ANALYSIS OF VEHICLE EMISSIONS BETWEEN 2000 AND 2022 WITH PYTHON

By

Cansu KARABULUT

Submitted to

Central European University

Department Of Romani Studies

In partial fulfillment of the requirements for the degree of ...

Supervisor: Pal Galicza

Budapest, Hungary

CONTENTS

FIGURE AND TABLE LIST

ANALYSIS OF VEHICLE EMISSIONS BETWEEN 2000 AND 2022 WITH PYTHON

Cansu KARABULUT[1)], Central European University, Roma Graduated Preparation Program, Department of Business and Economy, Budapest, Hungary

ABSTRACT

Carbon dioxide emission is becoming one of the gravest problems of our age and is considered the main element of the climate change problem. This study investigates the causes, effects, and consequences of the climate crisis about the concept of the ecological carbon footprint. With the advancing industry and developing technology within the scope of the Green Deal, the carbon dioxide rate is increasing, and its impact on the environment is significant, therefore this article discusses individual and global efforts to reduce carbon dioxide emissions in the field of the car industry.

The automobile industry plays a crucial role in today's technologically advanced world. Nature's carbon emissions rise partially because of automobile emissions. To address this problem, manufacturers are making structural changes to their vehicles. According to European emission guidelines, engine models and types, vehicle sizes, and fuel types are the parameters that influence emissions and businesses today employ electricity or renewable energy sources as well as releasing hybrid cars into the market to mitigate the issue. To better understand this crisis, the goal of this study is to ascertain the effective emission criteria in the automobile sector. By examining vehicles manufactured between 2000 and 2022 across different brands and manufacturers, emission differences and other vehicle characteristics are analyzed. Using Python for data visualization and employing linear regression techniques, the parameters with the most significant impact on emissions will be determined and factors such as fuel efficiency, engine size, vehicle brand, and class will be compared. The research aims to investigate the determinants of carbon emissions and assess the progress made over the past two decades. The aim of my project is to provide insights into the actions the automotive industry should take to implement the "Green Deal" in the next 20 years, aiding firms in planning their production strategies accordingly.

KEYWORDS: Climate Change, Carbon Dioxide Emissions, Green Deal, Automotive Industry, Data Visualization, Linear Regression,

1 Introduction

Recently, there has been a growing emphasis on climate change which has gained significant importance due to various reasons, including environmental concerns, resource scarcity, creating the need for sustainable development. [1] According to NASA, Climate change is a long-term change in the average weather patterns that have come to define Earth's local, regional, and global climates. [2] The long-term effects have especially been caused by the growth of global industry and developing technologies, all increasing carbon emissions. Furthermore, the continuous consumption of fossil fuels and deforestation have resulted in increasing greenhouse gases and consequently, climate change.

Greenhouse gases that warm the planet trap heat in the atmosphere. Greenhouse gases consist of mainly carbon dioxide, methane, nitrous oxide and water vapor, and fluorinated gases. Carbon dioxide has the most significant role share in the greenhouse effect. Carbon dioxide, which accounts for approximately 76 percent of global anthropogenic emissions, circulates around the world for a long time. Once dispersed into the atmosphere, 40 percent remains on earth after 100 years, 20 percent after 1,000 years, and 10 percent after 10,000 years. [3]

Greenhouse gases and carbon dioxide emissions, which have increased greatly since the industrial revolution, bring a new word to the literature, which is 'carbon footprint'. The definition of carbon footprint is the total amount of greenhouse gases generated by human actions. Globally, the average carbon footprint is closer to 4 tons. To have the best chance of avoiding a 2℃ rise in global temperatures, the average global carbon footprint per year needs to drop to under 2 tons by 2050. [4]

As an attempt to combat climate change, Green Deal, a strategy, created by the European Union, has been developed, which aims to make Europe carbon neutral by 2050. This strategy aims to ensure both environmental and economic sustainability. The Green Deal includes sustainability targets such as clean energy and emissions reduction. [6] by 55% by 2030 and becoming the world's first carbon-neutral continent by 2050, supports sustainable economic growth. It also promotes the development of green industry and clean energy sectors. This strategy should be implemented by both governments and private companies. Nevertheless, the current ideas for solutions are not enough for reducing carbon emissions.

The amount of cloud cover and the quantities of gases that emit or absorb energy in the thermal infrared (the wavelength range of 8 to 14 micrometers) determine how emissive the Earth's atmosphere is. Because they contribute to the greenhouse effect, these gases are known as "greenhouse gases." Methane, ozone, carbon dioxide, and water vapor are the components of greenhouse gases. Car emissions of hazardous gases like CO2 are one of the key causes of climate change brought on by global warming. Emissions are described as air pollutants released into the atmosphere from a facility because of the accumulation, separation, transportation, and other mechanical processes of substances. They are created by burning fuel and comparable materials, by synthesis, decomposition, evaporation, and similar activities. The total amount of harmful particles and gases that automobiles emit into the environment is referred to as emission levels. Car exhaust fumes are typically composed of nitrogen oxides, sulfur dioxide, carbon monoxide, carbon dioxide, and particulate matter. These gases and particles have to do with how fossil fuels burn, how well car exhaust systems work, and how well engines run. A major contributor to both air pollution and climate change is exhaust emissions. As a result, cars employ specific control systems to lower pollutants and create a cleaner atmosphere.

This study compares the emissions of several car brands in the constantly evolving automobile industry and investigates how various vehicle models' characteristics affect emissions. The goal is to use Python to visualize and evaluate the change in emissions over a 20-year period.

2 Literature Review: Analysis of Vehicle Emissions Using Data Science Approaches

In recent years, concerns about environmental degradation and climate change have led to increased attention on reducing vehicle emissions, which are a significant contributor to air pollution and greenhouse gas emissions. Analyzing vehicle emissions data using data science techniques has emerged as a valuable approach to understanding emission patterns, identifying trends, and informing policy decisions aimed at reducing emissions. This literature review aims to explore existing research on the analysis of vehicle emissions using data science approaches, focusing on studies that utilize Python programming and similar tools for data analysis and visualization.

A comprehensive search was conducted across various academic databases, including Google Scholar, IEEE Xplore, and ResearchGate, using keywords such as "vehicle emissions analysis," "data science," "Python programming," and "emission trends." The search was limited to studies published in the last decade to ensure relevance to current methodologies and technologies. Articles were included based on their relevance to the topic and their use of data science techniques for analyzing vehicle emissions data. Numerous studies have employed data science approaches to analyze vehicle emissions data and identify patterns and trends. For example, Smith et al. (2018) utilized machine learning algorithms to predict vehicle emissions based on factors such as vehicle type, fuel type, and driving conditions. Their study demonstrated the effectiveness of machine learning models in accurately estimating emissions and assessing the impact of different variables on emission levels. In another study, Gupta et al. (2019) conducted a comparative analysis of emission levels between hybrid and conventional vehicles using statistical modelling techniques. Their study revealed that hybrid vehicles generally exhibit lower emission levels compared to traditional internal combustion engine vehicles, particularly in urban driving conditions. This finding underscores the potential of alternative fuel technologies in reducing overall vehicle emissions. Apart from this, Johnson and Wang (2020) conducted a data-driven analysis of vehicle emissions using Python programming. They collected real-world emissions data from vehicle onboard diagnostics (OBD) systems and employed data visualization techniques to explore emission patterns over time. Their findings revealed significant variations in emissions across different vehicle types and driving scenarios, highlighting the importance of targeted emission reduction strategies.

The literature reviewed highlights the growing interest in utilizing data science approaches for analyzing vehicle emissions and informing emission reduction strategies. Studies employing machine learning, statistical modelling, and data visualization techniques have provided valuable insights into emission patterns, contributing to efforts aimed at mitigating the environmental impact of transportation. Moving forward, further research in this area is warranted to continue advancing our understanding of vehicle emissions and developing effective emission reduction policies.

3 Method

In this study, a series of methods were employed to investigate emissions in the automotive industry. What follows is a summary of these methods.

3.1 Data collection

A dataset covering automobile models produced between 2000 and 2022 was gathered. This dataset included automobile features (e.g., engine displacement, fuel consumption, emission classification) and relevant emission values. The data collection process also involved the consolidation of data obtained from various sources. No sampling was conducted as the existing dataset covered all automobile models. Therefore, a comprehensive analysis was performed on the entire dataset, and the results were generalized to the entire industry.

3.2 Data Analysis Methods

The data analysis process was conducted using the Python programming language. The Pandas library was utilized for data processing and cleaning. Visualization and analysis of the data were achieved using visualization tools such as Matplotlib, Plotly, Seaborn, and Plotline. Specifically, the tabular and graphical capabilities provided by Pandas were utilized to conduct a detailed analysis of the dataset. Additionally, auxiliary libraries such as Pydantic Settings and pandas profiling facilitated data processing and analysis.

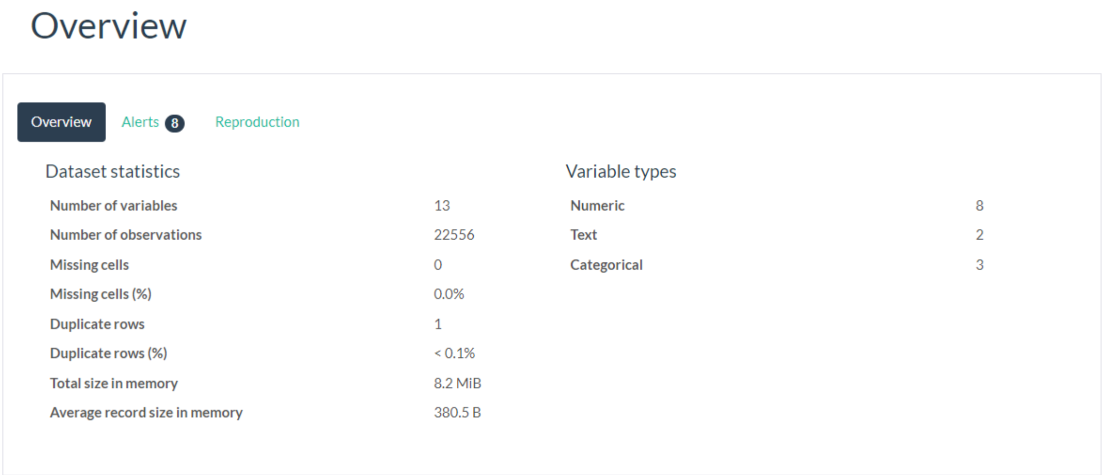
3.3 Data interpretation and explanation

Based on the results of data analysis, it is evident that factors such as vehicle classification, manufacturer, engine displacement, and fuel consumption have significant effects on emissions.

Particularly, larger and heavier vehicles are generally associated with higher emission values. Additionally, it has been observed that luxury and performance-oriented brands tend to exhibit higher emission levels.

Graphs and tables have played a pivotal role in the analysis of data. For instance, various types of graphs were utilized to visualize the relationship between vehicle classification and emissions. These graphical representations offer clearer insights into how vehicle characteristics influence emission levels.

The provided visual offers a general summary of dataset statistics and variable types. This visual was created using the pandas-profiling library. Pandas-profiling is a popular Python library used in data analysis that allows for a quick summary of the dataset.


Overview

Overview | Alerts 8 | Reproduction

Dataset statistics

Number of variables	13
Number of observations	22556
Missing cells	0
Missing cells (%)	0.0%
Duplicate rows	1
Duplicate rows (%)	< 0.1%
Total size in memory	8.2 MiB
Average record size in memory	380.5 B

Variable types

Numeric	8
Text	2
Categorical	3

Figure 1 Overview of Data from Pandas Profiling

The pandas-profiling library generates a comprehensive summary report of the dataset. This report allows for quick visualization of missing values, distributions, correlations, and many other important metrics. Here is the interpretation of the information presented in the visual within the context of pandas-profiling:

The dataset consists of a total of 13 variables. It includes 22,556 observations and contains no missing cells, indicating high data quality. Additionally, there is only one duplicate row in the dataset, demonstrating the consistency and integrity of the data. The total memory size of the dataset is 8.2 Megabytes, with each record averaging 380.5 bytes in size.

