

Objective

This report aims to explain why CO2 is released, their impacts on both our environment and daily lives, and potential solutions. We'll see the factors driving CO2 release. Then, we'll suggest some simple ways to reduce CO2 emissions and protect our planet.

Introduction

In today's world, carbon dioxide (CO2) emissions have become a pressing concern due to their significant impact on the environment and human health. It's a gas that we release into the air from various activities, like driving cars. CO2 is causing a lot of issues, like making our planet hotter and changing our weather. This report is all about understanding why CO2 happens, what it does to our world, and what we can do to make things better. We'll talk about why CO2 gets released, how it affects us, and some simple ideas to reduce it. Let's dive in and learn how we can protect our planet together.

Literature Review

Climate change is a big problem caused by things like carbon dioxide (CO2) in the air. Many books help us understand why it's happening and what we can do.

Bill Gates' 'How to Avoid a Climate Disaster' offers profound insights into the causes, impacts, and solutions for mitigating climate change, urging urgent action for a sustainable future."

"The Sixth Extinction" by Elizabeth Kolbert talks about how climate change is making animals and plants disappear.

Naomi Klein's "This Changes Everything" explains how our way of living and making money is hurting the planet.

In "Drawdown" edited by Paul Hawken, scientists give us ideas to stop climate change, like using cleaner energy and planting trees.

About Dataset

CONTENT

This dataset captures the details of how CO2 emissions by a vehicle can vary with the different features. The dataset has been taken from Canada Government official open data website. This is a compiled version. This contains data over a period of 7 years.

There are total 7385 rows and 12 columns. There are few abbreviations that has been used to describe the features. I am listing them out here. The same can be found in the Data Description sheet.

Model

4WD/4X4 = Four-wheel drive

AWD = All-wheel drive

FFV = Flexible-fuel vehicle

SWB = Short wheelbase

LWB = Long wheelbase

EWB = Extended wheelbase

Transmission

A = Automatic

AM = Automated manual

AS = Automatic with select shift

AV = Continuously variable

M = Manual

3 - 10 = Number of gears

Fuel type

X = Regular gasoline

Z = Premium gasoline

D = Diesel

E = Ethanol (E85)

N = Natural gas

Fuel Consumption

City and highway fuel consumption ratings are shown in litres per 100 kilometres (L/100 km) - the combined rating (55% city, 45% hwy) is shown in L/100 km and in miles per gallon (mpg)

ACKNOWLEDGEMENTS

The data has been taken and compiled from the below Canada Government official link https://open.canada.ca/data/en/dataset/98f1a129-f628-4ce4-b24d-6f16bf24dd64#wb-auto-6

Data Preprocessing

Data Cleaning

Duplicates

In our dataset, we've identified 1103 duplicate entries. To ensure data accuracy and integrity, our initial step involves removing these duplicates from the dataset.

After removing duplicates, the dataset contains 7385 entries, numbered from 0 to 7384.

Data Types

The data types include float64 for 4 columns, int64 for 3 columns, and object for 5 columns.

#	Column	Non-Null Count	Dtype
0	Make	6282 non-null	object
1	Model	6282 non-null	object
2	Vehicle Class	6282 non-null	object
3	Engine Size(L)	6282 non-null	float64
4	Cylinders	6282 non-null	int64
5	Transmission	6282 non-null	object
6	Fuel Type	6282 non-null	object
7	Fuel Consumption City (L/100 km)	6282 non-null	float64
8	Fuel Consumption Hwy (L/100 km)	6282 non-null	float64
9	Fuel Consumption Comb (L/100 km)	6282 non-null	float64
10	Fuel Consumption Comb (mpg)	6282 non-null	int64
11	CO2 Emissions(g/km)	6282 non-null	int64

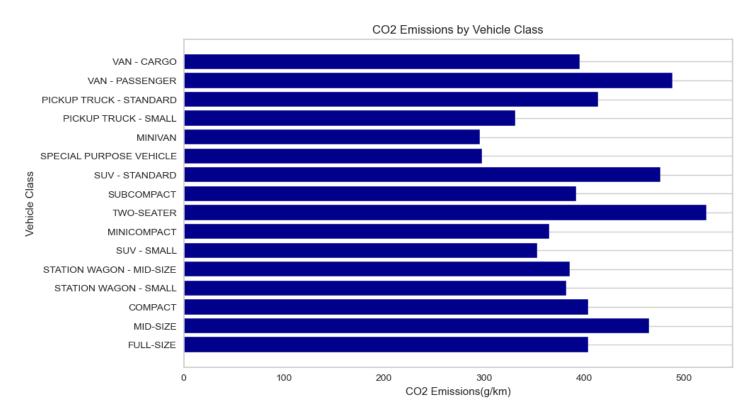
Missing Values

In our dataset, we've determined that there are no missing values present.

