A3: Individual Assignment

- Words

Aleksas Slavinskas

Computational Data Analytics with Python - DAT-5390 - LMBAN1

Professor: Dr. Tamie Salter

March 9th, 2025

Video link: 2263

Introduction

Nowadays, Environmental, Social and Governance (ESG) objectives remain highly important for companies and governments, as concerns regarding environment and social responsibility remain high. Environmental sustainability remains a particularly important piece of the framework, which addresses carbon emissions, climate change, and overall environmental pollution (British Business Bank, 2024). Government around the world are also concerned about contributing to ESG, including the Environmental aspect, however, are often criticized due to the taxes and laws being unfair, and often seen as a “cash grab” or biased towards some manufacturers, with one such example is Germany, where automakers have expressed concerns that the pollution tax for motorists of internal combustion engines (ICE) is unfair (Merten et al., 2022). This report will aim to analyze the fairness of the German personal passenger vehicle pollution tax, and if needed, provide recommendation for improvement.

Dataset preparation

To understand the fairness of German passenger vehicle pollution tax, Kfz-Steuer, it was first needed to acquire a dataset that would have a list of vehicles for sale, as well as fuel type, CO2 (carbon-dioxide) g/km emissions and engine size (in litres), since Kfz-Steuer tax is calculated based on these three variables, using the formula below (qmedia GmbH, 2025) (Figure 1).

Final Formula (Using Fuel Type)

$$\text{Tax} = (\max(0, \text{CO}_2 - 95) \times 2) + \left(\frac{\text{Engine Size (L)} \times 1000}{100} \times R \right)$$

Where:

- $R = 2$ for **Gasoline**
- $R = 9.50$ for **Diesel**

Figure 1. Formula summarization done by ChatGPT (OPENAI, 2025).

The dataset regarding vehicle information was found on Kaggle (Ranga, 2023), with all necessary information included: fuel type, emissions information, and engine size. The dataset was made to specifically summarize such information, and thus did not require any additional cleaning. However, two important features were added synthetically, with AI: vehicle average price and segment. The two columns were added to better understand tax changes based on vehicle price and segment, in order to test if more luxurious and expensive vehicles are indeed charged more, as otherwise, it would not be fair to those who cannot afford a vehicle more expensive than economy or standard. Further, the task of assigning average price and segment was given to AI, ChatGPT in this case, as otherwise it would have required to add every value manually by checking vehicle prices and segment they are aiming for, however AI could estimate the price of the vehicle quickly based on the manufacturer and model of the vehicle, as well as segment that the vehicle belongs to according to its category (sedan, SUV, etc.).

Once the dataset was synthetically expanded, it was time to add the last column, which was the pollution tax for each vehicle, which was done via python. The pollution tax for each vehicle was calculated following the formula, which would start of with an “if” function, as the multiplication values were different between petrol and diesel vehicles. Once this was done, and the “total tax” for each vehicle was calculated, it was time to do exploratory analysis and overview the fairness of the current tax system.

Overview of current tax system

Firstly, some data exploration can be done, such as the number of vehicles in each segment (the segment is based on the price and body type of vehicle). As it can be seen, even though economy segment is the cheapest, the segment with the most standard, with more than 2500 different options for a vehicle (including various trim levels), while the Super segment, that is the most expensive, still has more options to choose from than Economy, indicating that automakers prefer to focus on other segments, and not on economy (figure 2).

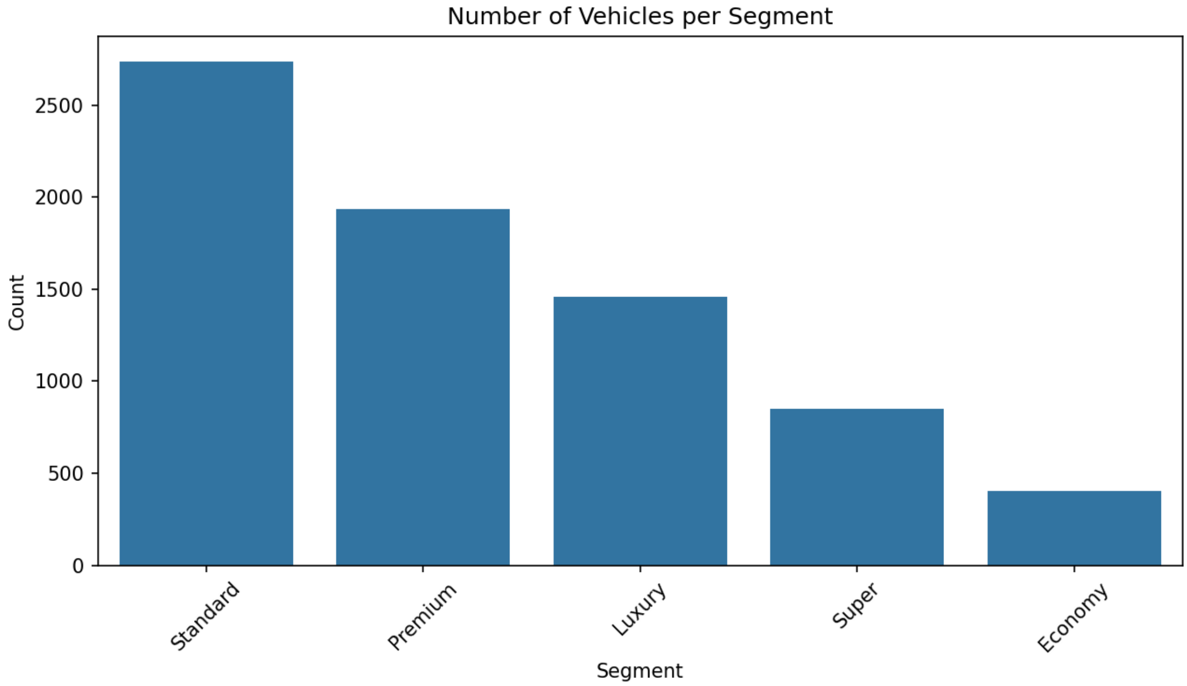

Figure 2. Number of vehicles per segment

Looking more closely at the prices of the vehicles, most vehicle offers are around 30'000 – 42'000 EUR price tag. It should be noted that the values were synthetically assigned with AI, which explains such distribution, however, it still indicates that the main focus for automakers is between 30 to 42 thousand EUR, with decent importance given to vehicles priced up to 70'000.