In terms of variable types, the dataset contains 8 numeric, 2 text, and 3 categorical variables. This diversity allows for the examination of the dataset through various analytical methods. Numeric variables can be used for quantitative analysis, while text and categorical variables can provide qualitative insights. Numeric data facilitates the analysis of the quantity and distribution of emission values, whereas text and categorical data offer insights into qualitative information such as manufacturer details and vehicle classification.

This dataset includes a significant number of observations and has no missing data. The presence of only one duplicate row indicates the high quality of the data. The memory size of the dataset is manageable, making it suitable for various analytical methods. The pandas-profiling report provides a quick summary of the dataset, helping data scientists understand the general characteristics of the data. This is an essential step in data cleaning, preprocessing, and analysis. By using pandas-profiling, the general status, missing values, anomalies, and distributions in the dataset can be quickly identified.

Overall, this type of dataset serves as a suitable foundation for the detailed analysis of emissions in the automobile industry. The combination of numeric, text, and categorical data enables complex data analytics and modelling processes. The use of pandas-profiling facilitates a deep understanding of this dataset and prepares for the necessary steps for future analyses.

4 Results and Discussion

This section is divided into two. These sections are Factors Affecting Emissions and Annual Emission Analysis. Each chapter is examined under four headings. These headings are Vehicle Class, Manufacturer, Engine Size, Fuel Consumption & Fuel Types.

In the first section, the parameters affecting emissions were examined and discussed. These parameters are determined as follows: Vehicle Class, Manufacturer, Engine Size, Fuel Consumption. In addition, the effect of fuel types commonly used in the automotive industry on these parameters was also investigated. These fuel types are regular gasoline, premium gasoline, diesel, ethanol (E85), natural gas. In the second part, the changing emission table between 2000 and 2022 was examined. The emission changes highlighted in the graph were examined and their relationships with the parameters were examined.

4.1 Factors Affecting Emissions

Emission is the amount of pollutant released into the atmosphere over a certain period of time. These pollutants include CO_2, CO, NO_x, SO_2, VOCs (volatile organic compounds), and PM (particulate matter). Emissions are often calculated using emission factors. These factors refer to the number of pollutants emitted per a given unit of activity. Emissions depend on a variety of factors, including industrial activities, transportation, energy consumption and government regulations. The automotive industry is a major source of emissions. Vehicle emissions vary depending on factors such as fuel type, vehicle type, engine technology, driving conditions and vehicle age. Accurate calculation and reduction of emissions is critical for environmental protection and improving human health. Therefore, it is necessary to make emission calculations and take measures based on these calculations.

In this section, the relevant parameters affecting emissions are investigated and the calculations in the existing data set are explained.

4.1.1 Vehicle Classification

Vehicle classification is one of the important parameters for emissions control because different vehicle types and sizes produce different amounts of emissions. Large vehicles generally produce more emissions, while small vehicles produce fewer emissions. This classification ensures the correct application of emission regulations and standards. Additionally, air quality improvement policies and incentives are determined by vehicle class and size, thus achieving more effective results. In short, vehicle classification is a critical parameter for properly managing and reducing emissions.

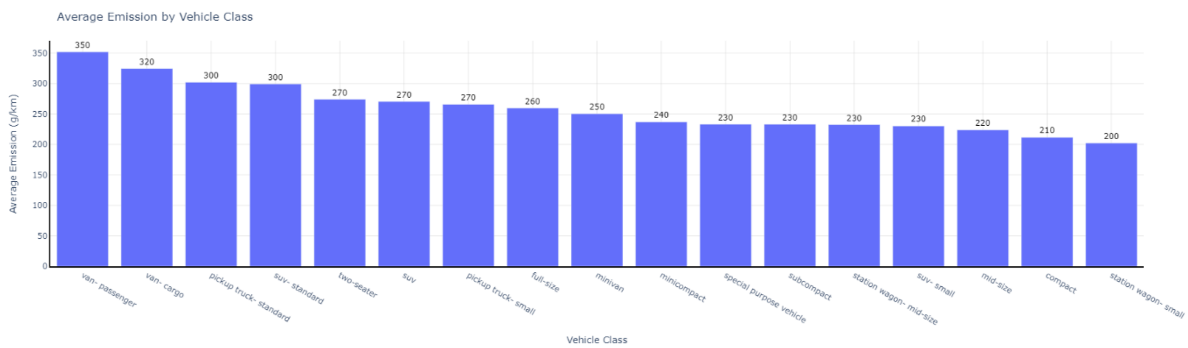


Figure 2 Average Emission by Vehicle Class

Figure 2 illustrates the average emission values of different vehicle classes, highlighting how vehicle size and weight affect emissions.

- Highest emission values: Passenger vans, with an average emission of about 350 g/km.
- High emission values: Cargo vans, passenger vans, standard SUVs, standard pickup trucks, and SUVs; these classes have average emissions ranging from 270 to 300 g/km.
- Moderate emission values: Two-seater vehicles, small pickups, and mid-size station wagons; these classes have average emissions between 260 and 270 g/km.
- Lower emission values: Minivans, mini compacts, and super compacts.
- Lowest emission values: Compact, mid-size, and small station wagons, with average emissions below 200 g/km.

This data indicates that generally, larger and heavier vehicles have higher emissions, while smaller and lighter vehicles have lower emissions.

4.1.2 Manufacturer

The manufacturer is known as another determining parameter. Emission differences from the manufacturer occur depending on various factors such as the type of raw materials and energy used, production technologies, process efficiency, waste management and local environmental regulations. These differences are of great importance for environmental sustainability, legal compliance and cost management. If a manufacturer uses more efficient technologies or turns to renewable energy sources, it can reduce its emissions. Likewise, manufacturers operating under strict environmental regulations may have lower emissions. Understanding these differences is critical to developing more effective emissions reduction strategies.

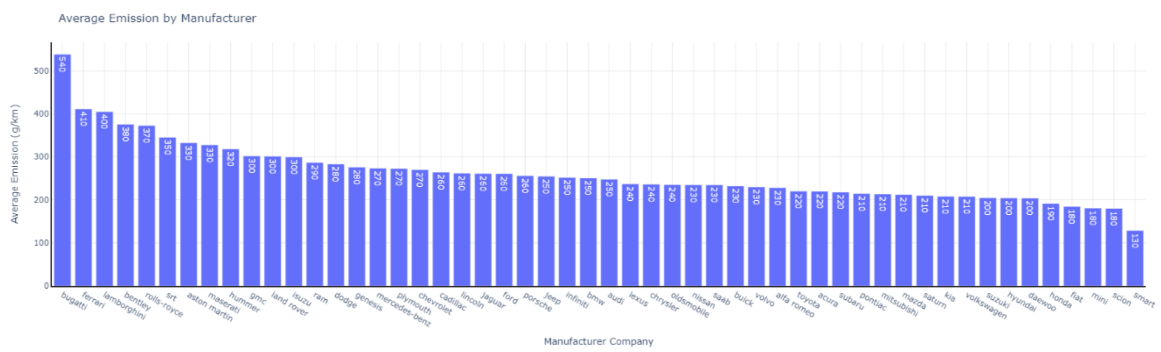


Figure 3 Average Emission by Manufacturer

Emission values are compared according to the brands in the data set. Figure 3 shows the average emissions of various car manufacturers. The brand with the highest emissions in the chart is Bugatti, with an average emission value of over 500 g/km. Following this, luxurious and performance-oriented brands such as Ferrari, Lamborghini, Bentley, Rolls-Royce and Porsche also have high emission values. Brands with mid-range emissions include more common manufacturers such as Chevrolet, Cadillac, BMW, Audi and Ford. Brands such as Toyota, Honda, Hyundai, Volkswagen and Subaru attract attention with their lower emission values. The brand with the lowest emission value was Smart, with an average value below 150 g/km. Figure 3 shows that high-performance vehicles generally have higher emission values, while more economical and environmentally friendly brands offer lower emission values.

4.1.3 Engine Size

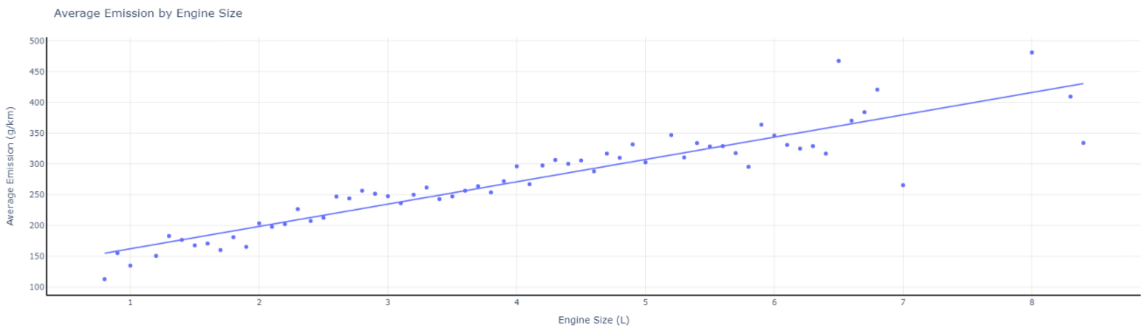


Figure 4 Average Emission by Engine Size

Engine size is known to be another factor affecting emissions. The relationship between engine size and emissions is generally that larger engines are associated with higher emissions, while smaller engines are associated with lower emissions, as can be observed in Figure 4. Larger engines consume more fuel and produce higher power, causing them to produce more emissions. But modern technologies and emissions control systems mean that larger engines can also produce lower emissions, and this relationship can become more complex. There are several reasons why emissions are not completely linear with engine size. While technological advances enable larger engines to be more efficient and less polluting, fuel type also affects emissions.

Depending on this situation, the effect of the fuel types in the data set on vehicles with the same engine volume is shown in Figure 5.

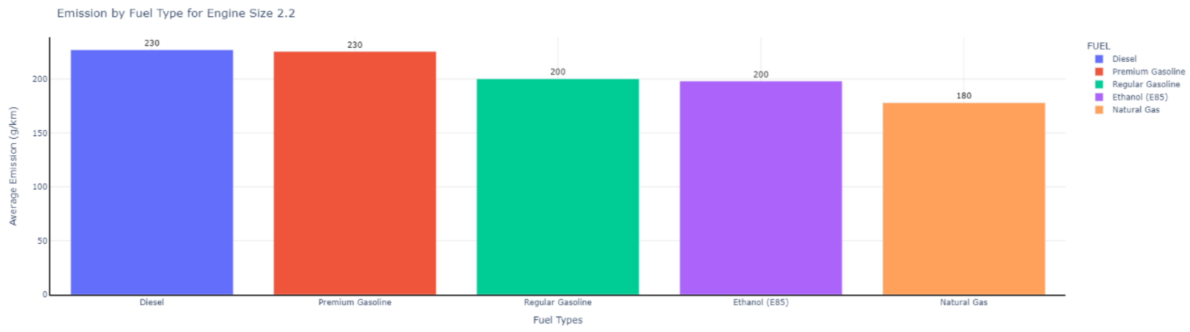


Figure 5 Average Emission by each Fuel Type for Engine Size 2.2 L

When we consider this comparison for vehicles with different fuel types with the same engine size which is 2.2 L, Diesel and Premium Gasoline have the highest emission values. This can be explained by the fact that both fuel types have high energy densities and cause more carbon emissions during combustion. Since Premium Gasoline provides high performance, it leads to more fuel consumption and therefore higher emissions. Regular Gasoline and Ethanol (E85) have lower emission values compared to Diesel and Premium Gasoline.

Regular Gasoline generally has a lower octane level, which can result in less energy density and lower emissions. Ethanol (E85), although considered a biofuel, can release carbon during combustion, but its emissions are still lower than Diesel and Premium Gasoline. Although ethanol has a cleaner combustion process, carbon emissions cannot be eliminated. Natural Gas has the lowest emission value. This shows that natural gas is a cleaner fuel and produces less carbon emissions during combustion. Natural Gas generally consists of methane and produces less carbon dioxide and other harmful emissions during combustion. This feature makes Natural Gas an environmentally friendly option and an attractive alternative for those looking to reduce their carbon footprint. As a result, the graph clearly shows the effects of different fuel types on emission values. While Diesel and Premium Gasoline have the highest emission values, Natural Gas has the lowest emission values. Regular Gasoline and Ethanol (E85) attract attention with their medium emission values. This data provides important clues for those who want to turn to more environmentally friendly fuel options.

