The Effect of Car Weight and Volume on CO2 Emissions

Sarah Bareh

Western Governors University

Table of Contents

Project Overview

A. Project Highlights

Research Question:

The goal of the project was to understand how the weight and volume of a car could impact its CO2 emissions, providing valuable insights to Future Cars, a car manufacturer that aimed to have a positive impact on the environment. Through the analysis of this relationship, Future Cars would be able to make informed decisions about the design and manufacturing of environmentally friendly cars.

Project Scope:

The scope of this project included an analysis of the relationship between car weight, volume, and CO2 emissions, based on a preexisting dataset. An analytical report was developed to explain how CO2 emissions were influenced by car weight and volume. The input for this report was the data points from the dataset representing the variables of interest. The output was scatterplots and multilinear regression table describing the relationship between car weight, volume and CO2 emissions.

Solution Overview - Tools and Methodologies

To develop a solution to understand the relationship between car weight, volume, and CO2 emissions. I used Python programming language and its included libraries such as Pandas, NumPy, Scikit, and Matplotlib for data manipulation, analysis, and visualization. The project also used a multi-linear regression model, which allowed for the interpretation of the relationship between the independent variables (car weight and volume) and the dependent variable (CO2 emissions). The solution involved analyzing a preexisting dataset obtained from Kaggle, specifically focused on CO2 emissions of cars.

Project Plan

B. Project Execution

Project Plan

The plan of the project was to write a python code in jupyter notebooks that would be used to analyze the relationship between car weight, volume and CO2 emissions. The plan was to collect data, clean it and prepare it for analysis. Then statistical analysis would be carried using multilinear regression table and scatter plot charts. Finally, the results would be conveyed through a written report and a visual presentation.

Project Planning Methodology

The waterfall methodology was used as expected. The order of steps presented in the proposal was followed. This sequential approach allowed for a systematic and organized progression where each phase was completed successfully before moving to the next.

Project Timeline and milestones

The following table indicates each milestone and the initial projected start and end dates, and projected duration as well as the actual timeline that was followed once the project was executed:

Milestone	Projected Start Date	Actual Start Date	Projected End Date	Actual End Date	Final Duration (days/hours)
Requirement gathering	06-15-2023	06-15-2023	06-17-2023	06-15-2023	1 days
Design	06-17-2023	06-16-2023	06-20-2023	06-16-2023	1 days
Implementation	06-20-2023	06-17-2023	06-21-2023	06-17-2023	1 days
Testing	06-21-2023	06-18-2023	06-22-2023	06-18-2023	1 days
Maintenance	06-21-2023	06-19-2023	Ongoing		N/A

Methodology

C. Data Collection Process

Data selection was performed as planned. The dataset provided in the proposal was used as intended. There were no obstacles to collecting or using the data since it is available publicly on Kaggle under a public domain license. There were no data governance issues experienced as additional obstacles. Discuss these elements; offer examples.

C1. Advantages and Limitations of Data Set

The dataset used for the purpose of this project is preexisting on Kaggle.com. Preexisting datasets are convenient and save time through eliminating the need to collect data from scratch. Data can be directly accessed and utilized without investing resources in data collection, cleaning, and validation. This convenience allows for a more efficient analysis process. Using a preexisting dataset is a more cost-effective option, since collecting data can be very expensive and resource exhaustive.

The biggest disadvantage experienced during this project was limited variables and scope in the dataset. The dataset chosen for the project was small with only 36 rows and 4 variables which limits the depth and breadth of the analysis. The small size of the dataset also decreases the accuracy and the generalizability of the results.

D. Data Extraction and Preparation Processes

The data was downloaded from Kaggle.com to my local machine. Jupyter notebook was selected as the working environment because of its ease of use and the ability to execute the code inline. Python was used as planned. Pandas library was used to create a data frame from the dataset so it can be easily parsed and manipulated. During preparation, a column consisting of all null values titles ‘Unnamed: 5’ was deleted from the data frame so we can have clean consistent dataset.

E. Data Analysis Process

E1. Data Analysis Methods

For this project, the python library seaborn was used to create 2 scatterplots. The first scatterplot was used to describe the relationship between car weight and CO2 emissions. The second one was used to describe the relationship between car volume and CO2 emissions. Then the python library statsmodel was used to create a multilinear regression model table to describe and analyze the relationship between all three variables of interest.