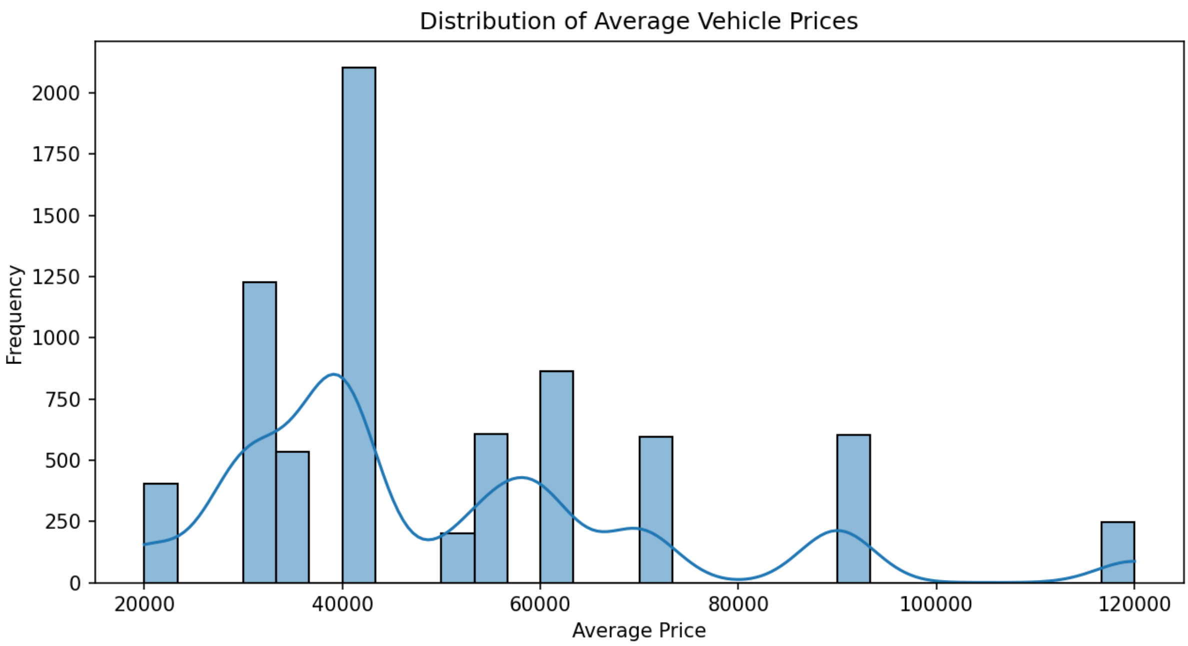

Figure 3. Distribution of Average Vehicle Prices.

Looking at the CO_2 emissions per segment, it can be seen that the cheapest segment, that is Economy, has indeed the lowest emissions, while the most popular Standard segment has second to lowest emissions, on average, however, the least polluting vehicles are also in this segment. From there, as the segment level increases, so do the emissions of the vehicle (Figure 4). Similarly, it seems that CO_2 emissions do indeed correlate with fuel consumption of a vehicle, as similar observations can be seen in the Figure 5: economy vehicles have the lowest fuel consumption, around 6L/100km, Standard around 8L/100km, and other segments being less fuel efficient (Figure 5).