The trend-line method was used to observe the difference between the two most effective fuel types, Diesel and Premium gasoline. In Figure 6 you can examine two fuel types with different inclinations. Figure 6 above shows the relationship between engine displacement and average emissions of two different fuel types, specifically diesel (blue) and premium gasoline (red).

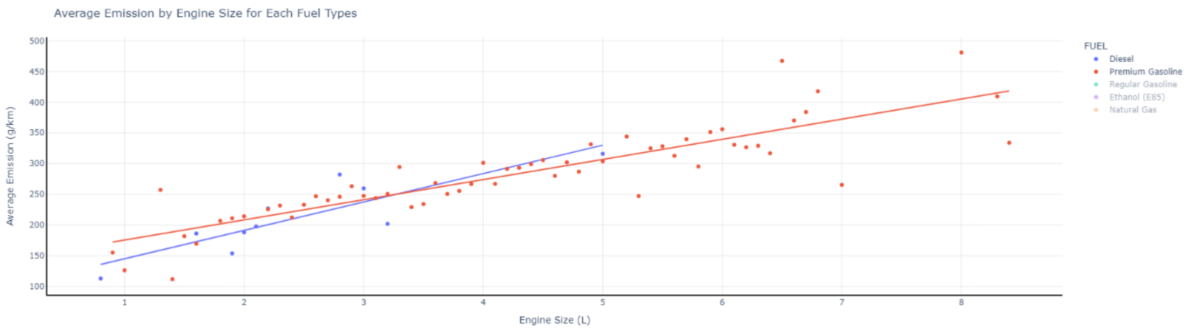


Figure 6 Average Emission by Engine Size for Two Fuel Types

The trend lines, derived using linear regression, illustrate the relationship between engine size and average emissions for both fuel types. The different slopes of these trend lines highlight that the rate of increase in emissions differs between vehicles with the same engine size but different fuel types. Specifically, the steeper slope for diesel indicates that as engine size increases, emissions from diesel vehicles rise more sharply compared to those from premium gasoline vehicles. This suggests that diesel engines, on average, produce higher emissions than premium gasoline engines as their size increases. In general, it has been observed that emissions increase as engine size increases, but this increase varies depending on fuel type. This analysis underscores the importance of fuel type and engine size in determining vehicle emissions.

In engines with the same engine size but different number of cylinders, the number of cylinders can affect emissions in several ways. Engines with more cylinders generally provide smoother combustion but can produce higher emissions due to more internal friction and surface area. On the other hand, engines with fewer cylinders can operate at higher pressure to optimize combustion efficiency, leading to lower emissions.

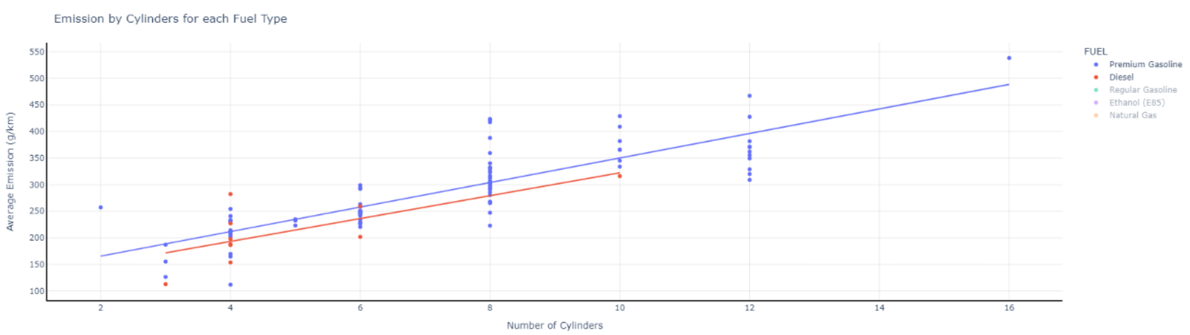


Figure 7 Average Emission by Number of Cylinders for Two Fuel Types

Figure 7 illustrates the impact of the number of cylinders on average emissions (g/km) for different fuel types. Focusing on Diesel and Premium Gasoline, we observe distinct trends and slopes in their respective trend lines. Diesel vehicles are represented by red dots and a red trend line, which has a more horizontal slope. This indicates that the increase in the number of cylinders has a relatively smaller effect on emissions for Diesel vehicles. Due to the generally higher efficiency of Diesel engines, emissions increase at a slower rate as the number of cylinders rises. On the other hand, Premium Gasoline vehicles are shown with blue dots and a blue trend line, which has a steeper slope. This suggests that the increase in the number of cylinders significantly boosts emissions for Premium Gasoline vehicles. More cylinders in Premium Gasoline vehicles typically mean higher performance and power, leading to higher emissions. Overall, while the number of cylinders increases emissions for both fuel types, the increase is more gradual for Diesel vehicles and more pronounced for Premium Gasoline vehicles. As the number of cylinders grows, fuel consumption and consequently emissions also rise. This graph clearly demonstrates the different effects of cylinder count on emissions for Diesel and Premium Gasoline fuel types.

4.1.4 Fuel Consumption & Fuel Types

One of the factors affecting the emissions of vehicles is fuel consumption. There is a direct relationship between fuel consumption, which is an important factor, and emissions.

Carbon emissions occur because of the burning of fossil fuels used in vehicles. As a result of the combustion of fossil fuels, not only CO2 but also various harmful substances such as methane (CH4), nitrogen oxides (NOx), Sulphur dioxide (SO2) and particulate matter (PM) are released. This varies depending on the type of fossil fuels.

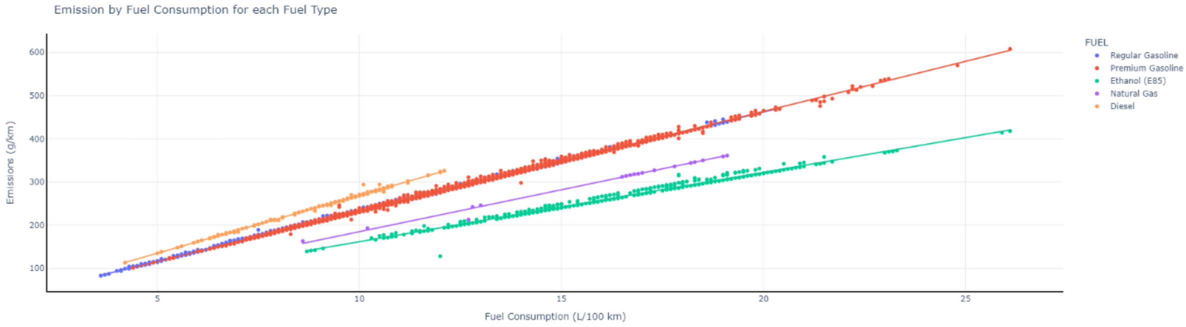


Figure 8 Emission by Fuel Consumption for Each Fuel Type

Regular Gasoline, Premium Gasoline, Diesel, Ethanol (E85), Natural Gas are the fuel types used by the vehicles in the data. And accordingly, the relationship between fuel consumption and emissions for each fuel type was examined. Linear regression formulas were obtained by adding trend-lines to the graphs.

Figure 8 shows the relationship between fuel consumption and emissions according to different fuel types. Each fuel type is represented by a different color, allowing us to distinguish the differences between emissions and fuel consumption. The trendline in the graph illustrates how emissions change as fuel consumption increases for each type of fuel. These trendlines clearly indicate the impact of fuel consumption (L/100 km) on emissions (g/km).

The slopes of the trendlines are expressed numerically in Table 1 below.

Table 1 Slopes of each Fuel Type

Fuel Type	Diesel	Premium Gasoline	Regular Gasoline	Natural Gas	Ethanol (E85)
Trend-Line (L/km)	26.823	23.201	23.057	19.415	16.074

In the case of diesel, emissions increase very rapidly as fuel consumption rises, represented by the highest slope among all fuel types. This indicates that diesel fuel has a significant negative impact on the environment. For premium gasoline, emissions also increase rapidly with higher fuel consumption, showing the second steepest slope after diesel. This means that premium gasoline also has a considerable environmental impact. Regular gasoline shows a more moderate slope, meaning that emissions increase at a slower rate as fuel consumption rises. This suggests that regular gasoline has a lower environmental impact compared to premium gasoline. Natural gas, depicted by a lower slope, maintains emissions at lower levels as fuel consumption increases, indicating it is a more sustainable option. Ethanol (E85) has the lowest slope, meaning that its emissions increase the least with higher fuel consumption. This shows that ethanol is the most environmentally friendly fuel among those analysed, with the least impact on emissions.

These slopes are crucial for understanding the environmental impacts of different fuel types. Fuels with steeper slopes cause more damage to the environment as fuel consumption increases. Diesel and premium gasoline, with their high slopes, result in higher emissions, while alternative fuels like natural gas and ethanol, with their lower slopes, emerge as more sustainable energy options. This data helps us to consider more sustainable energy choices and to understand their potential to reduce environmental impacts.

Moreover, in Figure 8, the points represent individual observations of fuel consumption and corresponding emissions for different fuel types. The trendline, on the other hand, shows the general trend or average relationship between fuel consumption and emissions for each fuel type. The fact that the points do not perfectly lie on the trendline indicates variability or error in the data, which can be attributed to several factors. Firstly, measurement and calculation errors can result from inaccuracies in measuring fuel consumption or emissions. These statistical errors can cause deviations, which is why the points do not align exactly with the trendline. Additionally, different vehicles using the same type of fuel may have varying efficiencies and emission rates due to differences in engine technology, maintenance levels, and manufacturing standards. Driving conditions are also a significant factor; variations such as city driving, highway driving, speed changes, or stop-and-go traffic can lead to substantial differences in emissions and fuel consumption. Environmental factors like temperature, altitude, and humidity can further impact fuel consumption and emissions. Differences in fuel quality, such as the presence of additives or impurities, can also cause variations in emissions and consumption.

Finally, there is always some degree of natural variability in any data set. The trendline provides a useful summary of the general trend but does not account for all individual variations. This variability or "error" is normal and expected in real-world data. The trendline is a helpful tool for summarizing the overall relationship, but individual points will naturally vary around this line due to the measurement and calculation errors, as well as various external factors. Therefore, when conducting your own calculations, it is essential to consider these measurement and calculation errors.

To avoid a possible calculation error due to the mentioned reasons, the emission value of the Diesel fuel type has been recalculated. The calculation was made via Python.

The calculation process of fuels is generally done as follows.

The emission factor (2.68 kg CO2/liter of diesel) was used to calculate CO2 emissions.

1. Convert Fuel Consumption to Fuel Used Per Kilometer:

$$\text{Fuel Consumption}\left(\frac{\text{L}}{\text{km}}\right) = \frac{\text{Fuel Consumption}\left(\frac{\text{L}}{100\text{km}}\right)}{100} \tag{1}$$

2. Calculate CO2 Emissions:

$$\text{CO2 Emissions}\left(\frac{\text{kg}}{\text{km}}\right) = \text{Fuel Consumption}\left(\frac{\text{L}}{\text{km}}\right) \times 2.68 \tag{2}$$

3. Convert CO2 Emissions to Per 100 km:

$$\text{CO2 Emissions}\left(\frac{\text{kg}}{100\text{ km}}\right) = \text{CO2 Emissions}\left(\frac{\text{kg}}{\text{km}}\right) \times 100 \tag{3}$$

Figure 9 displays the CO2 emission values obtained from calculations on the Y-axis and the emission values from the existing dataset on the X-axis. The points exhibit a regular linear relationship, indicating a significant alignment between the two data sets. This alignment suggests that the calculations are generally consistent with the existing dataset. However, factors such as unit differences and rounding errors might explain the small deviations between the data points.