Data Analysis

Analysis of Relationship Between Vehicle Class and CO2 Emissions

Based on the analysis, the bar chart depicting CO2 emissions by vehicle class reveals that the "Two-Seater" class exhibits the highest CO2 emissions, with levels reaching 500 g/km. Following closely are the "Van-Passenger," "SUV-Standard," "Mid-Size," "Pickup Truck-Standard," and "Full-Size" classes, each emitting approximately 400 g/km of CO2. Other vehicle classes exhibit lower emissions, typically around 300 g/km.



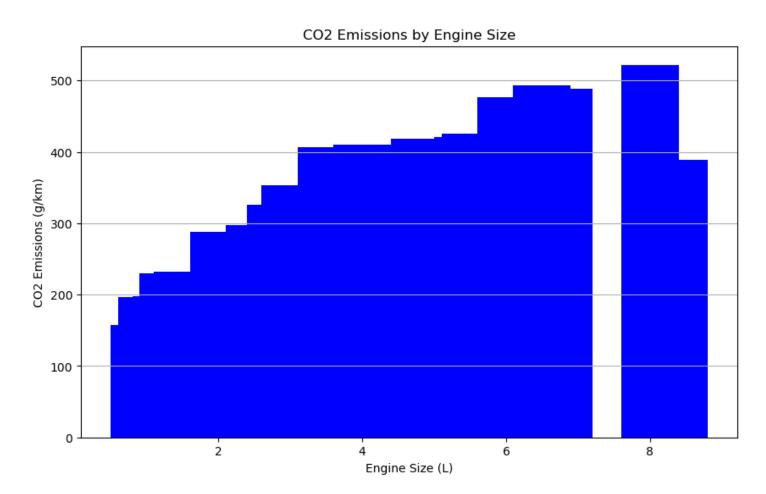
Analysis of Relationship Between Engine Size and CO2 Emissions

The bar chart clearly shows that as the engine size of vehicles increases, their CO2 emissions also go up. This matches what we know about how cars work.

Bigger engines burn more fuel each time they work, so they release more CO2. This happens because they can burn a larger amount of air and fuel together in each cycle, which makes more energy and more emissions. On the other hand, smaller engines burn less fuel, so they release less CO2.

This relationship between engine size and CO2 emissions in the bar chart reminds us that the size of the engine is an important factor when we talk about how much pollution a car makes. It shows that bigger engines usually make more greenhouse gas emissions.

This analysis highlights the importance of making engines more efficient and using technology to reduce CO2 emissions from cars. As car companies work on new ideas, finding ways to make engines better and using different types of fuel become really important for helping the environment.



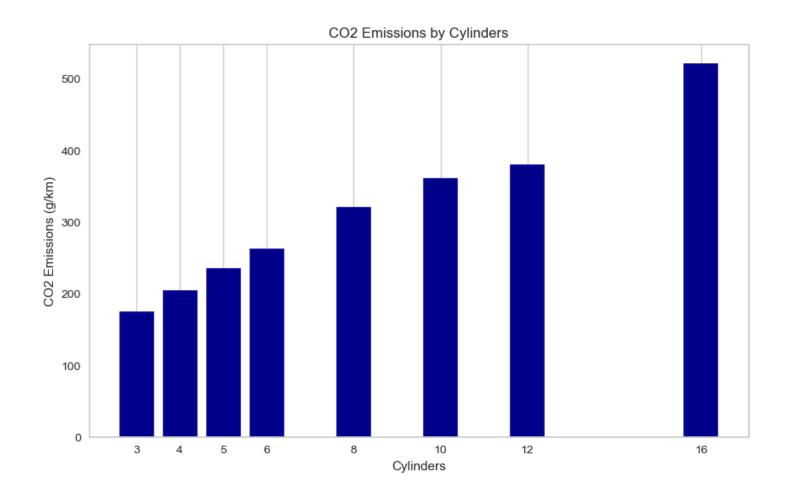
Analysis of Relationship Between Cylinders and CO2 Emissions

The bar chart clearly shows that as the number of cylinders in a vehicle increases, so do its CO2 emissions. This happens because engines with more cylinders burn more fuel and release more greenhouse gases.

This relationship is because engines with more cylinders can burn more fuel each time they work, making more CO2. Also, engines with more cylinders usually make more power, so they need more fuel, which leads to higher emissions.

This has important implications for designing cars and environmental rules. It reminds us to think about cylinder size when we talk about how much pollution cars make. It also shows the need for new technology and rules to make cars cleaner.