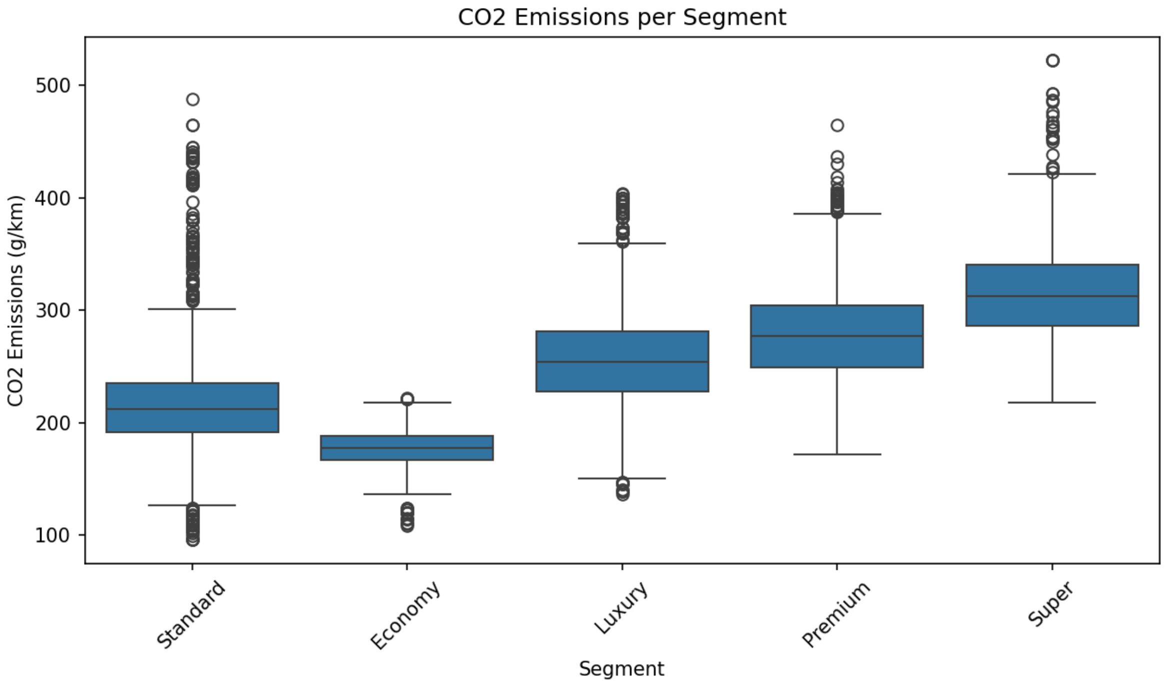

Figure 4. CO_2 emissions (g/km) per segment.

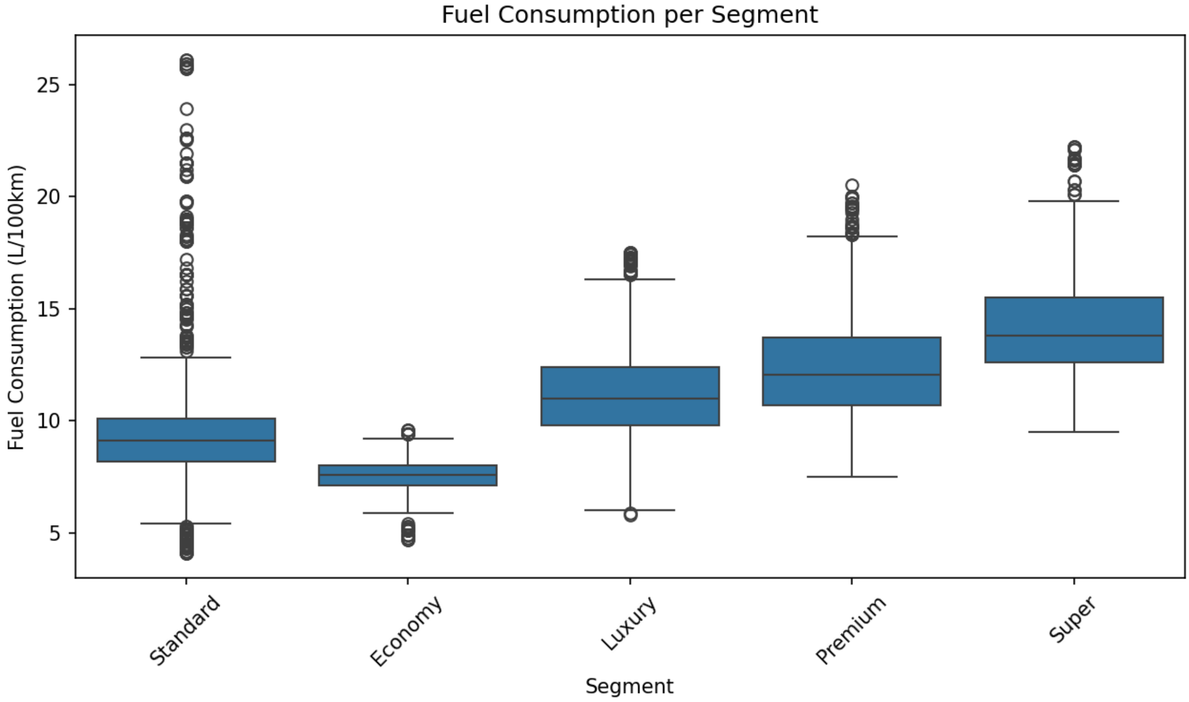

Figure 5. Fuel consumption per segment.

Looking at Total pollution tax per vehicle, it is also quite similar to the previous two graphs, with the tax being around 200 EUR per Economy vehicle, nearly 300 for Standard, and increasing all the way to 500 EUR for Super segment. From this perspective, it seems that the pollution tax is indeed quite accurate and fair.