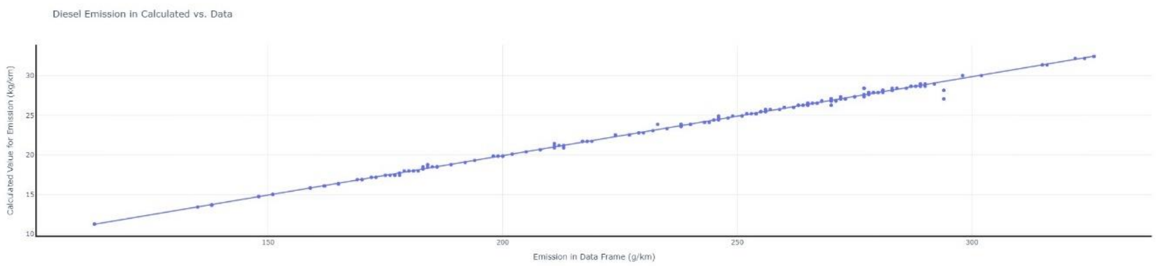


Figure 9 Diesel Emission Comparison

Firstly, it is essential to assess whether differences in units are causing these deviations. If the units used in the dataset and the calculations are different, the values may not be directly comparable. For instance, one set might be measured in kilograms while the other is in grams. In such cases, proper unit conversions must be performed. Secondly, the diesel emission factor used in the calculations may not be a fixed value, contributing to the variations.

The diesel emission factor can vary depending on parameters such as engine efficiency, fuel quality, and driving conditions, leading to minor differences in the calculations. Additionally, rounding errors should also be considered. Using a different number of decimal places or applying various rounding methods can result in slight discrepancies in the results. Ensuring consistency in the number of decimal places used in calculations can minimize these errors.

Finally, although the trend line in Figure 9 generally aligns with the data, some points deviate from this line. These deviations can be attributed to minor variability in the calculations, unit conversions, or rounding errors. To better understand the causes of these deviations and make necessary corrections, a thorough examination of the calculations and dataset is required.

4.2 Yearly Emissions Analysis

The 2000s mark an important era of environmentally friendly technologies and regulations in the automotive industry. During this period, increasing environmental concerns globally led automobile manufacturers to produce lower-emission vehicles and led to the development of various emission standards around the world. Since the early 2000s, the European Union has focused on significantly reducing the number of pollutants that vehicles release into the atmosphere by implementing Euro 4, Euro 5 and most recently Euro 6 emission standards. In Figure 10 below, you can observe the changing average emission amount between 2000 and 2022.

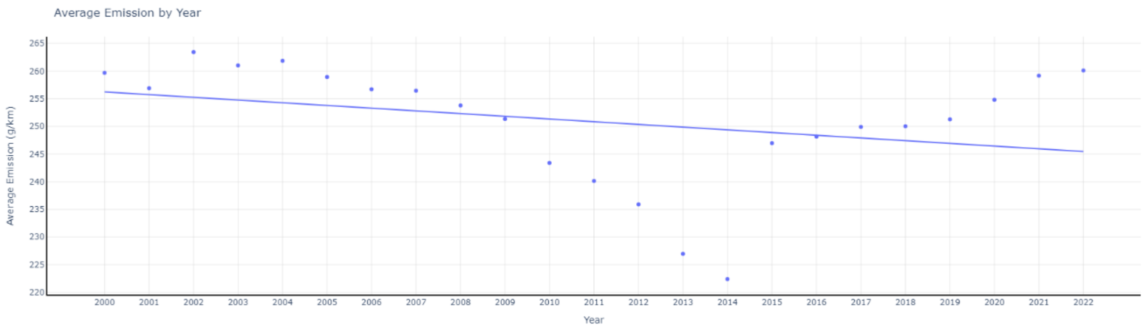


Figure 10 Average Emission by Year

According to Figure 10, the decline between 2009 and 2014 was driven by several factors in the automotive industry. First, the 2008-2009 global financial crisis caused a significant decline in consumer spending and automobile demand. During this period, as people preferred to spend less, automobile sales were also negatively affected. Additionally, during this period, many countries and regions began to impose stricter emissions regulations.

Automobile manufacturers had to make large investments to comply with these new regulations, and the implementation of Euro 5 emission standards contributed to this process. Euro 5 standards included stricter emissions limits and directed the auto industry to produce cleaner and less polluting vehicles. While this increased production costs, it also affected vehicle prices. As a result, there was a decline in the automotive industry between 2009 and 2014 due to the economic crisis, tightening regulations and Euro 5 standards. The sudden rise between 2014 and 2015 was a major turning point in the automotive industry. In 2014, the European Union moved to the implementation of Euro 6 emission standards, but the impact of this transition was contrary to expectations. Although the transition to Euro 6 led automakers to produce cleaner and more environmentally friendly vehicles, it took time to fully implement and comply with these standards. During the same period, the need to conduct driving tests in real-world conditions was also emphasized, which helped emission tests provide more realistic results. However, another important event during this period was Volkswagen's diesel emissions scandal that emerged in 2015. It was revealed that the company manipulated emission tests on its vehicles, raising serious questions about reliability. As a result, there was a period of general distrust and uncertainty in the industry. These events have once again shown that the industry needs to be more diligent about environmental compliance and transparency. What happened during this period once again revealed the difficulties faced by the automotive industry and its importance in environmental sustainability. The period between 2000 and 2022 has been characterized by significant changes and transformations in the automotive industry. During this time, the development of environmentally friendly technologies and tightening of emission regulations led automobile manufacturers to produce cleaner vehicles. In particular, the European Union's adoption and implementation of Euro emission standards has contributed to increased environmental compliance in the automotive industry. However, during this period, factors such as global economic fluctuations, oil price fluctuations and pandemics caused fluctuations in emissions in the automotive industry. For example, the 2008 financial crisis led to declines in auto sales and production, while the COVID-19 pandemic caused a severe pause in the automotive industry and declines in emissions in 2019 and 2020. During this period, regulatory changes, technological advances, economic conditions and global events had significant impacts on emissions in the automotive industry.

In line with the reasons and explanations for this change over the years and especially this increase in 2014 and 2015, it is important to examine the data for the relevant parameters and discuss what effects it has.

4.2.1 Vehicle Class

The issue of whether there has been a major change in the dimensions of vehicles with the developing technology is important to observe the increase in emissions in these years. For this reason, the effect of vehicle classes in 2014 and 2015 on the emission value was examined.

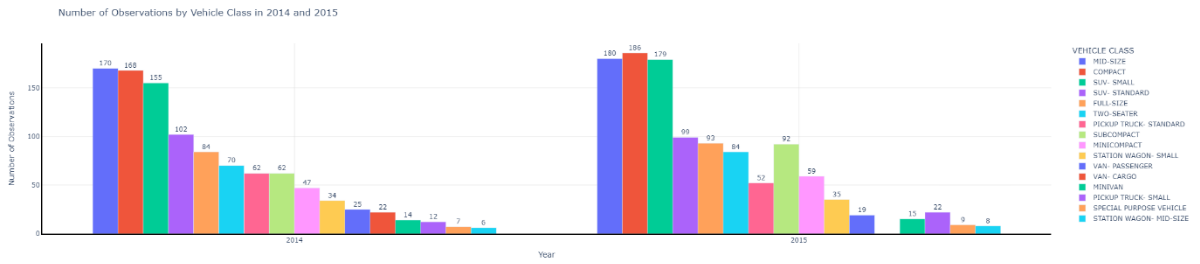


Figure 11 Number of Observations by Vehicle Class in 2014 and 2015

Figure 11 presents a comparative analysis of the number of observations made over two years for different vehicle classes. Compact and mid-size vehicles have the highest number of observations in both years, with an increase seen from 2014 to 2015. Observed vehicle numbers for most vehicle classes show a similar trend in both years, but there are some differences. For example, full-size cars and standard pickup trucks show a significant increase in 2015 compared to 2014, indicating that interest or focus on these vehicle classes increased in the following year.

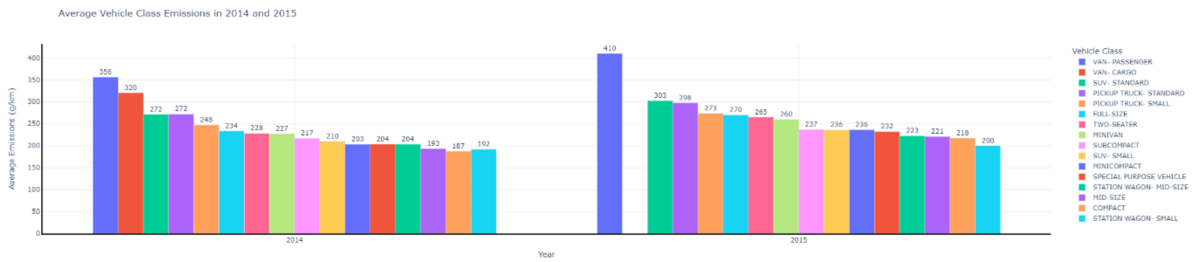


Figure 12 Average Emission by Vehicle Class in 2014 and 2015

Figure 12 shows the average emissions for each vehicle class over two years. It has been observed that emissions for each vehicle class are relatively consistent over the years. Larger vehicle classes such as vans and SUVs typically exhibit higher average emissions than smaller vehicle classes such as subcompacts and mini compacts. This trend is associated with larger engines and higher fuel consumption. The chart shows that there may be intra-year fluctuations in certain classes, but overall emissions patterns remain stable.

4.2.2 Manufacturer

If the impact of brands on emission values is investigated, it may be logical to look at Bugatti values of 2009 and later for Euro5 and later. According to the data, it has been observed that the largest share in emissions originates from Bugatti.

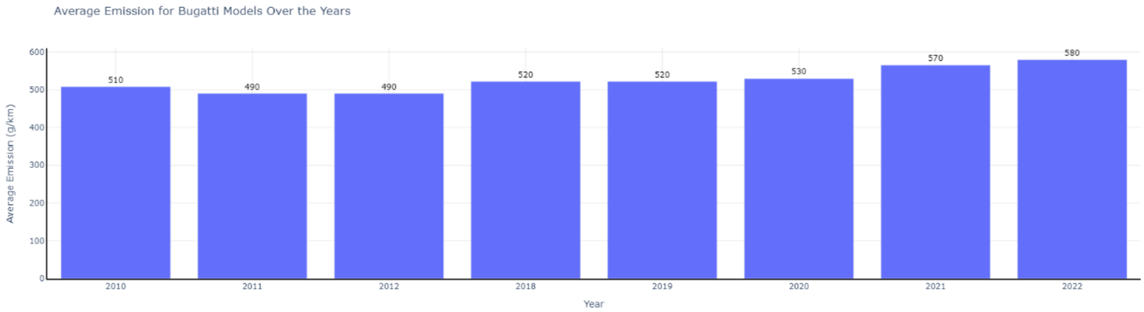


Figure 13 Average Emission for Buggati Model Over the Years

Accordingly, to observe the impact of Bugatti on the decreasing emission values and the sudden increase in emission values as of 2014. For this sudden increase, a visual was created to see whether there was an increase in emission values especially in 2014-2015. production of the brand. Average emission values of Bugatti models are shown by year. Contrary to expectations in Figure 13, a particularly striking period is the non-production period between 2012 and 2018. There is no data in the chart between these years that could mean that no new Bugatti model was released or that no new Bugatti model was released during that period. There has been no significant change in emissions of current models. This can be considered as a break or a transition period during which Bugatti takes a break to develop new models. Developing new technologies and designs in the automotive industry can take time, and such a break may be appropriate. In addition, this period may be a transition period in terms of discontinuing old models and developing new models. The company may have worked on new engine technologies and environmentally friendly features during this process. This non-production period can also be a strategic decision for the company. They may have waited a certain period for new models to be released or evaluated the competitive conditions in the market. As a result, the non-production period between 2012 and 2018 can be considered as an innovative and strategic preparation period for Bugatti's future models. The increase in emissions in models released after this period may be a result of these preparations and new technologies. Although more information is needed to understand the details of this process, this gap in the chart points to periods of innovation and transition that are sometimes necessary in the automotive industry.

