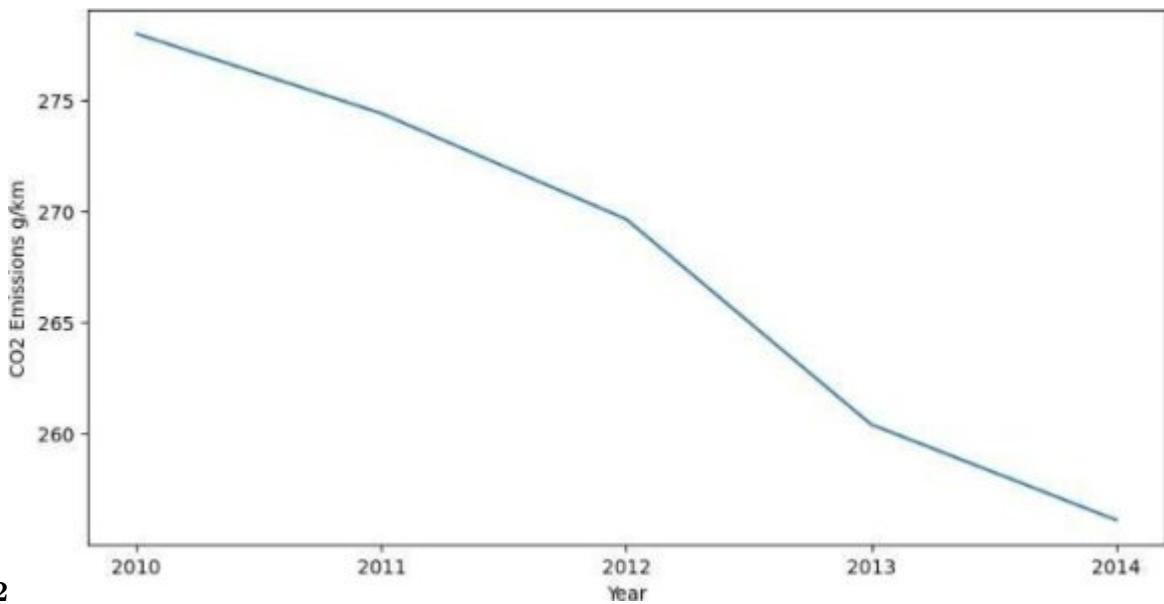
10

Figure 10: Figure 10 :

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

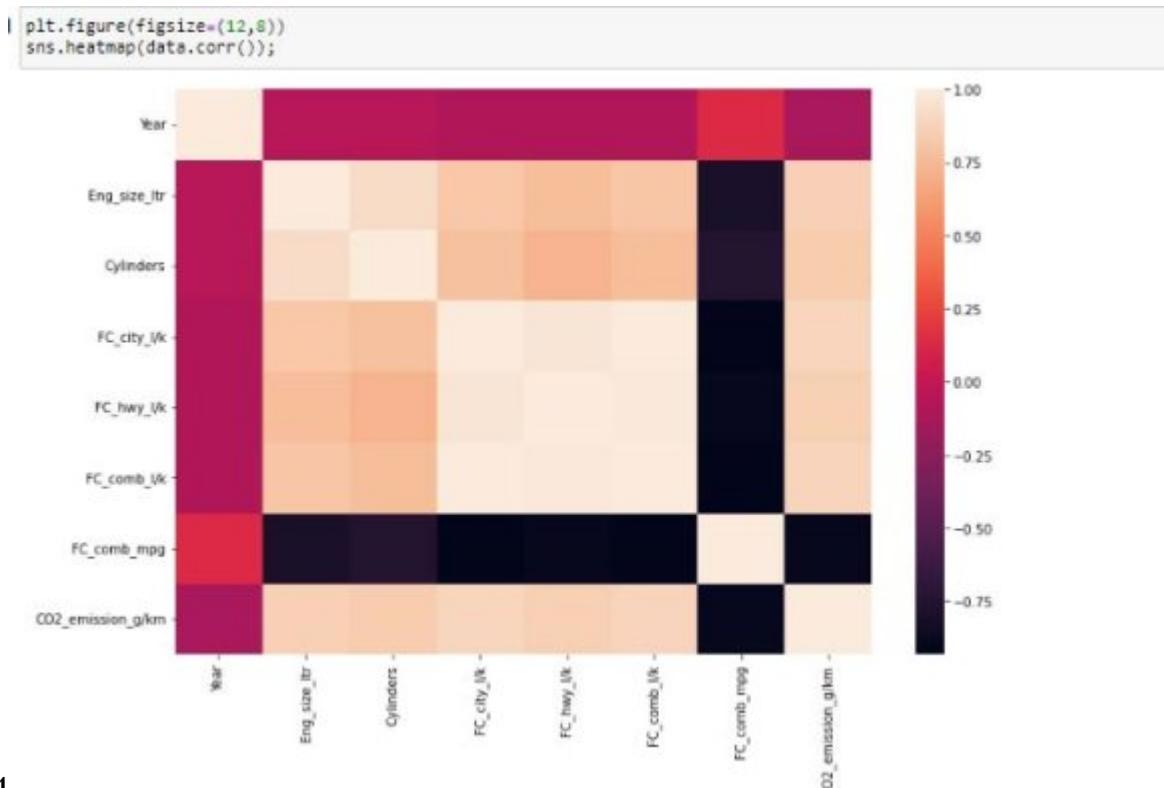
11 45

Figure 11: Figure 11 :