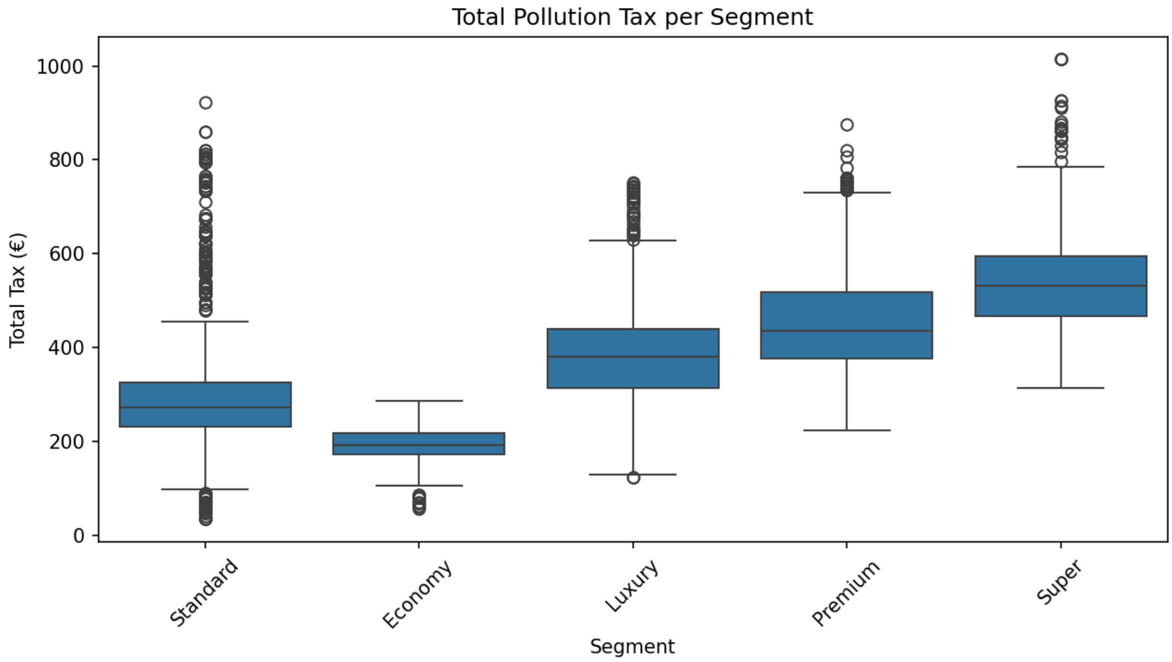

Figure 6. Pollution tax per segment.

To confirm if there was a correlation between pollution tax, fuel consumption and CO2 emissions, a correlation heatmap was done, including the vehicle price as well. As it can be seen, the pollution tax, labeled as Total Tax, has a 0.99 correlation with CO2 emissions, meaning that the tax system is very accurate and indeed fair, at least by looking such variables and summarizing all data. It can be also seen that Fuel Consumption and Engine Size does highly correlate with Total tax as well, while Average Price of the vehicle, while also having a high correlation of 0.58, has a lower value, most likely because a more expensive vehicle does not always constitute that it has an engine that pollutes more.

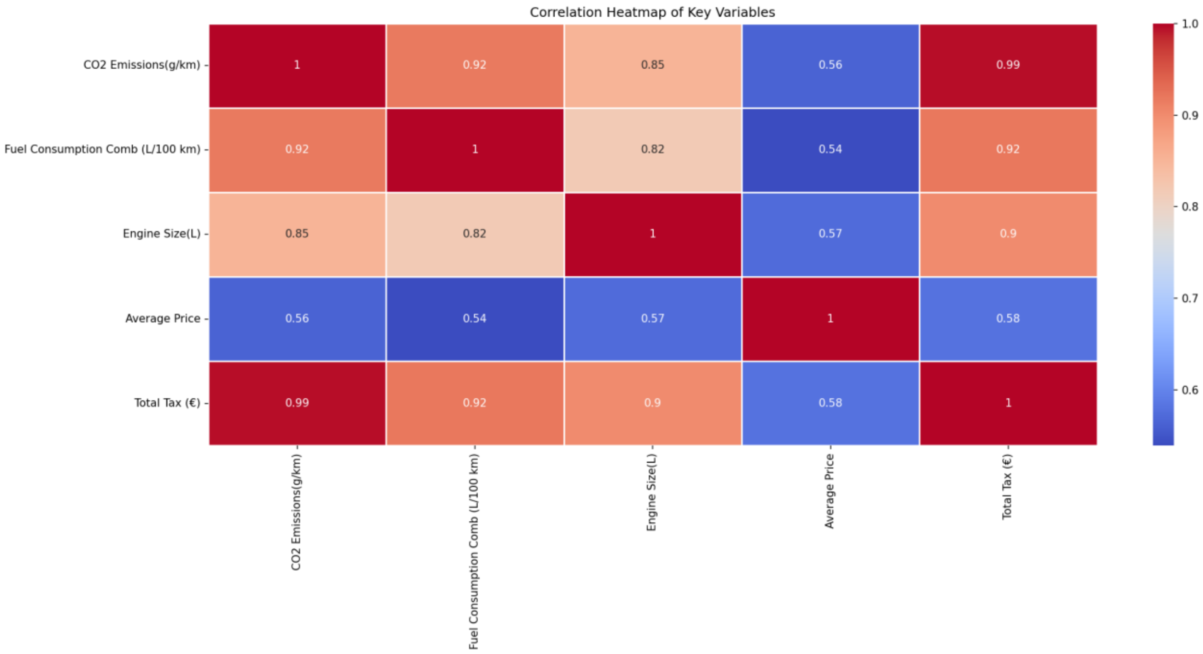

Figure 7. Pollution heatmap.

While the correlation heatmap indicates that there might the current German passenger vehicle pollution tax system is fair, there might be a case for unfair discrimination regarding diesel vehicles, as seen from the Figure 8 and Figure 9. As it can be seen, Diesel vehicles (indicated by D) consume less fuel than petrol ones (indicated by letter Z) and well as less by flex-fuel vehicles (indicated by letter X) or Ethanol powered (indicated by letter E). However, while diesel vehicles are indeed most efficient in terms of fuel consumption, diesel vehicles seem to be more pollutant, as according to Figure 9 findings, the average pollution of a diesel vehicle is almost the same as from a petrol one, while flex-fuel vehicles are the least polluting, on average.