4.2.3 Engine Size

New sustainability laws implemented over the years have caused changes in the design of vehicles and especially their engines. Examining the changes in these emission values over the years can be considered as one of the reasons for the sudden increase in emissions. Accordingly, what is expected is that the engine size will increase significantly during that period.

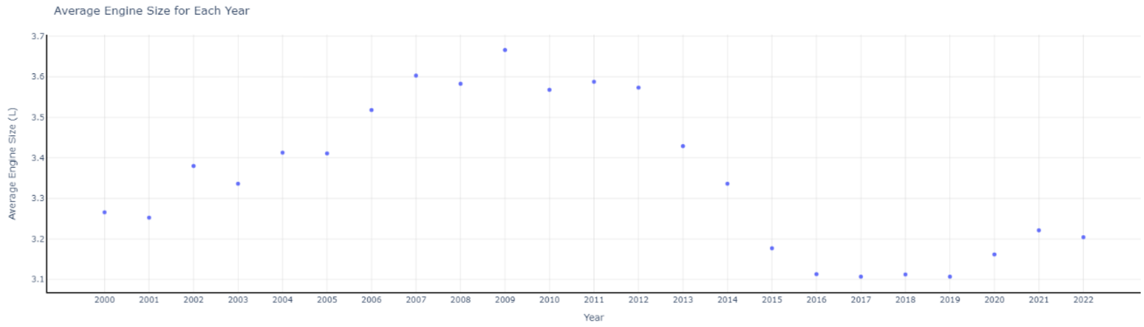


Figure 14 Average Engine Size for Each Year

The provided Figure 14 depicts the average engine size (in liters) for each year from 2000 to 2022. Here's an analysis of the trends observed, with a particular focus on the years 2014 and 2015:

From 2000 to 2011, there is a general upward trend in the average engine size, peaking around 2009-2010. However, after 2011, there is a noticeable decline in engine size, reaching its lowest point in 2016. Post-2016, the average engine size appears to remain relatively stable with slight fluctuations.

2014-2015: The graph shows a significant decline in average engine size during these years. This period marks one of the steepest drops, indicating a major shift in the automotive market. Possible reasons for this decline include the implementation of stringent environmental policies, increased popularity of smaller, more efficient engines, and a push from manufacturers to meet new fuel economy standards.

In this case, it may be thought that the engine volume does not have a big effect or that small engines consume more fuel and therefore there are fluctuations in emission values. However, when we look at the total graph, it is predicted that small engines are more efficient and therefore the volume is getting smaller.

4.2.4 Fuel Consumption & Fuel Types

Fuel consumption is a factor that directly affects emissions, and knowing the amount used over the years is especially important for this increase and decrease. The meaning of this increase, especially in 2014 and 2015, in terms of fuel consumption is a matter of curiosity.

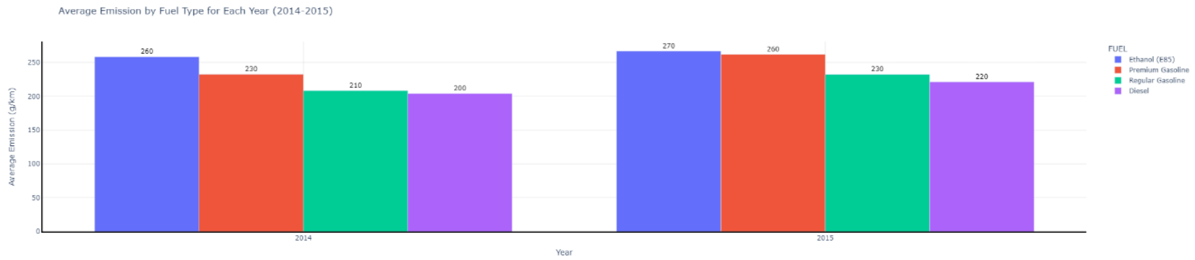


Figure 15 Average Emission by Fuel Types for 2014 and 2015

Figure 15 displays the average emission vs the lowest emission values in both years. Diesel engines generally have higher fuel efficiency values for different fuel types in 2014 and 2015. In both years, Ethanol (E85) has the highest emission values. This can be explained by the fact that Ethanol (E85) produces more carbon emissions during combustion. Ethanol typically has a higher-octane rating, which can enhance engine performance but also increase emissions. ha, resulting in lower emissions. However, there is an observed increase in Diesel emissions in 2015. This increase could be attributed to higher usage of Diesel vehicles, or these vehicles being operated under heavier loads. There is also an increase in the emission values for Premium Gasoline and Regular Gasoline in 2015. This rise could be due to increased vehicle usage or changes in fuel quality. Premium Gasoline usually offers higher performance, which can lead to higher emissions. Regular Gasoline, while remaining consistent, might show increased emissions due to higher usage. Overall, there is an observed increase in emissions across all fuel types in 2015. This can be explained by factors such as the increase in the number of vehicles, higher traffic density, or changes in fuel quality. These factors can lead to increased fuel consumption and consequently higher emissions. However, there is no significant change in average emissions between these two years. This indicates that while there are fluctuations in emissions for different fuel types, the overall average emissions remained relatively stable. Therefore, it is important to consider the change in the number of vehicles over the years and the prevalence of different fuel types in the vehicle population. In summary, the emission values observed in both years are directly related to the combustion properties and usage conditions of the fuel types, but there is no serious difference in average emissions between the two years.

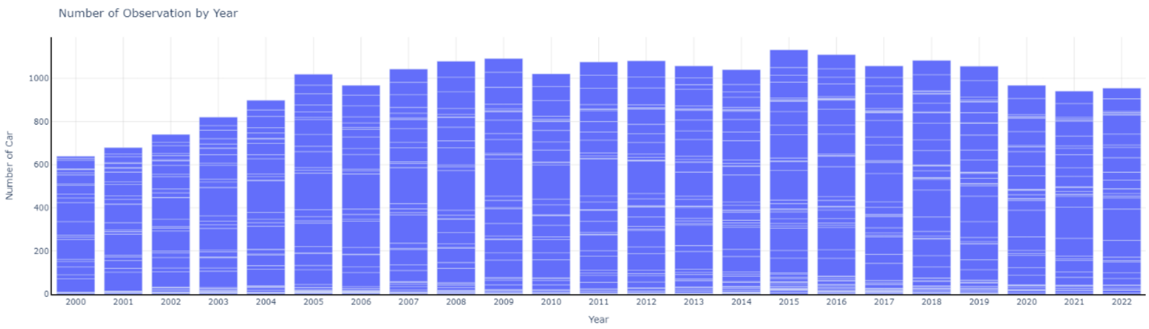


Figure 16 Number of Observation by Year

Figure 16 displays the number of car observations by year from 2000 to 2022. When looking specifically at 2014 and 2015, the number of observations is quite similar, around 1000 cars each year. These years are part of a period (2005-2016) characterized by generally high observation counts.

- Consistent Observations: There is no significant difference in the number of car observations between 2014 and 2015, indicating that the data collected in these years is comparable in volume.
- High Observation Period: Both years fall within a period where vehicle observations were consistently high.
- Emission Data: The changes in emission values observed in the Figure 16 for 2014 and 2015 are not due to differences in the number of vehicles observed but likely due to other factors such as vehicle usage and fuel types.

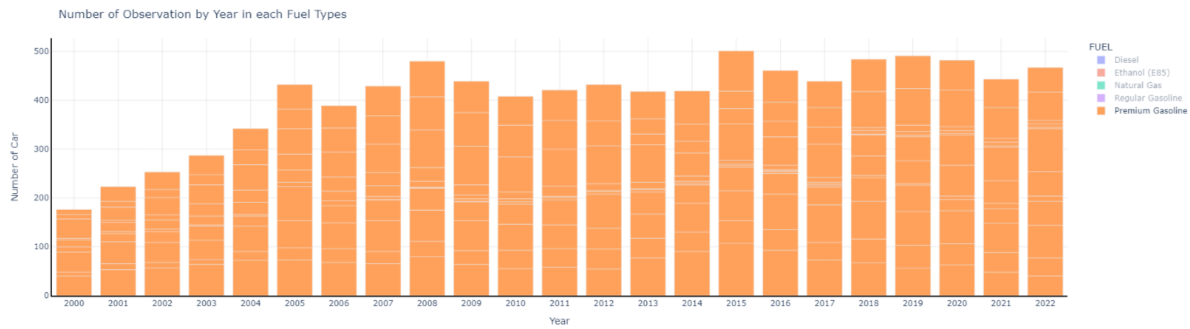


Figure 17 Number of Observation by Year in Premium Gasoline

Figure 17 shows the number of vehicle observations using Premium Gasoline by year, from 2000 to 2022.

When we look specifically at the years 2014 and 2015, we see that the number of observations was around 450 vehicles in both years. These years are part of a period (2005-2016) characterized by generally high numbers of observations.

Consistent Observations: There is no significant difference in the number of vehicle observations between 2014 and 2015, indicating that the data collected in these years is comparable in volume.

High Observation Period: Both years coincide with a period when vehicle observations are consistently high.

Emissions Data: Changes in emissions values observed in Figure 17 for 2014 and 2015 are likely due to other factors such as vehicle usage and fuel types, and not to differences in the number of vehicles observed.

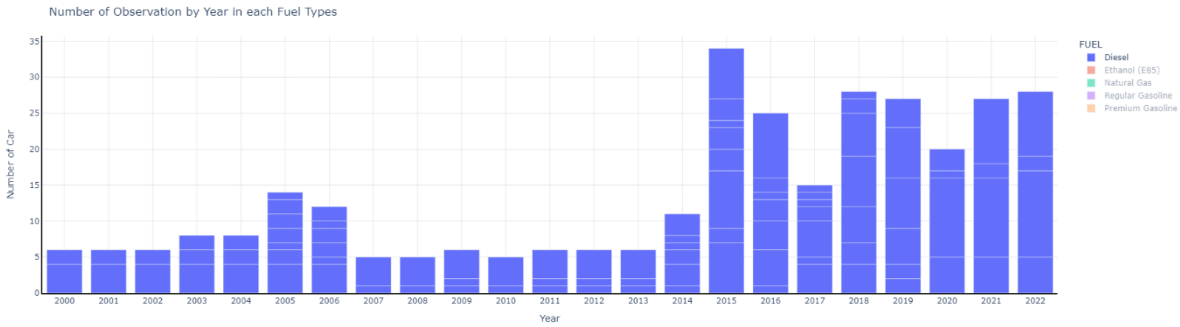


Figure 18 Number of Observation by Year in Premium Diesel

Figure 18 shows the number of vehicle observations using Diesel by year, from 2000 to 2022. When we look specifically at the years 2014 and 2015, we see that the number of observations was approximately 20 vehicles in 2014 and 30 vehicles in 2015. These years are part of a period (2014-2022) characterized by generally high numbers of observations.

Consistent Observations: There is a significant increase in the number of vehicle observations between 2014 and 2015, indicating that more data was collected in 2015 than in 2014.

High Observation Period: Both years coincide with a period when vehicle observations are consistently high.

Emissions Data: Changes in emissions values observed in Figure 18 for 2014 and 2015 may be due to differences in the number of vehicles observed and may also be influenced by other factors such as vehicle usage and fuel types.

Due to this interesting change in diesel, it becomes an important point to re-examine the Volkswagen Diesel Scandal in 2015.