In short, the bar chart highlights how cylinder size affects CO2 emissions from cars, reminding us to consider engine features when discussing vehicle pollution and ways to make cars better for the environment.



Analysis of CO2 Emissions by Transmission Type

The analysis of CO2 emissions by transmission type shows some interesting findings about how different transmissions affect the environment.

Automatic (A) Cars: Automatic cars have the highest CO2 emissions among the types considered. These transmissions use torque converters and hydraulic systems, making them less efficient and leading to more fuel consumption.

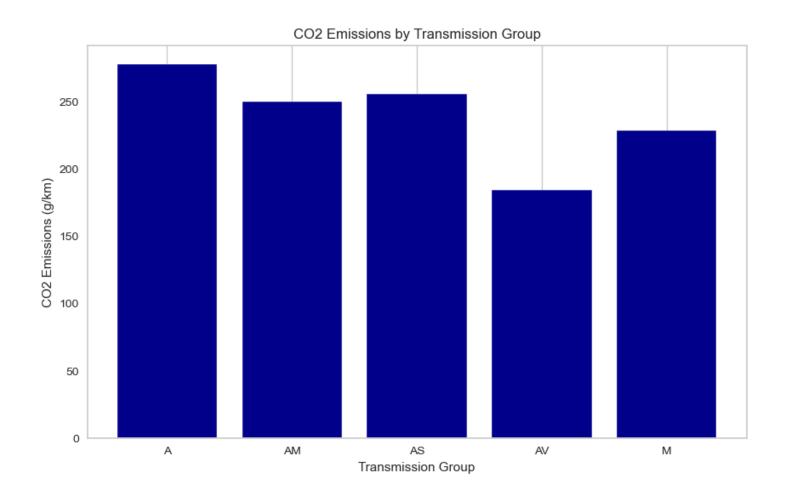
Automatic with Select Shift (AS) Cars: These cars come next in CO2 emissions after automatic ones. They offer drivers the option to manually shift gears, but still rely on traditional automatic transmission technology, resulting in higher emissions.

Automatic Manual (AM) Cars: Cars with automatic manual transmissions rank third in CO2 emissions. Although they're more efficient than traditional automatics, their complex operation can still lead to higher emissions.

Manual (M) Cars: Manual cars, where the driver shifts gears manually, have lower CO2 emissions compared to automatics. They're simpler and offer better fuel efficiency, making them more environmentally friendly.

Continuously Variable (AV) Cars: Cars with continuously variable transmissions (CVT) have the lowest CO2 emissions. CVTs use a system of pulleys and belts to optimize engine efficiency, resulting in reduced fuel consumption and emissions.

In conclusion, the analysis shows that transmission type has a significant impact on CO2 emissions. While automatic transmissions tend to produce more emissions, manual and continuously variable transmissions offer greener alternatives, with manual transmissions having the lowest emissions.



Analysis of CO2 Emissions by Fuel Type

X = regular gasoline
Z = premium gasoline
D = diesel
E = ethanol (E85)
N = natural gas

Premium Gasoline: Premium gasoline releases the most CO2 among the fuels we looked at. This is because it's refined and has a higher octane rating, leading to more CO2 per unit of fuel burned. Although it might improve performance for some cars, it also adds more to greenhouse gas emissions compared to other fuels.

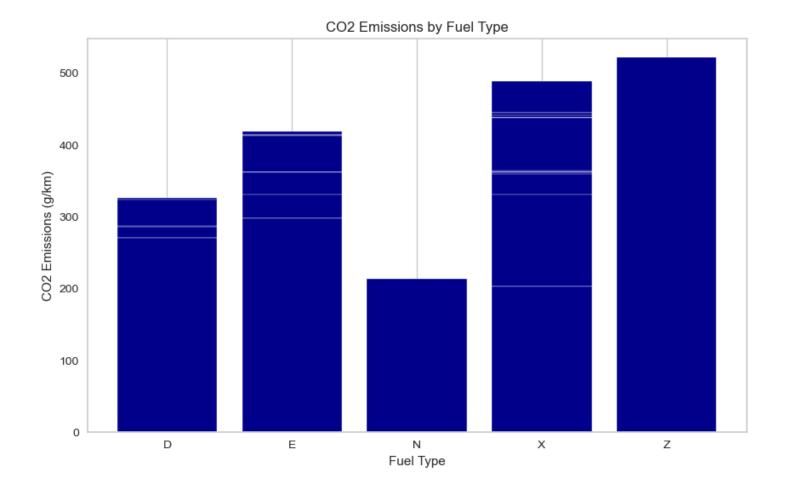
Regular Gasoline Regular gasoline comes next after premium gasoline in CO2 emissions. It's used in most cars and has a lower octane rating. While it emits less CO2 than premium gasoline, it's still a big source of greenhouse gases, especially in less efficient engines.