12

Figure 12: Figure 12 :



14

Figure 13: Figure 14 :

15

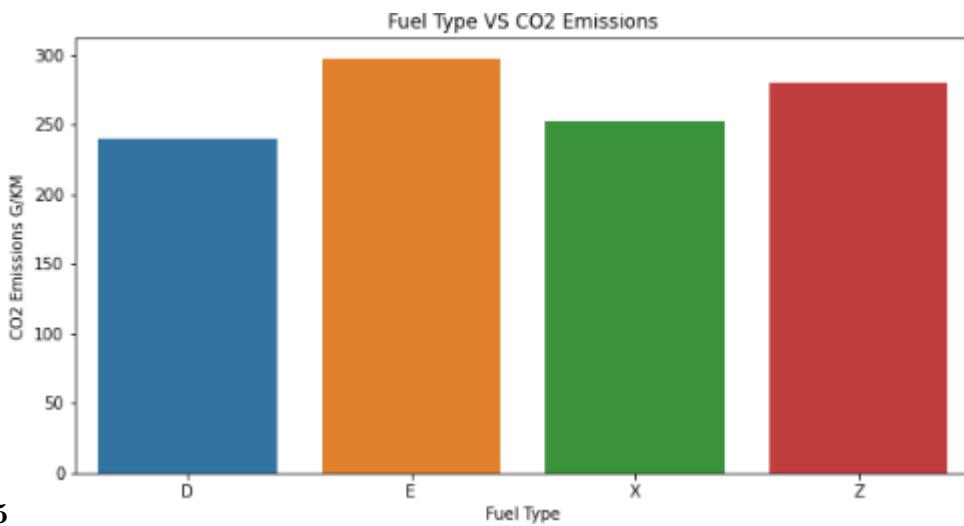


Figure 14: Figure 15 :