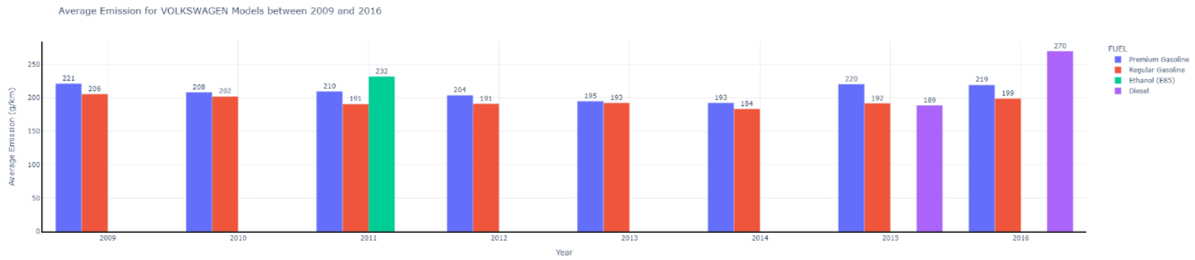


Figure 19 Average Emission for Volkswagen Models between 2009 and 2016

Figure 19 shows the average emission values of Volkswagen models according to different fuel types between 2009 and 2016. Emission values of Premium Gasoline and Regular Gasoline fuel types have generally remained stable throughout these years, indicating that the engine technology used in these fuel types is relatively stable. However, there was a significant increase in the emission values of the Diesel fuel type in 2016. This increase in diesel emissions in 2016 may be due to vehicles operating at higher loads that year or changes in diesel engine technology, or it may be a result of the Volkswagen diesel scandal that emerged in 2015. The scandal revealed that Volkswagen manipulated emissions tests for its diesel-powered vehicles, which may have led to more accurate reporting of diesel emissions in 2016. Following the scandal, regulatory authorities began to impose more stringent emission tests, which caused the emission values of diesel vehicles to increase significantly in 2016.

Overall, the chart shows that there are some differences in emissions between fuel types and years, but that emissions remain relatively constant for Premium Gasoline and Regular Gasoline fuel types. The Volkswagen diesel scandal has caused a noticeable increase in diesel emissions and led to a re-evaluation of the environmental impact of diesel-powered vehicles. This situation not only shook the confidence in Volkswagen's diesel vehicles, but also caused regulatory authorities to impose stricter emission controls. The scandal has led to more accurate reporting of emissions from diesel vehicles and more careful scrutiny of the environmental impacts of engine technologies.

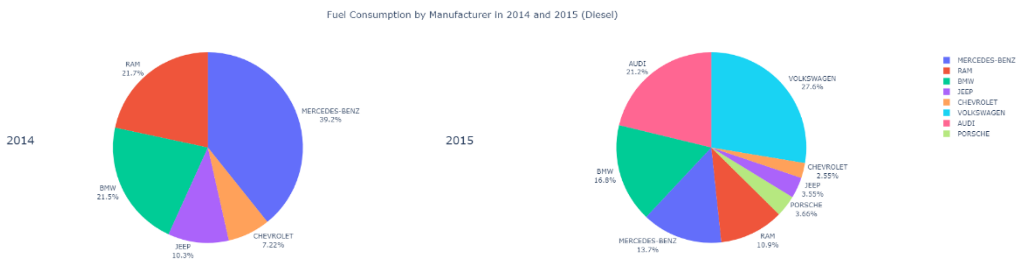


Figure 20 Diesel Fuel Consumption by Manufacturer in 2014 and 2015

Figure 20 shows diesel fuel consumption for just a few major manufacturers in both years:

Year 2014: Mercedes-Benz, RAM, BMW, Jeep and Chevrolet are among the manufacturers with the highest diesel fuel consumption. While Mercedes-Benz has the highest share with 39.2%, RAM has 21.7%, BMW 21.5%, Jeep 10.3% and Chevrolet 7.22%.

Year 2015: In this year, Volkswagen has the highest share with 27.4%. It is followed by Audi with 21.4%, BMW with 16.6%, Mercedes-Benz with 13.7%, RAM with 10.9% and Chevrolet with 2.95%.

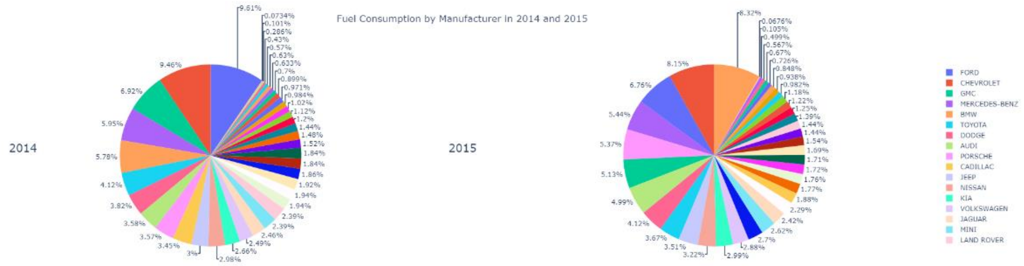


Figure 21 Total Fuel Consumption by Manufacturer in 2014 and 2015

Figure 21 has been expanded to include more manufacturers and shows each manufacturer's share of total fuel consumption in more detail:

Year 2014: Ford, Chevrolet, GMC, Mercedes-Benz, Toyota, Dodge, Audi, Porsche, Cadillac, Acura and many other manufacturers are included. Here, each manufacturer's share is less than 10%, with Ford having the highest share at 9.46%, Chevrolet at 6.92%, and GMC at 5.95%.

Year 2015: There are many manufacturers this year as well, and the manufacturer with the highest share is Ford with 8.13%. It is followed by Chevrolet with 6.76%, GMC with 5.46%, Mercedes-Benz with 5.37% and BMW with 4.99% total fuel consumption.

All in all, despite Volkswagen having the highest share of diesel fuel consumption in 2015 (27.4%), its overall share of total fuel consumption is not significant when compared to other manufacturers like Ford and Chevrolet.

Year 2014:

- Volkswagen is not listed among the top manufacturers for total fuel consumption.
- Ford holds the highest share at 9.46%.

Year 2015:

- Ford remains the leader in total fuel consumption with 8.13%.
- Volkswagen’s total fuel consumption share is negligible and not listed among the top manufacturers.

This highlights that while Volkswagen may have a high share of diesel consumption, it does not translate to a significant impact on the overall fuel consumption landscape, dominated by manufacturers like Ford, Chevrolet, and GMC.

4.3 Impact of Volkswagen's Diesel Scandal and Other Potential Problems

When looking at these graphic sets, it is possible to see the effects of Volkswagen's diesel scandal in 2015. It is observed that there is a significant increase in Volkswagen's fuel consumption rate, and it has a higher share than other major manufacturers. However, it is difficult to make definitive judgments about other potential scandals or secret situations based on these graphs. However, some notable points in the charts are:

Volkswagen's Remarkable Increase: The significant increase in Volkswagen's share of diesel fuel consumption in 2015 compared to 2014 may indicate the impact of the diesel scandal on fuel consumption. This may suggest that Volkswagen may have sold more diesel vehicles through emissions cheating.

Shares of Other Manufacturers: The fact that there are more manufacturers in the second set and the fuel consumption shares of each of these manufacturers are low may suggest that there may be a general fraud or hidden problem in the industry. However, more detailed data and analysis are needed to confirm this.

Shares of Mercedes-Benz and BMW: In both years, the high shares of manufacturers such as Mercedes-Benz and BMW attract attention. Investigating whether these manufacturers resort to similar tricks may be useful in understanding the prevalence of the diesel scandal throughout the industry.

Different Manufacturers and Changes Between Years: In the second set, more manufacturers and changes between years attract attention. Investigating the reasons for these changes can help identify possible fraud or manipulation in the industry.

As a result, Volkswagen's diesel scandal is having a significant impact on the charts. However, more data and analysis are needed to make a definitive judgment about other manufacturers. Changes in the shares of other producers can be a starting point to investigate whether there are other potential problems in the industry. Therefore, it is important to conduct comprehensive research and examine the practices of other manufacturers in the industry to better understand the effects of the diesel scandal.

5 Conclusion

5.1 Summary of the Research

This research examined emissions in the automotive industry, particularly those that contribute to global warming. Although many studies have been conducted highlighting various parameters, this study focused on a data set covering vehicle models produced between 2000 and 2022. This study analyses emissions in vehicles and the affecting parameters using Python visualization.

The results of the study show an in-depth examination of various factors affecting automobile emissions. Variables such as vehicle classification, manufacturer, engine displacement and fuel consumption have significant effects on emissions, and each of these has been analyzed in detail. In terms of vehicle classification, larger and heavier vehicles have generally been found to have higher emissions. When examined on a manufacturer basis, it is seen that luxury, and performance-oriented brands generally have higher emission values. Technical characteristics such as engine displacement and fuel consumption also have a decisive impact on emissions. While larger engines are generally associated with higher emissions, the use of different fuel types also causes changes in emission values. When emission trends are examined by year, it is seen that environmentally friendly technologies and regulations have increased significantly in the automotive industry since the early 2000s. However, factors such as economic fluctuations, changes in oil prices and pandemics have caused fluctuations in emissions in industry. However, with the introduction of Euro 5 emissions standards in 2009, there has been a marked shift in emissions regulations in the automotive industry. These standards have encouraged the production of cleaner and more environmentally friendly vehicles and forced automakers to comply with stricter emissions limits.

However, in this case, when looking at the big picture, it should not only be limited to a specific brand, but also other large companies should be researched based on numerical data. Such a big change should not be limited to a single brand.

Overall, the study highlights the complexity and importance of efforts to reduce emissions in the automotive industry and encourages the consideration of more sustainable energy options. It aims to encourage the automotive industry to emit less emissions within the scope of the Green Deal in the next 20 years and to ensure that the events that caused emission changes are not repeated by examining the past 20 years in more detail.

5.2 Limitations of the Study

In this research I could work on its different types of fuel such as electricity or hydrogen which are environmentally friendly. However, in my data set there are only fuels that have higher emissions than the others.

5.3 Further Research Possibilities

There are several potential ways to expand the study. First, it would be important to examine the emission profiles of alternative fuel types, especially other environmentally friendly options such as electric vehicles or hydrogen fuel cell vehicles. Additionally, a study could be conducted to evaluate the effects of current and future emissions regulations. A detailed analysis can be conducted to understand how technological advances will affect emissions trends. Institutional and social awareness activities can increase awareness about switching to environmentally friendly vehicles. Finally, validating the real-world emissions performance of specific vehicle models through laboratory or field studies could further strengthen the study's findings and contribute to more definitive conclusions.

Furthermore, it is crucial to investigate whether similar diesel emissions scandals could be present in other firms. Expanding the study to include a comprehensive examination of other major manufacturers would help to uncover any potential discrepancies or manipulations in emissions data across the industry. By doing so, we can ensure a more transparent and reliable understanding of the true environmental impact of diesel vehicles. This broader investigation would not only provide a clearer picture of the industry's overall emissions practices but also drive the adoption of more stringent regulations and cleaner technologies across the board.

REFERENCES

NASA. *What is Climate Change? | Facts – Climate Change: Vital Signs of the Planet*, 2024, from https://climate.nasa.gov/

Yapı Kredi. *What are Greenhouse Gases? What You Need to Know About Greenhouse Gases*. 2024, from https://www.yapikredi.com.tr/blog/surdurulebilirlik/cevre/detay/sera-gazlari-nelerdir-sera-gazlari-hakkinda-bilmeniz-gerekenler

Waternet. *What is Carbon Footprint?* 2024, from https://www.waternet.com.tr/faydali-bilgiler/karbon-ayak-izi-nedir/

The Nature Conservancy. *What is your carbon footprint? | Carbon Footprint Calculator*. 2024, from https://www.nature.org/en-us/get-involved/how-to-help/carbon-footprint-calculator/

4EN. (n.d.). *What is Carbon Footprint and How to Reduce It?*.2024, from https://4en.com.tr/karbon-ayak-izi-nedir-ve-nasil-azaltilir

Mattila, T., Kujanpaa, M., Dahlbo, H., Soukka, R., & Myllymaa, T. (2011). Uncertainty and sensitivity in the carbon footprint of shopping bags. *Journal of Industrial Ecology*, *15*(2), 217-227.

Jones, C. M., & Kammen, D. M. (2011). Quantifying carbon footprint reduction opportunities for U.S. households and communities. *Environmental Science and Technology*, *45*(9), 4088-4095.