Ethanol (E85): Ethanol, especially E85, emits less CO2 than gasoline. It's made from renewable sources like corn or sugarcane and burns cleaner. But because it has less energy per volume than gasoline, cars not made for ethanol might use more fuel and emit more CO2.

Diesel: Diesel fuel is fourth in CO2 emissions among the fuels we studied. Diesel engines are efficient but produce more nitrogen oxides and particles than gasoline engines. Though they've gotten cleaner with technology, they still add to CO2 emissions, though not as much as gasoline.

Natural Gas: Natural gas has the lowest CO2 emissions among the fuels we looked at. It's mostly methane and burns cleaner than liquid fuels, releasing less CO2 per unit of energy. Cars using compressed or liquefied natural gas are cleaner and emit fewer pollutants than gasoline or diesel vehicles.

To sum up, our analysis shows different fuels have different CO2 emissions. While premium and regular gasoline add the most, ethanol, diesel, and natural gas offer cleaner options, with natural gas being the cleanest among them.



Conclusion

In conclusion, the analysis reveals valuable insights into the relationship between vehicle characteristics and CO2 emissions. Engine size, cylinder count, and transmission type all play significant roles in determining the environmental impact of vehicles. Larger engines and automatic transmissions tend to result in higher CO2 emissions, while manual transmissions and smaller engines contribute less to greenhouse gas pollution. Additionally, the type of fuel used also influences emissions, with premium gasoline and diesel being the highest contributors and natural gas being the cleanest option. These findings underscore the importance of considering various factors when designing and regulating vehicles to minimize their environmental footprint. Efforts to promote fuel-efficient technologies and cleaner fuel alternatives are essential for mitigating the adverse effects of vehicle emissions on the environment and human health.

Regression Model Preparation

Assumptions of linear regression

Linearity: The relationship between variables is linear.

Independence: Observations are independent of each other.

Homoscedasticity: Residuals have constant variance.

Normality of Residuals: Residuals are normally distributed.

No Perfect Multicollinearity: No perfect relationship among predictors.

No Autocorrelation: Residuals are not correlated with each other.

Additivity: Effects of predictors on the response are additive.

Zero Mean Residuals: Mean of residuals is zero.

Descriptive Statistics

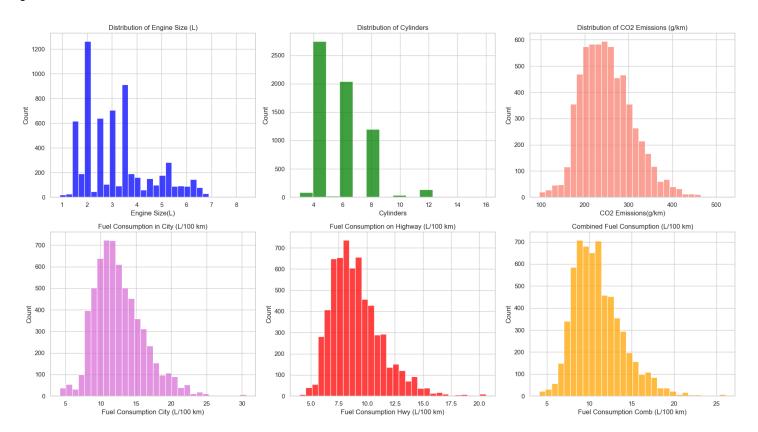
	Engine Size(L)	Cylinders	Fuel Consumption City (L/100 km)	Fuel Consumption Hwy (L/100 km)	Fuel Consumption Comb (L/100 km)	Fuel Consumption Comb (mpg)	CO2 Emissions(g/km)
count	7385.000000	7385.000000	7385.000000	7385.000000	7385.000000	7385.000000	7385.000000
mean	3.160068	5.615030	12.556534	9.041706	10.975071	27.481652	250.584699
std	1.354170	1.828307	3.500274	2.224456	2.892506	7.231879	58.512679
min	0.900000	3.000000	4.200000	4.000000	4.100000	11.000000	96.000000
25%	2.000000	4.000000	10.100000	7.500000	8.900000	22.000000	208.000000
50%	3.000000	6.000000	12.100000	8.700000	10.600000	27.000000	246.000000
75%	3.700000	6.000000	14.600000	10.200000	12.600000	32.000000	288.000000
max	8.400000	16.000000	30.600000	20.600000	26.100000	69.000000	522.000000