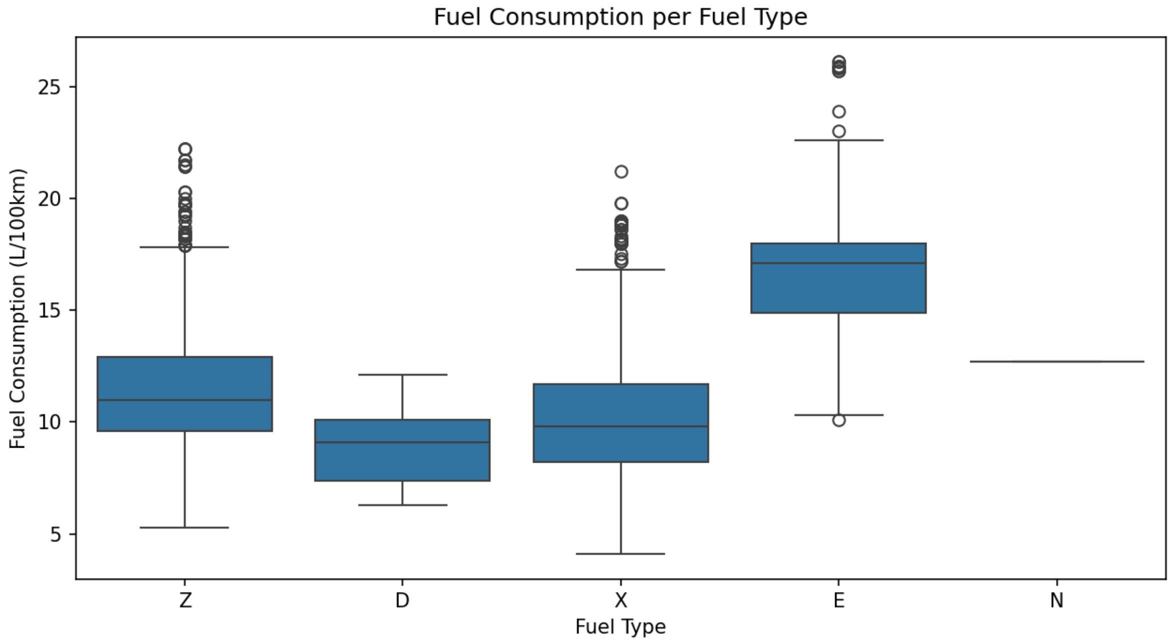

Figure 8. Average Fuel Consumption per Fuel Type.

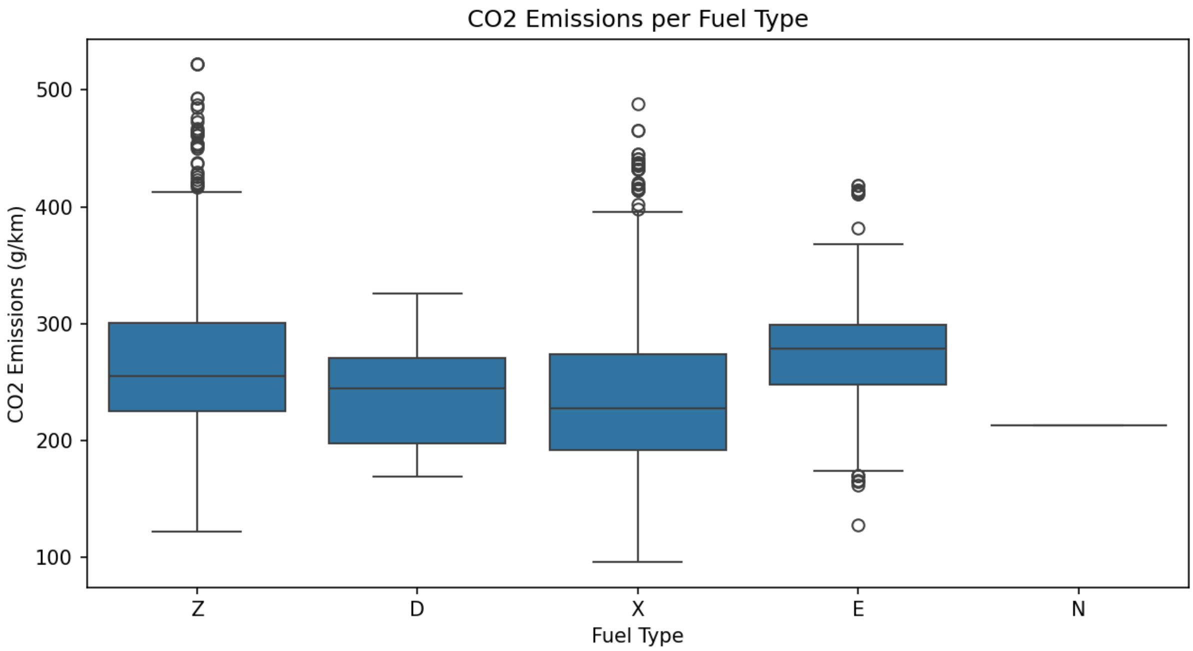

Figure 9. Average CO2 emissions per Fuel Type.

Looking at Figure 10, in broad terms, diesel vehicle owners do pay around similarly less in pollution tax, when compared to petrol vehicle owners, thus debunking a well-known myth. Further, flex-fuel vehicle owners seem to be paying the least, on average, for pollution tax, and the overall tax system seems to be fair.