Stack Overflow user. (2023, July 9). *Import Vaex error: PydanticImportError: BaseSettings has been moved to the pydantic-settings package*. Stack Overflow. 2024, from https://stackoverflow.com/questions/76664231/import-vaex-error-pydanticimporterror-basesettings-has-been-moved-to-the-py

CodeAI. (2023, March 25). *Data Visualization with Plotly*. 2024, from https://medium.com/codeai/plotly-ile-veri-g%C3%B6rselle%C5%9Ftirme-64541fbdd81b

APPENDIX

```
# -*- coding: utf-8 -*-

"""FINAL PROJECT FOR RGPP.ipynb adlı not defterinin kopyası

Automatically generated by Colab.

Original file is located at

    https://colab.research.google.com/drive/1fnGyzn2jYVhhhtSFNSZmaE9HGGMAXsX-

"""

!pip install pydantic-settings



!pip install https://github.com/pandas-profiling/pandas-profiling/archive/master.zip



import seaborn as sns



from pydantic_settings import BaseSettings

from pandas_profiling import ProfileReport



import plotly.graph_objects as go

import pandas as pd

import plotly.express as px

from plotnine import *
```

```
from plotly.subplots import make_subplots

from tabulate import tabulate



# Fuel Consumption dataset

df=
pd.read_csv('https://raw.githubusercontent.com/cansukarabulut/carbonemission/main/Fue
l_Consumption.csv')

df['VEHICLE CLASS']= df['VEHICLE CLASS'].str.replace(':', '-').str.replace(' -','-')



profile = ProfileReport(df, explorative=True, dark_mode=True)

profile.to_file('output.html')



df



df['VEHICLE CLASS']= df['VEHICLE CLASS'].str.replace(':', '-').str.replace(' -','-')

number_of_observation_by_year= df.groupby(['YEAR','FUEL','VEHICLE CLASS'],
as_index = False).agg(number_of_observation = ('YEAR', 'count'))

number_of_observation_by_year



p= px.bar(number_of_observation_by_year, x= 'YEAR', y= 'number_of_observation')

p.update_layout( title =  'Number of Observation by Year', xaxis_title= 'Year',
yaxis_title='Number of Car', plot_bgcolor = 'White' )
```

```
p.update_xaxes(type='category')

p.update_xaxes(showline=True, linecolor='black', linewidth=2, gridcolor='lightgray',
gridwidth=0.5)

p.update_yaxes(showline=True, linecolor='black', linewidth=2, gridcolor='lightgray',
gridwidth=0.5)



number_of_observation_by_year_replaced = number_of_observation_by_year.replace(

    {'FUEL': {'X': 'Regular Gasoline', 'Z': 'Premium Gasoline', 'D': 'Diesel', 'E': 'Ethanol
(E85)', 'N': 'Natural Gas'}}

)



# Create the bar chart using the replaced DataFrame

p = px.bar(number_of_observation_by_year_replaced, x='YEAR',
y='number_of_observation', color='FUEL')



# Update the layout of the chart

p.update_layout(

    title='Number of Observation by Year in each Fuel Types',

    xaxis_title='Year',

    yaxis_title='Number of Car',

    plot_bgcolor='White'

)
```

```
# Set the x-axis type to category

p.update_xaxes(type='category')

p.update_xaxes(showline=True, linecolor='black', linewidth=2, gridcolor='lightgray', gridwidth=0.5)

p.update_yaxes(showline=True, linecolor='black', linewidth=2, gridcolor='lightgray', gridwidth=0.5)

# Show the plot

p.show()



df



p=px.scatter(df.replace({'FUEL': {'X': 'Regular Gasoline', 'Z': 'Premium Gasoline', 'D': 'Diesel', 'E': 'Ethanol (E85)', 'N':'Natural Gas'}}), x='COMB (L/100 km)', y='EMISSIONS', trendline='ols', color='FUEL')



p.update_layout(

    title='Emission by Fuel Consumption for each Fuel Type',

    xaxis_title='Fuel Consumption (L/100 km)',

    yaxis_title='Emissions (g/km)',

    plot_bgcolor='White'

)
```

```
p.update_xaxes(showline=True, linecolor='black', linewidth=2, gridcolor='lightgray', gridwidth=0.5)

p.update_yaxes(showline=True, linecolor='black', linewidth=2, gridcolor='lightgray', gridwidth=0.5)



filter = df['FUEL'] == 'D' #diesel

filtered_df = df.loc[filter].copy()



# Diesel emission factor (kg CO2/litre)

diesel_emission_factor = 2.68  # kg CO2/gallon / litre/gallon



# Multiply the values in litre units corresponding to the filtered entries with the diesel emission factor

filtered_df.loc[:, 'CO2_Emission_kg/km'] = (filtered_df['COMB (L/100 km)'] /100) * diesel_emission_factor*100



print(tabulate(filtered_df, headers='keys', tablefmt='grid'))



scatter_df = filtered_df[['EMISSIONS', 'CO2_Emission_kg/km']]
```

```
p = px.scatter(scatter_df, x='EMISSIONS', y='CO2_Emission_kg/km', title='Diesel Emission in Calculated vs. Data', trendline= 'ols')

p.update_layout(title= 'Diesel Emission in Calculated vs. Data',xaxis_title='Emission in Data Frame (g/km)', yaxis_title='Calculated Value for Emission (kg/km)', plot_bgcolor= 'white')

p.update_xaxes(showline=True, linecolor='black', linewidth=2, gridcolor='lightgray', gridwidth=0.5)

p.update_yaxes(showline=True, linecolor='black', linewidth=2, gridcolor='lightgray', gridwidth=0.5)


engine_size_emission= df.groupby('ENGINE SIZE', as_index = False).agg(average_emission = ('EMISSIONS', 'mean'))

p= px.scatter(engine_size_emission, x= 'ENGINE SIZE', y= 'average_emission', trendline= 'ols')

p.update_layout( title = 'Average Emission by Engine Size', xaxis_title= 'Engine Size (L)', yaxis_title='Average Emission (g/km)', plot_bgcolor = 'White' )

p.update_xaxes(showline=True, linecolor='black', linewidth=2, gridcolor='lightgray', gridwidth=0.5)

p.update_yaxes(showline=True, linecolor='black', linewidth=2, gridcolor='lightgray', gridwidth=0.5)


engine_size_emission_by_type= df.groupby(['ENGINE SIZE', 'FUEL'] , as_index = False).agg(average_emission = ('EMISSIONS', 'mean'))


engine_size_emission_by_type
```

```
engine_size_emission_by_type.replace({'FUEL': {'X': 'Regular Gasoline', 'Z': 'Premium Gasoline', 'D': 'Diesel', 'E': 'Ethanol (E85)', 'N':'Natural Gas'}}, inplace=True)

p= px.scatter(engine_size_emission_by_type, x= 'ENGINE SIZE', y= 'average_emission', color= 'FUEL' , trendline= 'ols')

p.update_layout( title =  'Average Emission by Engine Size for Each Fuel Types', xaxis_title= 'Engine Size (L)', yaxis_title='Average Emission (g/km)', plot_bgcolor = 'White' )

p.update_xaxes(showline=True, linecolor='black', linewidth=2, gridcolor='lightgray', gridwidth=0.5)

p.update_yaxes(showline=True, linecolor='black', linewidth=2, gridcolor='lightgray', gridwidth=0.5)


cylinders_by_type= df.groupby(['CYLINDERS', 'ENGINE SIZE', 'FUEL'] , as_index = False).agg(average_emission = ('EMISSIONS', 'mean'))

cylinders_by_type


cylinders_by_type.replace({'FUEL': {'X': 'Regular Gasoline', 'Z': 'Premium Gasoline', 'D': 'Diesel', 'E': 'Ethanol (E85)', 'N':'Natural Gas'}}, inplace=True)

p= px.scatter(cylinders_by_type, x= 'CYLINDERS', y= 'average_emission', color= 'FUEL' , trendline= 'ols')

p.update_layout(title= 'Emission by Cylinders for each Fuel Type',xaxis_title= 'Number of Cylinders', yaxis_title='Average Emission (g/km)', plot_bgcolor= 'white')

p.update_xaxes(showline=True, linecolor='black', linewidth=2, gridcolor='lightgray', gridwidth=0.5)
```

```
p.update_yaxes(showline=True, linecolor='black', linewidth=2, gridcolor='lightgray', gridwidth=0.5)

fuel_names = {'X': 'Regular Gasoline', 'Z': 'Premium Gasoline', 'D': 'Diesel', 'E': 'Ethanol (E85)', 'N': 'Natural Gas'}

engine_size_2_2 = df[df['ENGINE SIZE'] == 2.2].groupby('FUEL', as_index=False).agg(average_emission=('EMISSIONS', 'mean'))

engine_size_2_2['FUEL'] = engine_size_2_2['FUEL'].replace(fuel_names)

engine_size_2_2_sorted = engine_size_2_2.sort_values(by='average_emission', ascending=False)

engine_size_2_2_sorted

p = px.bar(engine_size_2_2_sorted, x='FUEL', y='average_emission', color='FUEL', text_auto= '.2s')

p.update_traces(textfont=dict(color='black'), textposition='outside')

p.update_layout(title='Emission by Fuel Type for Engine Size 2.2', xaxis_title= 'Fuel Types', yaxis_title='Average Emission (g/km)', plot_bgcolor='white')
```

```
p.update_xaxes(showline=True, linecolor='black', linewidth=2, gridcolor='lightgray',
gridwidth=0.5)

p.update_yaxes(showline=True, linecolor='black', linewidth=2, gridcolor='lightgray',
gridwidth=0.5)



"""# LINEAR REGRESSION"""



results = px.get_trendline_results(p)

print(results)



results.px_fit_results.iloc[0].summary()



df.groupby(['YEAR', 'MAKE'])['EMISSIONS'].mean().reset_index()



df['MAKE'] = df['MAKE'].str.lower()

df['YEAR'] = df['YEAR'].astype(str).str.lower()



result = df.groupby(['MAKE', 'YEAR'],
as_index=False).agg(average_emission=('EMISSIONS', 'mean'))



result.rename(columns={'YEAR': 'MODEL_YEAR'}, inplace=True)
```

```
result.sort_values('average_emission', ascending=False, inplace=True)

print(result)

df['MAKE'] = df['MAKE'].str.lower()
result = df.groupby(['MAKE'], as_index = False).agg(average_emission = ('EMISSIONS', 'mean'))
#result['MAKE'] = result['MAKE'].str.lower()
result.sort_values('average_emission', ascending = False, inplace = True)

result

p= px.bar(result, x='MAKE', y='average_emission', text_auto= '.2s')
p.update_layout( title =  'Average Emission by Manufacturer', xaxis_title= 'Manufacturer Company', yaxis_title='Average Emission (g/km)', plot_bgcolor = 'White' )
p.update_xaxes(showline=True, linecolor='black', linewidth=2, gridcolor='lightgray', gridwidth=0.5)
p.update_yaxes(showline=True, linecolor='black', linewidth=2, gridcolor='lightgray', gridwidth=0.5)

df['VEHICLE CLASS'] = df['VEHICLE CLASS'].str.lower()
```

```
vehicle_class_emission = df.groupby('VEHICLE CLASS', as_index = False).agg(average_emission = ('EMISSIONS', 'mean'))

vehicle_class_emission.sort_values('average_emission', ascending = False, inplace = True)

vehicle_class_emission



p= px.bar(vehicle_class_emission, x='VEHICLE CLASS', y='average_emission', text_auto= '.2s')

p.update_traces(textfont=dict(color='black'), textposition='outside')

p.update_xaxes(showline=True, linecolor='black', linewidth=2, gridcolor='lightgray', gridwidth=0.5)

p.update_yaxes(showline=True, linecolor='black', linewidth=2, gridcolor='lightgray', gridwidth=0.5)

p.update_layout( title =  'Average Emission by Vehicle Class', xaxis_title= 'Vehicle Class', yaxis_title='Average Emission (g/km)', plot_bgcolor = 'White' )
```

```
"""# YEARLY ANALYSIS"""



yearly_emission= df.groupby('YEAR', as_index = False).agg(average_emission = ('EMISSIONS', 'mean'))

p= px.scatter(yearly_emission, x= 'YEAR', y= 'average_emission', trendline= 'ols')

p.update_layout( title =  'Average Emission by Year', xaxis_title= 'Year', yaxis_title='Average Emission (g/km)', plot_bgcolor = 'White' )

p.update_xaxes(type='category')

p.update_xaxes(showline=True, linecolor='black', linewidth=2, gridcolor='lightgray', gridwidth=0.5)