16

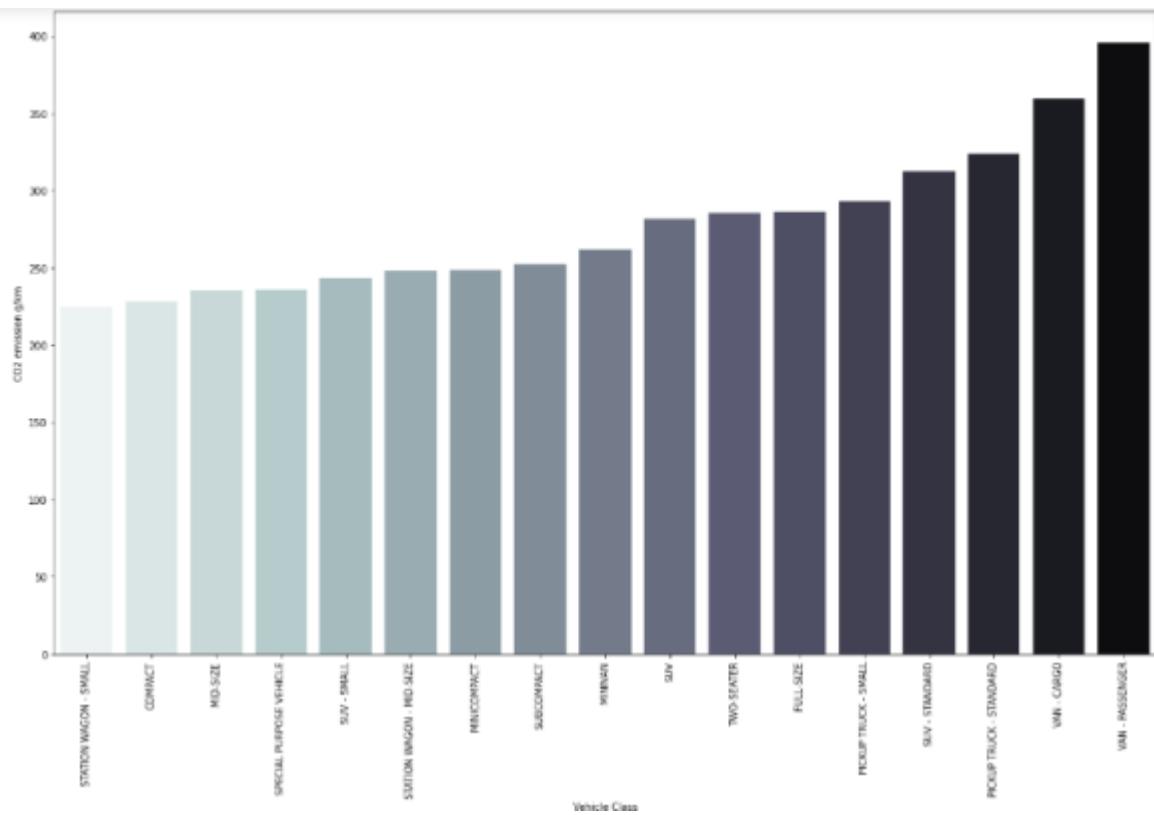
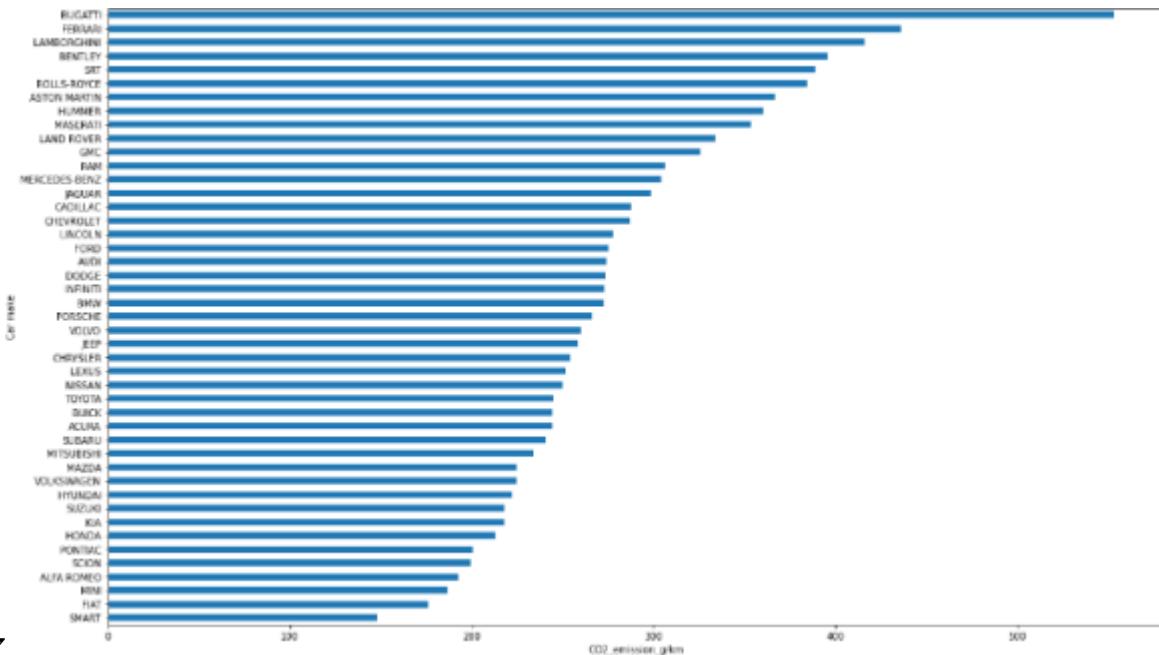


Figure 15: Figure 16 :



17

Figure 16: Figure 17 :

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

b) Scatter Plot

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.

Figure 17:

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 CO<sub>2</sub> 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 CO<sub>2</sub> emissions and fuel consumption of a diesel engine driven car in the WEDC. Proceedings of the 1<sup>st</sup> Iberic Conference on Theoretical and Experimental Mechanics and Materials. 11<sup>th</sup> 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. Analysis 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

Journal of Science Frontier Research ( H ) XXIII Issue VI Version IYear 2023

Figure 18: and Visualization of Fuel Consumption Against Co<sub>2</sub> Emission Global

---

148 [Visualisierung-Grundlagen and Methoden] , Allgemeine Visualisierung-Grundlagen , Methoden . Berlin:  
149 Springer-Verlag.

150 [Muller ()] , H Muller , W . 2000.

151 [Spencer and Emil ()] 'A notebook format for the holistic design of embedded systems'. P Spencer , S Emil .  
152 *Formal Integrated Development Environment EPTC* 2018. 2018. 284 p. . (Tool Paper)

153 [Xi and Xinyu ()] 'Data Visualization in Smart Grid and Low-Carbon energy systems: A review'. C Xi , C Xinyu  
154 . 10.1002/2050-7038.12889. <https://doi.org/10.1002/2050-7038.12889> *International Transactions  
155 on Electrical Energy Systems* 2021. (7) p. 13.

156 [Wohlecker et al. ()] *Determination of weight elasticity of fuel economy for ICE hybrid and fuel cell vehicles*,  
157 R Wohlecker , M Johannaber , M Espig . 10.4271/2007-01-0343. SAE Technical Paper 2007-01-0343.  
158 <https://doi.org/10.4271/2007-01-0343> 2007.

159 [Useap (2022)] *Greenhouse Gas Emissions from a typical passenger vehicle. Greenhouse Gas Emissions from a  
160 Typical Passenger Vehicle / US EPA*, Useap . 2022. 21 November 2022.

161 [Tufte ()] *The visual display of quantitative information*, E Tufte . 1983. Cheshire, Connecticut: Graphics Press.