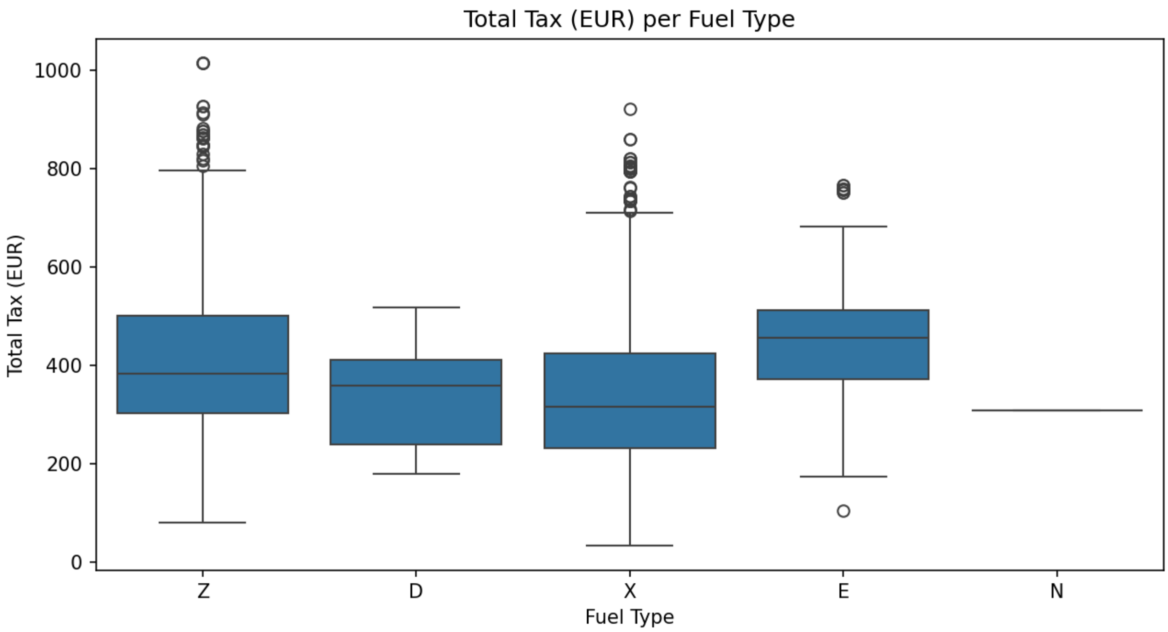

Figure 10. Total Tal tax per Fuel Type.

However, considering that in the formula, diesel vehicle owners have to pay much more tax, than those who own petrol cars, however, the correlation of 0.99 indicates otherwise, it might be the case that Diesel owners are overpaying around the same as much as petrol vehicle owners, or owners of other types of vehicles, are being undertaxed. Looking at Figure 11, which displays each vehicle individually, it seems that it might not be the case, however, it is not fully clear, due to the number of vehicles that there are. Another interesting insight is that technically, if environmental pollution in what matters, then the engine size should not be a part of the calculation, as in this case, a car owner with a bigger engine might be taxed more even though their car is polluting less than others because of, for example, more sophisticated engine. In the case where the pollution tax is indeed fair, the scatter plot should for a nearly perfect line when displayed in such C02 emissions vs Total tax form, however, it seems that there are small inconsistencies throughout nearly all vehicles, which means, that some might be paying more or less than others, whose cars are polluting identically the same. For this to be fully confirmed, a linear regression model will have to be applied.

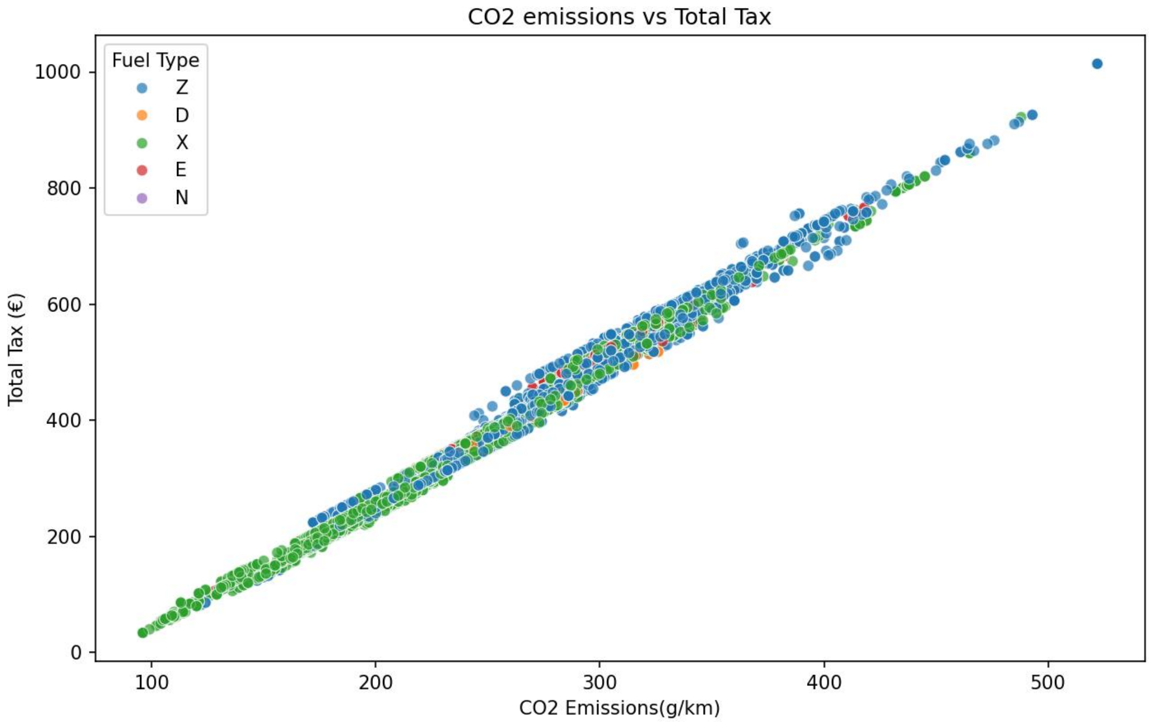

Figure 11. CO2 emissions vs Total Tax per fuel type.