p.update_yaxes(showline=True, linecolor='black', linewidth=2, gridcolor='lightgray', gridwidth=0.5)



bugatti_model = df[df['MAKE'] == 'bugatti'].groupby('YEAR', as_index=False).agg(average_emission=('EMISSIONS', 'mean'))

p = px.bar(bugatti_model, x='YEAR', y='average_emission', title='Average Emission for Bugatti Models Over the Years', text_auto= '.2s')

p.update_traces(textfont=dict(color='black'), textposition='outside')

p.update_layout(xaxis_title= 'Year', yaxis_title='Average Emission (g/km)', plot_bgcolor = 'White')

p.update_xaxes(showline=True, linecolor='black', linewidth=2, gridcolor='lightgray', gridwidth=0.5)

p.update_yaxes(showline=True, linecolor='black', linewidth=2, gridcolor='lightgray', gridwidth=0.5)
```

```
number_of_observation_by_year= df.groupby(['YEAR','FUEL','VEHICLE CLASS'],
as_index = False).agg(number_of_observation = ('YEAR', 'count'))

number_of_observation_by_year



number_of_observation_by_year_class = df[df['YEAR'].isin([2014,
2015])].groupby(['VEHICLE CLASS', 'YEAR']).size().reset_index(name='Number of
Observations')

number_of_observation_by_year_class



df_filtered = df[df['YEAR'].isin([2014, 2015])]



number_of_observation_by_year_class = df_filtered.groupby(['YEAR', 'VEHICLE
CLASS'], as_index=False).size()

number_of_observation_by_year_class.rename(columns={'size': 'Number of
Observations'}, inplace=True)

sorted_2014 =
number_of_observation_by_year_class[number_of_observation_by_year_class['YEAR']
== 2014].sort_values('Number of Observations', ascending=False)

sorted_2015 =
number_of_observation_by_year_class[number_of_observation_by_year_class['YEAR']
== 2015].sort_values('Number of Observations', ascending=False)



sorted_classes = pd.concat([sorted_2014, sorted_2015])
```

```
fig = px.bar(sorted_classes, x='YEAR', y='Number of Observations',
             color='VEHICLE CLASS',
             barmode='group',
             text='Number of Observations',
             labels={'YEAR': 'Year', 'Number of Observations': 'Number of Observations'},
             title='Number of Observations by Vehicle Class in 2014 and 2015')


fig.update_traces(texttemplate='%{y:.0f}', textposition='outside')
p.update_layout(xaxis_title= 'Year', yaxis_title='Number of Observation', plot_bgcolor = 'White' )


fig.update_xaxes(showline=True, linecolor='black', linewidth=2, gridcolor='lightgray', gridwidth=0.5)
fig.update_yaxes(showline=True, linecolor='black', linewidth=2, gridcolor='lightgray', gridwidth=0.5)
fig.update_layout(plot_bgcolor='white')
fig.show()


df_filtered = df[df['YEAR'].isin([2014, 2015])]
average_emissions_by_class = df_filtered.groupby(['YEAR', 'VEHICLE CLASS'], as_index=False)['EMISSIONS'].mean()
```

```
sorted_classes = average_emissions_by_class.groupby('VEHICLE CLASS')['EMISSIONS'].mean().sort_values(ascending=False).index

p = px.bar(average_emissions_by_class, x='YEAR', y='EMISSIONS', color='VEHICLE CLASS',
        barmode='group', labels={'EMISSIONS': 'Average Emissions (g/km)', 'YEAR': 'Year', 'VEHICLE CLASS': 'Vehicle Class'},
        title='Average Vehicle Class Emissions in 2014 and 2015',
        category_orders={'VEHICLE CLASS': sorted_classes})

p.update_traces(texttemplate='%{y:.0f}', textposition='outside')
p.update_xaxes(showline=True, linecolor='black', linewidth=2, gridcolor='lightgray', gridwidth=0.5)
p.update_yaxes(showline=True, linecolor='black', linewidth=2, gridcolor='lightgray', gridwidth=0.5)
p.update_layout(plot_bgcolor='white')

p.show()

average_engine_size = df.groupby('YEAR')['ENGINE SIZE'].mean().reset_index()

p = px.scatter(average_engine_size, x='YEAR', y='ENGINE SIZE')
```

```
p.update_layout(plot_bgcolor='white', xaxis_title='Year', yaxis_title='Average Engine Size (L)', title= 'Average Engine Size for Each Year')

p.update_xaxes(type='category', categoryorder='category ascending')

p.update_xaxes(showline=True, linecolor='black', linewidth=2, gridcolor='lightgray', gridwidth=0.5)

p.update_yaxes(showline=True, linecolor='black', linewidth=2, gridcolor='lightgray', gridwidth=0.5)

p.show()



df.replace({'FUEL': {'X': 'Regular Gasoline', 'Z': 'Premium Gasoline', 'D': 'Diesel', 'E': 'Ethanol (E85)', 'N': 'Natural Gas'}}, inplace=True)



yearly_fuel_emission = df.groupby(['YEAR', 'FUEL'], as_index=False).agg(average_emission=('EMISSIONS', 'mean'))



yearly_fuel_emission_filtered = yearly_fuel_emission[(yearly_fuel_emission['YEAR'] >= 2014) & (yearly_fuel_emission['YEAR'] <= 2015)]

yearly_fuel_emission_filtered = yearly_fuel_emission_filtered.sort_values(by='average_emission', ascending=False)



p = px.bar(yearly_fuel_emission_filtered, x='YEAR', y='average_emission', color='FUEL', barmode='group',

        title='Average Emission by Fuel Type for Each Year (2014-2015)', labels={'average_emission': 'Average Emission'}, text_auto='.2s')
```

```
p.update_traces(textfont=dict(color='black'), textposition='outside')

p.update_xaxes(type='category', categoryorder='category ascending')

p.update_layout(plot_bgcolor='white', xaxis_title='Year', yaxis_title='Average Emission (g/km)')

p.update_xaxes(showline=True, linecolor='black', linewidth=2, gridcolor='lightgray', gridwidth=0.5)

p.update_yaxes(showline=True, linecolor='black', linewidth=2, gridcolor='lightgray', gridwidth=0.5)

p.show()



average_fuel_consumption = df.groupby('YEAR')['FUEL CONSUMPTION'].mean().reset_index()

average_fuel_consumption



"""# 2009-2016 VW"""



df_filtered = df[(df['MAKE'] == 'VOLKSWAGEN') & (df['YEAR'].isin(range(2009,2017)))]

vw_model = df_filtered.groupby(['YEAR', 'FUEL'], as_index=False).agg(average_emission=('EMISSIONS', 'mean'))

p = px.bar(vw_model, x='YEAR', y='average_emission', color='FUEL', barmode='group',

        title='Average Emission for VOLKSWAGEN Models between 2009 and 2016', text_auto= '.2s')
```

```
p.update_xaxes(type='category', categoryorder='category ascending')

p.update_traces(texttemplate='%{y:.0f}', textposition='outside')

p.update_layout(plot_bgcolor='white', xaxis_title='Year', yaxis_title='Average Emission (g/km)')

p.update_xaxes(showline=True, linecolor='black', linewidth=2, gridcolor='lightgray', gridwidth=0.5)

p.update_yaxes(showline=True, linecolor='black', linewidth=2, gridcolor='lightgray', gridwidth=0.5)

p.show()



df_filtered = df[(df['MAKE'] == 'VOLKSWAGEN') & (df['YEAR'].isin(range(2014,2017)))]

vw_model = df_filtered.groupby(['YEAR', 'FUEL'], as_index=False).agg(average_emission=('EMISSIONS', 'mean'))

p = px.bar(vw_model, x='YEAR', y='average_emission', color='FUEL', barmode='group',

        title='Average Emission for VOLKSWAGEN Models between 2014 and 2016', text_auto= '.2s')

p.update_xaxes(type='category', categoryorder='category ascending')

p.update_traces(texttemplate='%{y:.0f}', textposition='outside')

p.update_layout(plot_bgcolor='white', xaxis_title='Year', yaxis_title='Average Emission (g/km)')

p.update_xaxes(showline=True, linecolor='black', linewidth=2, gridcolor='lightgray', gridwidth=0.5)
```

```
p.update_yaxes(showline=True, linecolor='black', linewidth=2, gridcolor='lightgray',
gridwidth=0.5)

p.show()



df_2014_diesel = df[(df['YEAR'] == 2014) & (df['FUEL'] == 'D')]



p_2014_diesel = px.pie(df_2014_diesel, names='MAKE', values='COMB (L/100 km)',
title='Fuel Consumption by Manufacturer in 2014 (Diesel)')



p_2014_diesel.update_traces(textinfo='percent+label', textposition='outside')

p_2014_diesel.update_layout(

    title='Fuel Consumption by Manufacturer in 2014 (Diesel)',

    plot_bgcolor='White'

)



p_2014_diesel.show()



df_2015_diesel = df[(df['YEAR'] == 2015) & (df['FUEL'] == 'D')]
```

```
p_2015_diesel = px.pie(df_2015_diesel, names='MAKE', values='COMB (L/100 km)',
title='Fuel Consumption by Manufacturer in 2015 (Diesel)')


p_2015_diesel.update_traces(textinfo='percent+label', textposition='outside')

p_2015_diesel.update_layout(

    title='Fuel Consumption by Manufacturer in 2015 (Diesel)',

    plot_bgcolor='White'

)


p_2015_diesel.show()


df_2014_diesel = df[(df['YEAR'] == 2014) & (df['FUEL'] == 'D')]

df_2015_diesel = df[(df['YEAR'] == 2015) & (df['FUEL'] == 'D')]


fig = make_subplots(

    rows=1, cols=2,

    subplot_titles=('2014', '2015'),

    specs=[[{'type': 'domain'}, {'type': 'domain'}]]

)
```

```
fig.add_trace(go.Pie(
    labels=df_2014_diesel['MAKE'],
    values=df_2014_diesel['COMB (L/100 km)'],
    textinfo='percent+label',
    textposition='outside'
), row=1, col=1)
fig.add_trace(go.Pie(
    labels=df_2015_diesel['MAKE'],
    values=df_2015_diesel['COMB (L/100 km)'],
    textinfo='percent+label',
    textposition='outside'
), row=1, col=2)
fig.update_layout(
    title={
        'text': 'Fuel Consumption by Manufacturer in 2014 and 2015 (Diesel)',
        'y': 0.95,
        'x': 0.5,
    },
    plot_bgcolor='White',
    margin=dict(t=100),
```

```
    annotations=[
        dict(
            text='2014',
            x=0.02,
            y=0.5,
            font_size=20,
            showarrow=False,
            xref="paper", yref="paper"
        ),
        dict(
            text='2015',
            x=0.5,
            y=0.5,
            font_size=20,
            showarrow=False,
            xref="paper", yref="paper"
        )
    ]
)
```

```
fig.show()


df_2014 = df[df['YEAR'] == 2014]

df_2015 = df[df['YEAR'] == 2015]


fig = make_subplots(

    rows=1, cols=2,

    subplot_titles=('Fuel Consumption by Manufacturer in 2014', 'Fuel Consumption by
Manufacturer in 2015'),

    specs=[[{'type': 'domain'}, {'type': 'domain'}]]

)

fig.add_trace(go.Pie(

    labels=df_2014['MAKE'],

    values=df_2014['COMB (L/100 km)'],

    textinfo='percent',

    textposition='outside'

), row=1, col=1)


fig.add_trace(go.Pie(

    labels=df_2015['MAKE'],

    values=df_2015['COMB (L/100 km)'],
```

```
    textinfo='percent',
    textposition='outside'
), row=1, col=2)
fig.update_layout(
    title={
        'text': 'Fuel Consumption by Manufacturer in 2014 and 2015',
        'y': 0.95,
        'x': 0.5,
    },
    plot_bgcolor='White',
    margin=dict(t=100),
    annotations=[
        dict(
            text='2014',
            x=0.02,
            y=0.5,
            font_size=20,
            showarrow=False,
            xref="paper", yref="paper"
        ),
```

```
        dict(
            text='2015',
            x=0.5,
            y=0.5,
            font_size=20,
            showarrow=False,
            xref="paper", yref="paper"
        )
    ]
)

fig.show()
```