E2. Advantages and Limitations of Tools/Techniques

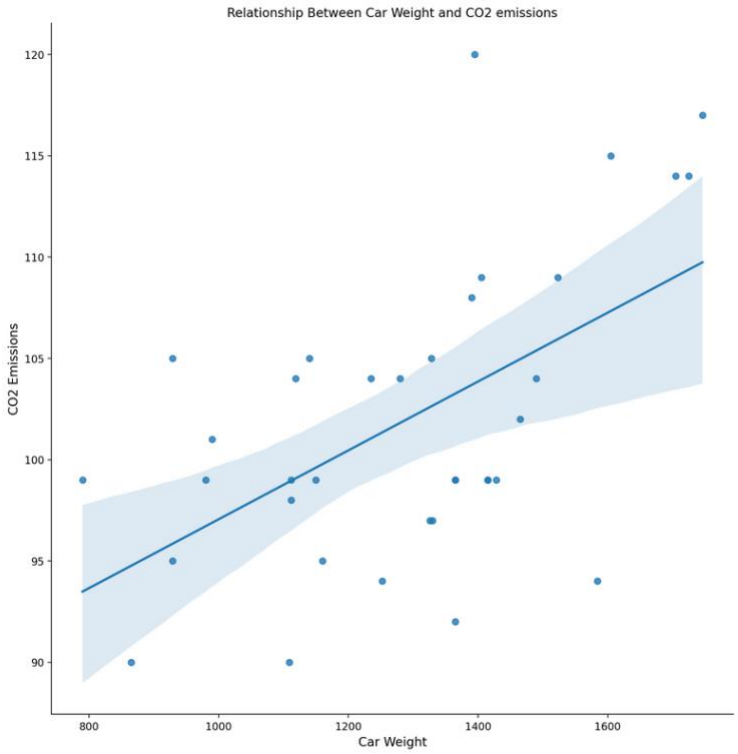
The seaborn library provides is very powerful tool used for creating visually appealing and informative plots. The scatterplots generated using seaborn allow for a clear visualization of the relationship between car weight and CO2 emissions, as well as car volume and CO2 emissions. Visual representations help in better understanding patterns, trends, and potential outliers in the data.

The statsmodel library offers a comprehensive set of statistical models and methods for analyzing data. In this project, using the multilinear regression model from statsmodel allowed for studying the relationship between all three variables of interest simultaneously. It provided insights into the individual and combined effects of car weight and volume on CO2 emissions.

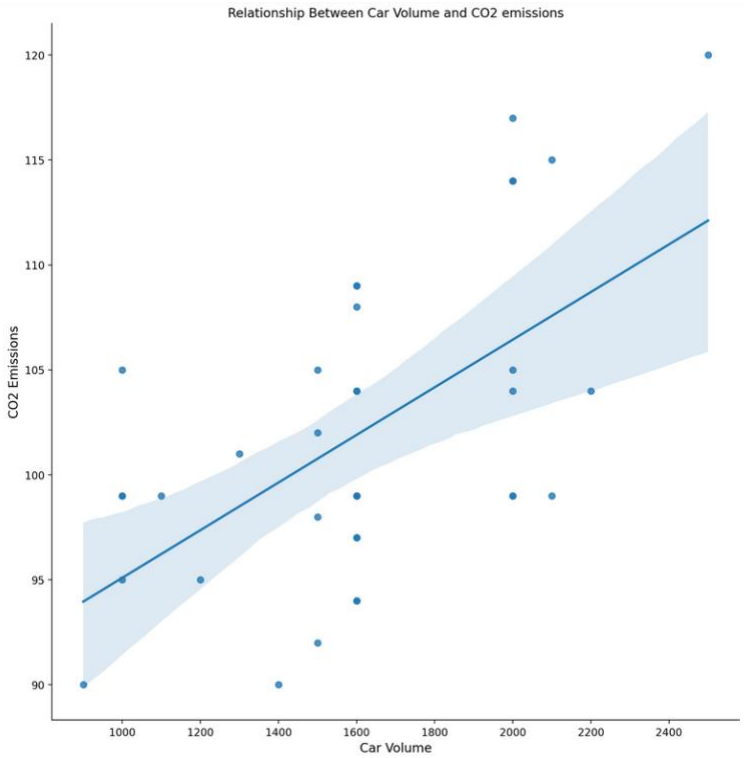
On the other hand, the scatterplots and regression analysis provide a simplified understanding of the relationships between variables. They do not account for potential confounding factors or consider the complex interactions that may exist in the real world. Additionally, multilinear regression assumes certain assumptions, such as linearity and independence. Violations of these assumptions can affect the validity and reliability of the regression results.

E3. Application of Analytical Methods

The following scatter plot shows a strong positive correlation between car weight and CO2 emissions:



The following scatter plot shows a strong positive correlation between car volume and CO2 emissions:



The following table represents the multilinear regression model used to explain the relationship between the variables of interest:

OLS Regression Results

Dep. Variable:	CO2	**R-squared (uncentered):**	0.977
Model:	OLS	**Adj. R-squared (uncentered):**	0.975
Method:	Least Squares	**F-statistic:**	714.9
Date:	Tue, 20 Jun 2023	**Prob (F-statistic):**	1.67e-28
Time:	11:05:40	**Log-Likelihood:**	-149.96
No. Observations:	36	**AIC:**	303.9
Df Residuals:	34	**BIC:**	307.1
Df Model:	2		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Weight	0.0648	0.013	4.850	0.000	0.038	0.092
Volume	0.0097	0.011	0.919	0.365	-0.012	0.031

Omnibus:	2.430	**Durbin-Watson:**	0.718
Prob(Omnibus):	0.297	**Jarque-Bera (JB):**	2.239
Skew:	0.550	**Prob(JB):**	0.327
Kurtosis:	2.470	**Cond. No.**	13.4

Results

F. Project Success

F1. Statistical Significance

To Assess statistical significance of the findings, a null and an alternative hypothesis were created. The null hypothesis states that, “There is no connection between car weight, volume, and CO2 emissions. It suggests that the weight and volume of a car do not have any effect on the amount of CO2 emissions it produces.”. While the alternative hypothesis states that, “There is a relationship between car weight, volume, and CO2 emissions. It proposes that the weight and volume of a car do have an impact on the amount of CO2 emissions it produces.”. The Type I error rate was set at 5% which means that if the p-value is more than 0.05 then we would fail to

reject the null hypothesis. On the other hand, if the p-value is less than 0.05 then would reject the null hypothesis and accept the alternative hypothesis.