- Engine Size (L): The engine size ranges from 0.9 to 8.4 liters, with a mean value of approximately 3.16 liters. The standard deviation is around 1.35, indicating some variability in engine sizes.
- **Cylinders**: The number of cylinders ranges from 3 to 16, with a mean of approximately 5.62 cylinders. The standard deviation is approximately 1.83, suggesting some variability in cylinder counts.
- Fuel Consumption City (L/100 km): The average fuel consumption in the city is about 12.56 liters per 100 kilometers, with a standard deviation of approximately 3.5. The values range from 4.2 to 30.6 liters per 100 kilometers.
- Fuel Consumption Hwy (L/100 km): The average fuel consumption on the highway is approximately 9.04 liters per 100 kilometers, with a standard deviation of around 2.22. The values vary between 4.0 and 20.6 liters per 100 kilometers.
- Fuel Consumption Comb (L/100 km): The combined fuel consumption averages around 10.98 liters per 100 kilometers, with a standard deviation of about 2.89. The range extends from 4.1 to 26.1 liters per 100 kilometers.
- Fuel Consumption Comb (mpg): The combined fuel consumption in miles per gallon (mpg) has a mean of approximately 27.48, with a standard deviation of around 7.23. Values range from 11 to 69 mpg.
- CO2 Emissions (g/km): The CO2 emissions range from 96 to 522 grams per kilometer, with a mean value of approximately 250.58 g/km. The standard deviation is about 58.51, indicating some variability in emission levels.

These statistics provide an overview of the central tendency, variability, and range of each numerical variable in your dataset. They can help in understanding the distribution and characteristics of the data.

Distribution Analysis

Engine Size, Cylinders, Fuel Consumption, and CO2 Emissions: All these key variables show a skewed distribution. This indicates variability in vehicle types, with some having significantly larger engines, higher fuel consumption, and greater CO2 emissions.



Outliers

The boxplots for the selected independent variables — 'Engine Size(L)', 'Cylinders', 'Fuel Consumption City (L/100 km)', and 'Fuel Consumption Hwy (L/100 km)' — indicate the presence of outliers in the dataset:

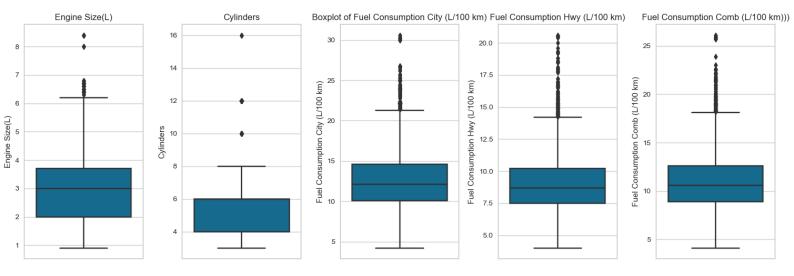
Engine Size (L): There are several outliers indicating engines significantly larger than the median engine size.

Cylinders: The presence of outliers suggests some vehicles have a much higher number of cylinders than the typical range.

Fuel Consumption City (L/100 km): There are outliers indicating that some vehicles have significantly higher fuel consumption in the city than most.

Fuel Consumption Hwy (L/100 km): Similar to city fuel consumption, highway fuel consumption also shows outliers, indicating higher-than-typical fuel usage.

Fuel Consumption Comb (L/100 km): The Fuel Consumption Comb (L/100 km) variable represents a combined rating, with 55% of the consumption attributed to city driving and 45% to highway driving. It has also shows outliers, indicating higher-than-typical fuel usage.



Correlation Analysis

Engine size: 0.85 correlated

Cylinders: 0.83 correlated

Fuel Consumption City(100L/km): 0.92 correlated

• Fuel Consumption Hwy(100L/km): 0.88 correlated

Fuel Consumption Comb(100g/km): 0.92 correlated

Fuel Consumption Comb(mpg): -0.91correlated

These correlation coefficients suggest strong positive relationships between each of these features and CO2 emissions. Specifically:

- Engine size and cylinders: Larger engine sizes and higher cylinder counts tend to result in higher CO2 emissions, as they typically indicate more powerful and fuel-consuming vehicles.
- Fuel consumption City & Hwy(100L/km): Both city and highway fuel consumption rates are strongly correlated with CO2 emissions. Higher fuel consumption rates lead to increased CO2 emissions, as more fuel is burned to power the vehicle.
- Fuel consumption combined(100g/km): This is the combined fuel consumption rate in (100g/km), which accounts for both city and highway driving conditions. As expected, it is also highly correlated with CO2 emissions.
- Fuel consumption combined(mpg): This is the combined fuel consumption rate in (100g/km), which accounts for both city and highway driving conditions. It is negatively correlated with CO2 emissions.