Applying linear regression to improve the tax system

To finally pin-point and accurately measure the fairness of the current tax system, a linear regression model can be applied in order to extract values such as regression coefficient and intercept value and compare it against total tax paid by CO2 emissions. As mentioned before, in this case, engine size will not be taken into account, as even though larger engines do seem to have higher emissions, the two same-sized engines will not have the same pollution, as indicated by Figure 11. Furthermore, similar can be said about vehicles of different fuel types, as if it is the CO2 pollution only that matters (as indicated by tax formula), then technically, it should not matter based on what sort of fuel the emissions are coming from.

The linear regression was run by fitting "CO2 emissions (g/km)" value as X, and "Total Tax (€)" as Y value. Further, a line was programmed to be displayed over the same type of graph as seen in Figure 11. The result, which can be seen in Figure 12, displays that indeed the current tax system is "approximately" fair as significant amount of vehicles are positioned around the linear regression line. This indicates that owners of vehicles below the regression line are being undertaxed for pollution, while those above the line, overtaxed. Further, we can see that vehicles with lowest CO2 emissions are all overtaxed, while the owners of most polluting vehicles are not contributing enough. Remembering that cheapest, economy segment, that is the cheapest vehicles are the ones that usually have the lowest CO2 emissions, yet the owners pay more than they should, while owners of “Super” segment are paying less than they should in a fair system. Such insights indicate that previously found correlation of 0.99 between indeed false in terms of fairness, as some vehicles are indeed canceling out each other. This also calls for improvement of the formula.

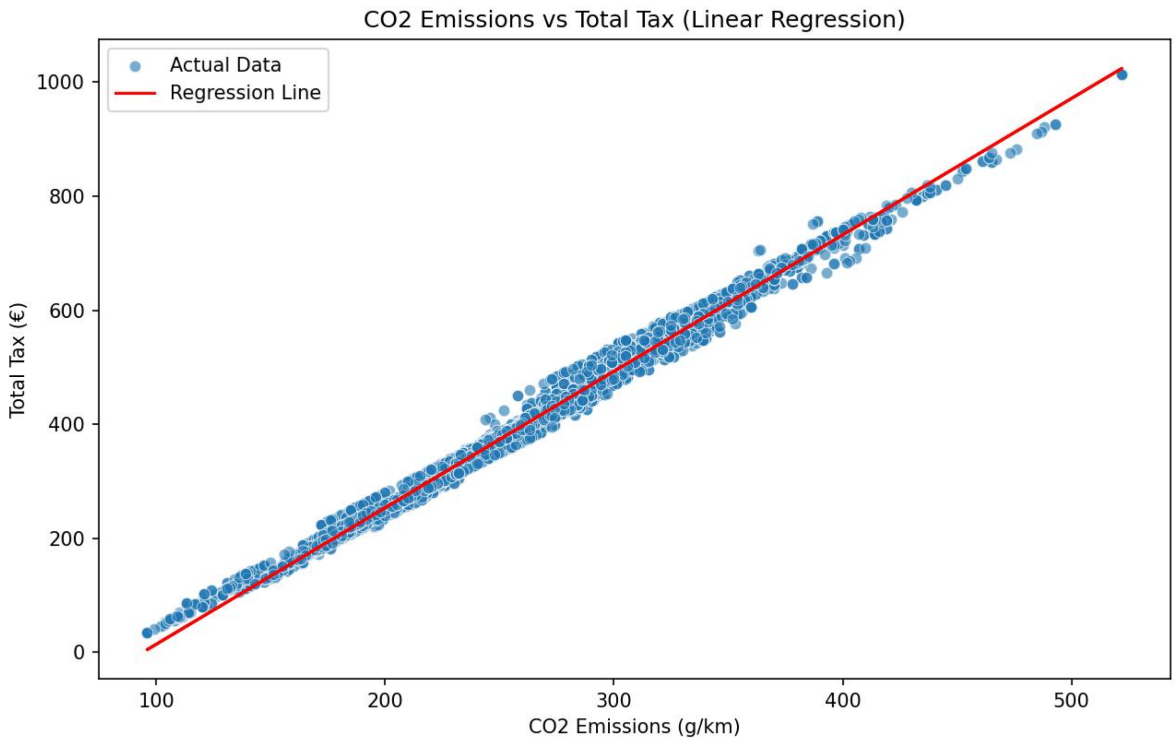

Figure 12. CO2 emissions vs Total Tax.

The current formula of such regression line had values of 2.39 for coefficient and intercept of -225.52, indicating that under normal circumstances, vehicles with emissions ~94 g/km or less, should not be taxed, however, according to Figure 12, they still were. With this information, it can be assumed that the government wants to promote non-fossil fuel-based modes of transportation, thus, the new formula for calculating pollution tax might not need an intercept value, and could do with just a different coefficient, especially if only the tax is based solely on CO2 emissions.

The new tax formula was found by first recalculating pollution tax for each vehicle by using the coefficient of 2.39, which was found before. The total recalculated tax and current tax were summed and divided, this way finding the adjustment factor, as otherwise, the total sum of tax collected might be lower or higher than before. The recalculated pollution tax values were then multiplied by the adjustment factor, thus matching the total sum of tax to the one it was before. This new recalculated tax was saved as a new column for each vehicle. Finally, this new column was used to run a linear regression analysis again, similar to the one before, just on the newly adjusted tax.