By creating a multilinear regression model and fitting the summary into the table previously mentioned we can deduct the following results:

- For the effect of the independent variables (car weight and volume) on the dependent variables (CO2 emissions)
    - The r-squared was 0.977, indicating that approximately 97.7% of the variation in CO2 can be explained by the independent variables.
    - The f-statistic was 714.9, with a very low p-value (1.67e-28), suggesting that the overall model is statistically significant.
    - Weight: The coefficient estimate is 0.0648, with a standard error of 0.013. The t-value is 4.850, indicating that the weight variable has a statistically significant effect on CO2 emissions (p-value < 0.05).
    - Volume: The coefficient estimate is 0.0097, with a standard error of 0.011. The t-value is 0.919, indicating that the volume variable does not have a statistically significant effect on CO2 emissions (p-value > 0.05).

F2. Practical Significance

In practical terms, these results imply that car weight plays a significant role in determining the amount of CO2 emissions. For example, a heavier car is likely to produce higher CO2 emissions compared to a lighter car. This finding has implications for car manufacturers, as it suggests that reducing the weight of vehicles can lead to a decrease in CO2 emissions and an improvement of the environmental effects of cars. Future Cars, the car manufacturer interested in this study, can use this insight to guide their design and manufacturing processes. They may consider using lighter materials or implementing weight-reducing technologies to develop more environmentally friendly cars.

F3. Overall Success

I believe that this project was successful. The reason for this lies in the fact that the main goal of the project was to investigate the relationship between car weight, volume, and CO2 emissions. This goal was achieved by using the appropriate analytical methods that provided insights into the significance of these variables and their impact on emissions. The analysis showed both statistical and practical significance in the relationship between car weight and CO2 emissions. The high R-squared value and the statistically significant coefficient estimate for weight suggests a strong association between these variables. This finding has practical implications for Future Cars and other car manufacturers, enabling them to make informed decisions to reduce emissions and improve their environmental impact.

G. Key Takeaways

G1. Summary of Conclusions

In conclusion, the key takeaway from this project is that there is a significant statistical relationship between car weight and the amount of CO2 emissions. On the contrary, the car volume did not have any statistical significance based on the multilinear regression model generated. These findings can help Future Cars make informed decisions in designing and manufacturing more environmentally friendly cars.

G2. Effective Storytelling

The first chosen tool was scatter plots which are effective in visually representing the relationship between variables, allowing viewers to quickly grasp the patterns and trends in the data. The data is plotted on a two-dimensional graph, where the x-axis represents the independent variable (car weight or volume), the y-axis represents the dependent variable (CO2 emissions). This simple chart allows viewers to easily understand the relationship between the variables and identify any potential patterns or outliers. The other tool that was chosen for this project was a multilinear regression table, which provide numerical information about the coefficients, standard errors, t-

values, and p-values. These values offer quantitative insights into the significance and impact of the variables. This information helps to support the narrative by providing statistical evidence and reinforcing the conclusions drawn from the visual representations.

G3. Findings-based Recommendations

Based on the analysis and findings, the following are two logical recommendations related to the research question and organizational need of the project:

1. Optimize Car Design: The analysis showed that car weight had a statistically significant effect on CO2 emissions, which indicates that reducing the weight of cars can help reduce their environmental impact. Therefore, one logical course of action is for Future Cars to focus on optimizing car design by possibly using lightweight materials in cars. This can lead to the development of fuel-efficient and eco-friendly cars which will lead to a significant reduction in CO2 emissions and a positive impact on the environment.

2. Further Investigate Volume Effect: The analysis showed that car volume does not have a statistically significant effect on CO2 emissions. Future Cars can conduct additional research or experiments to explore the potential impact of car volume on emissions in specific scenarios. This can involve analyzing data from different car models, conducting controlled experiments, or collaborating with industry experts and researchers. By further investigating the volume effect, Future Cars can improve their design strategies and make informed decisions regarding car volume and its influence on CO2 emissions.

H. Panopto Presentation

https://wgu.hosted.panopto.com/Panopto/Pages/Viewer.aspx?id=ad439f90-0139-49f4-9601-b02901292a07

Appendices

I. Evidence of Completion

Included in my submission:

- Jupyter notebook: capstone.ipynb
- Recommendation Report Slides: capstone.pptx
- Original dataset: DATA.csv

Sources

L, M. D. (2022, July 18). *CO2 Emission of Cars Dataset*. Kaggle.

https://www.kaggle.com/datasets/midhundasl/co2-emission-of-cars-dataset