	Correlation Heatmap								1.00
Engine Size(L)	1.00	0.93	0.83	0.77	0.82	-0.76	0.85		- 0.75
Cylinders	0.93	1.00	0.80	0.72	0.78	-0.72	0.83		- 0.50
Fuel Consumption City (L/100 km)	0.83	0.80	1.00	0.95	0.99	-0.93	0.92		- 0.25
Fuel Consumption Hwy (L/100 km)	0.77	0.72	0.95	1.00	0.98	-0.89	0.88		- 0.00
Fuel Consumption Comb (L/100 km)	0.82	0.78	0.99	0.98	1.00	-0.93	0.92		0.25
Fuel Consumption Comb (mpg)	-0.76	-0.72	-0.93	-0.89	-0.93	1.00	-0.91		0.50
CO2 Emissions(g/km)	0.85	0.83	0.92	0.88	0.92	-0.91	1.00		- -0.75
	Engine Size(L)	Cylinders	Fuel Consumption City (L/100 km)	Fuel Consumption Hwy (L/100 km)	Fuel Consumption Comb (L/100 km)	Fuel Consumption Comb (mpg)	CO2 Emissions(g/km)		

As we Seen All Independent Variables are Correlated with our Dependent Variable CO2 Emissions(g/km)

Positive Correlated: Engine Size(L), Cylinders, Fule Consumption City(L/100km), Fule Consumption Hwy(L/100km) Negative Correlated: Fule Consumption Comb(L/100km)

Regression Model

I'll create two models to predict CO2 emissions:

1- Using All Features

Include all features in your dataset, including categorical ones like 'Make', 'Model', 'Vehicle Class', 'Transmission', 'Fuel Type', and numerical ones like 'Engine Size(L)', 'Cylinders', 'Fuel Consumption

City (L/100 km)', and 'Fuel Consumption Hwy (L/100 km)'.

2- Using Most Correlated Features

Engine Size(L), Cylinders, Fule Consumption City(L/100km), Fule Consumption Hwy(L/100km)

I'm excluding the 'Fuel Consumption Comb (L/100 km)' feature because it represents a combined rating of 55% city and 45% highway fuel consumption, presented in liters per 100 kilometers. It's not a direct combination of both features. Additionally, 'Fuel Consumption Comb (mpg)' is essentially the same as 'Fuel Consumption Comb (mpg)', with only a change in units.

Model 1

Uses all features from the original dataset.

Prepare Data:

Include all features in dataset, including categorical ones like 'Make', 'Model', 'Vehicle Class', 'Transmission', 'Fuel Type', and numerical ones like 'Engine Size(L)', 'Cylinders', 'Fuel Consumption City (L/100 km)', 'Fuel Consumption Hwy (L/100 km)', and 'Fuel Consumption Comb (mpg)'.

Encode Categorical Features:

• These Columns are categorical ('Make', 'Model', 'Vehicle Class', 'Transmission', 'Fuel Type'). Use one-hot encoding to convert categorical variables into numerical representations. This ensures that the model can understand them.

Train the Model:

• Fit a linear regression model to the training data using all features, including both the encoded categorical features and the scaled numerical features.

Evaluate the Model:

 Assess the performance of the model on the testing data to see how well it predicts CO2 emissions.

Result

Root Mean Squared Error (RMSE): 0.00012186278387644087

It appears that the Root Mean Squared Error (RMSE) for your Model 1 is approximately 0.00012186278387644087. This metric indicates the average deviation of the predicted CO2 emissions from the actual values. A lower RMSE suggests that the model's predictions are closer to the actual values, indicating better performance.

Model 2

Utilizes only the most correlated numerical features.

Multicollinearity

```
Feature VIF

Engine Size(L) 28.396042

Cylinders 183.084254

Fuel Consumption City (L/100 km) 881.394942

Fuel Consumption Hwy (L/100 km) 764.106286
```

Based on the VIF values you provided, it appears that there are high levels of multicollinearity present in your features. Specifically:

Fuel Consumption City (L/100 km) and Fuel Consumption Hwy (L/100 km) have VIF values exceeding 10, indicating strong multicollinearity. This suggests that these two features are highly correlated with each other.

Cylinders also has a very high VIF value, indicating a significant degree of multicollinearity. This suggests that the number of cylinders is strongly correlated with other features in the dataset.

Engine Size(L) has a VIF value above 10, indicating moderate multicollinearity. While not as severe as the other features, it still suggests some level of correlation with other features.

When multicollinearity is present, it can lead to unstable coefficient estimates and difficulty in interpreting the model. It's essential to address multicollinearity before proceeding with your analysis.

Dealing with Multicollinearity

Due to the presence of multicollinearity between the 'Fuel Consumption City (L/100 km)' and 'Fuel Consumption Hwy (L/100 km)'

features, I have opted to utilize the 'Fuel Consumption Comb (L/100 km)' feature as a solution. By incorporating the combined fuel consumption, I aim to mitigate the issue of multicollinearity and ensure the robustness of the model.