As it can be seen for the figure 13, the newly adjusted tax formula calculates taxes for each vehicle near-perfectly, and thus, can be indeed labeled as “fair”, or at least “fairer”, as all the tax paid by current vehicle owners will directly and accurately correlate with the CO2 emissions of their vehicle. The new formula for the pollution tax is also very simple:

“P.Tax = 1.49 * C02 emissions (g/km)”

With this method of pollution tax calculation, the previous problem of unfair tax charges is removed, since, as it can be seen, the correlation between CO2 and total tax paid is equal to 1. Further, compared to the previous tax formula, it decreases the financial burden of economy class vehicles, while making it more fair for those who can afford excessively luxurious vehicles. Lastly, with this value, the total amount of tax would change by less than 0.5%, thus, not overcharging the owners and providing the same amount of tax income to the government as before.

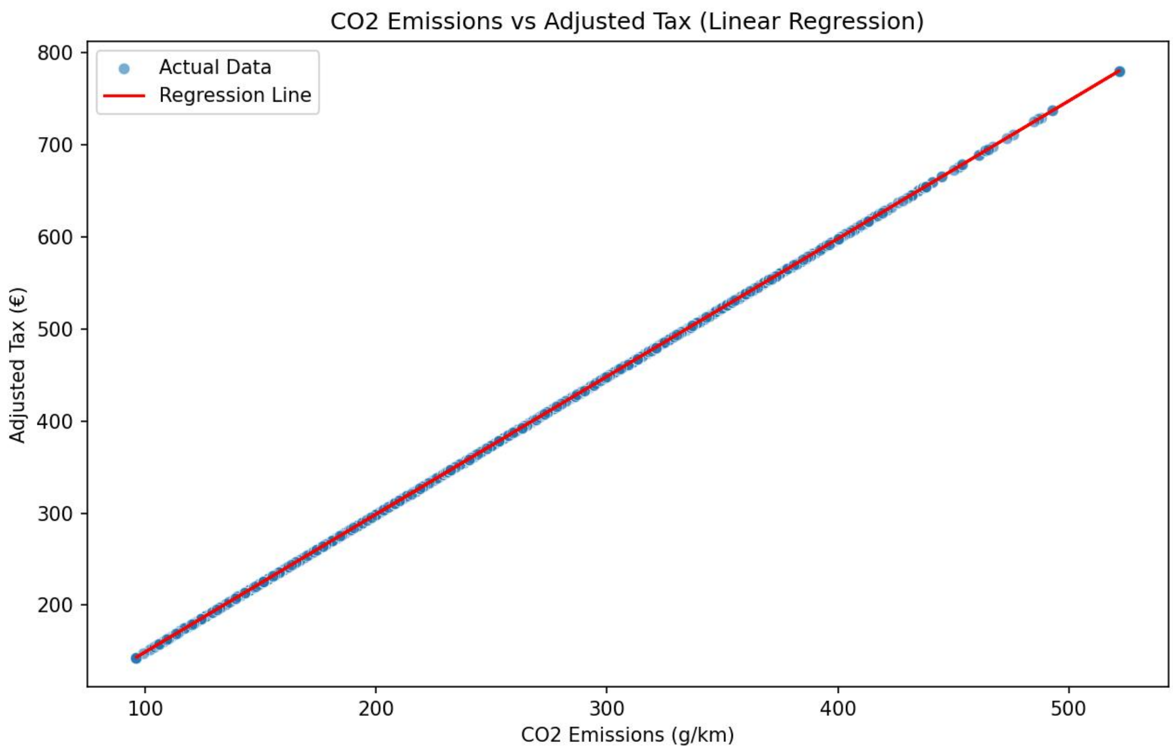

Figure 13. CO2 emissions vs Adjusted Tax.

Limitations & issues

A few limitations must be mentioned for the end. First of all, data regarding vehicle prices and segment was added synthetically, and might not resemble the real life costs and segments of the vehicles, for which reason, the fact that economy vehicle owners were being overcharged should be taken with a “grain of salt”. Further, the dataset only includes new vehicles on sale, however, many vehicles on the road are years, even decades old, and presumably, pollute more in terms of different particles, when having the same engine and fuel metrics. The issue around different particles beyond CO2, such as CO (carbon-monoxide) or NOx (nitrogen oxides) is relevant for older vehicles that have older particle filters, as well as for diesel vehicles which are known to emit more NOx particles (Green Vehicle Guide, 2023). However, on the other hand, the original formula used in Kfz-Steuer does not consider such particles as well, only different tax on engine size for vehicles of different fuel type.

Refence list

British Business Bank. (2024). What is ESG? A guide for businesses | British Business Bank. Www.british-Business-Bank.co.uk. https://www.british-business-bank.co.uk/business-guidance/guidance-articles/sustainability/what-is-esg-a-guide-for-smaller-businesses

Green Vehicle Guide. (2023). Vehicle emissions | Green Vehicle Guide. Www.greenvehicleguide.gov.au. https://www.greenvehicleguide.gov.au/pages/UnderstandingEmissions/VehicleEmissions

Merten, M. J., Becker, A. C., & Matthies, E. (2022). What explains German consumers' acceptance of carbon pricing? Examining the roles of pro-environmental orientation and consumer coping style. Energy Research & Social Science, 85, 102367. https://doi.org/10.1016/j.erss.2021.102367

OpenAI. (2025). *ChatGPT*. ChatGPT; OpenAI. https://chatgpt.com/

qmedia GmbH. (2025). Car tax calculator for Germany» kfz-steuer.wiki. Kfz-Steuer.wiki. https://kfz-steuer.wiki/en/car-tax-germany/

Ranga, B. (2023). CO2 Emissions. Www.kaggle.com. https://www.kaggle.com/datasets/bhuviranga/co2-emissions