- Step 1: We apply the np.log2 transformation to the specified columns to handle outliers.
- Step 2: We split the data into independent (X) and dependent (y) variables.
- Step 3: We use StandardScaler to standardize the features (independent variables).
- Step 4: We split the data into training and testing sets using train_test_split.
- Step 5: We fit a linear regression model to the training data.

Step 6: We evaluate the model's performance on both the training and testing data by calculating the R-squared scores.

Result

Training R^2 Score: 0.9110462502540878

Testing R^2 Score: 0.9078127033934025

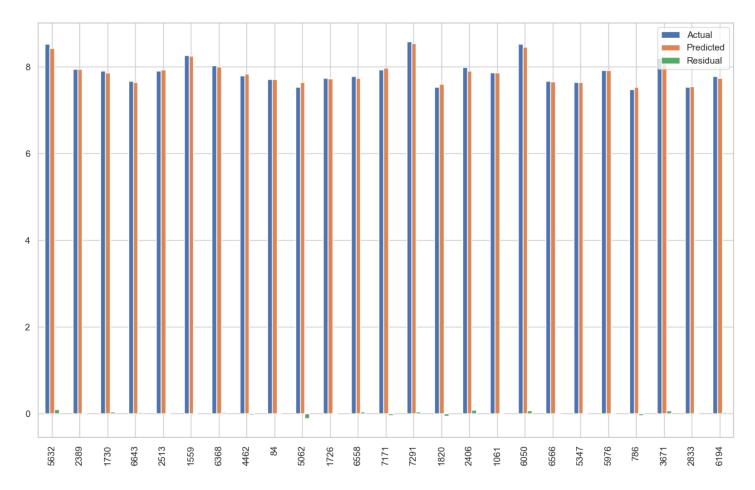
- The training R^2 score of approximately 0.911 suggests that your model explains about 91.10% of the variance in the training data. This indicates a strong fit of the model to the training data.
- The testing R^2 score of approximately 0.908 suggests that your model explains about 90.78% of the variance in the testing data. This indicates that your model generalizes well to unseen data, as it performs similarly well on the testing data as it did on the training data.

Overall, these R^2 scores indicate that your linear regression model is performing well and is a good fit for your data. However, it's essential to consider other metrics and perform further validation to ensure the robustness of your model.

Residuals

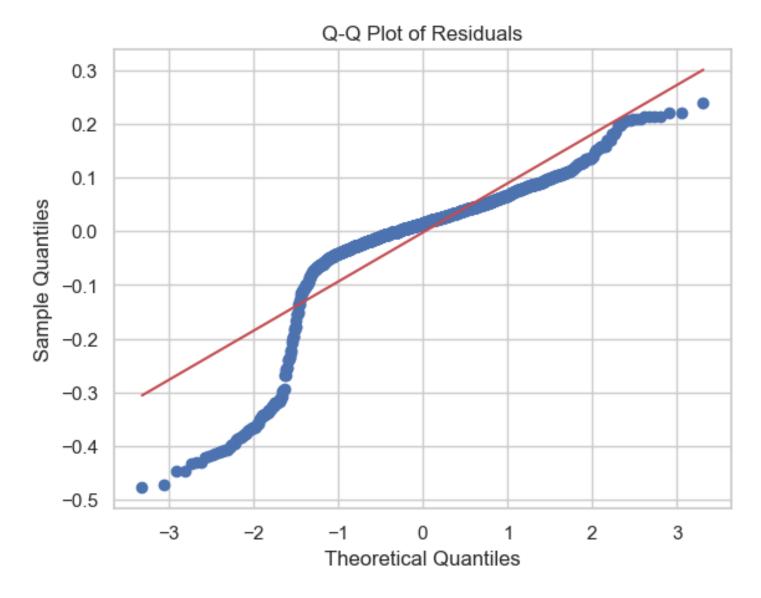
Bar chart represents the actual, predicted, and residual values for each observation in the testing dataset. The actual values are shown in blue, predicted values in orange, and residuals in green.

Residuals represent the differences between the actual and predicted values, with positive residuals indicating that the predicted values are higher than the actual emissions and negative residuals indicating the opposite. Overall, the residuals are relatively small, suggesting that the linear regression model provides reasonably accurate predictions of CO2 emissions.



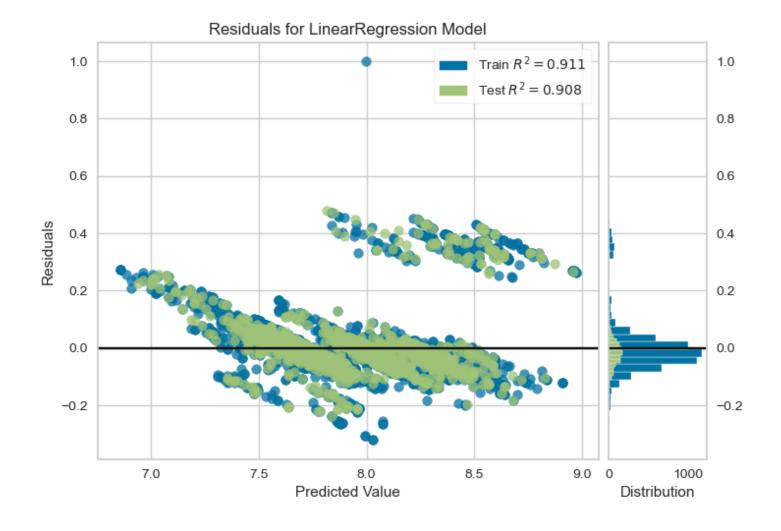
Q-Q Plot of Residuals

The Q-Q plot of residuals visually depicts discrepancies between observed and expected quantiles, indicating deviations from the linear regression line. Specifically, some residual values fall above the line, indicating that the corresponding predicted values are higher than expected based on the regression model. Conversely, other residual values fall below the line, suggesting that the predicted values are lower than expected. These deviations highlight areas where the linear regression model may not fully capture the variability in the data and may warrant further investigation or model refinement.



Residuals Plot

In the Residuals Plot, the blue markers represent the training data, while the green markers represent the testing data. The plot reveals that many residuals are above the zero line, indicating that the model tends to overestimate the target variable in some instances. Despite this, the model performs well overall, as evidenced by the high training R^2 score of 0.911 and testing R^2 score of 0.908. These scores indicate that the model explains a significant portion of the variance in the data and generalizes well to unseen data.



Hyperparameter Tuning

Hyperparameter tuning is a process used to optimize the performance of machine learning models by adjusting the hyperparameters, which are parameters set before the learning process begins. This optimization involves systematically testing different hyperparameter configurations and selecting the combination that yields the best performance metrics, such as accuracy or R^2 score. By fine-tuning hyperparameters, we can enhance the model's predictive capabilities and improve its ability to generalize to unseen data.

Let's conduct hyperparameter tuning to evaluate whether it improves the model fitting. In linear regression, we have two regularization techniques:

Linear Regression Hyperparameter Tuning

Hyperparameter tuning for Linear Regression typically involves optimizing parameters such as the regularization strength (alpha) in regularization techniques.

Ridge regression

Lasso regression

We will apply each of them separately to assess their impact on the model performance.

Lasso Regression & Ridge Regression

- We import Ridge and Lasso from sklearn.linear_model.
- We fit a Ridge regression model (ridge_model) and a Lasso regression model (lasso_model) to the training data (X_train, y_train) using default alpha values.
- We evaluate both models on the training and testing data and print their R^2 scores.
- You can adjust the alpha parameter for Ridge and Lasso to control regularization strength.

Lasso Regression

Lasso Training R^2 Score: 0.0

Lasso Testing R^2 Score: -0.0010657095620385526

The training R^2 score of 0.0 indicates that the Lasso regression model explains none of the variance in the training data. This suggests that the Lasso model is not able to capture the relationship between the independent and dependent variables in the training data.

The testing R^2 score of approximately -0.001 suggests that the Lasso regression model performs even worse on the testing data, explaining slightly less than 0% of the variance. A negative R^2 score indicates that the model is performing worse than a horizontal line fitting the data, essentially providing no predictive power.

These results indicate that the Lasso regression model is not suitable for dataset.

Ridge Regression

Ridge Training R^2 Score: 0.9110462015627051

Ridge Testing R^2 Score: 0.9078143172355687

The Ridge regression model achieved a training R^2 score of approximately 0.911 and a testing R^2 score of around 0.908. This indicates that the model performs well in both training and testing datasets, with slightly higher performance on the training set compared to the testing set.

Comparison

Linear Regression

Training R^2 Score: 0.9110462502540878

Testing R^2 Score: 0.9078127033934025

Ridge Regression

Training R^2 Score: 0.9110462015627051

Testing R^2 Score: 0.9078143172355687

The very slight difference in R^2 scores between the linear regression and Ridge regression models suggests that the regularization introduced by the Ridge regression did not lead to significant improvements in performance for this dataset. However, the Ridge regression model might still offer benefits such as increased robustness to multicollinearity or improved generalization to unseen data in other scenarios or datasets.

Comparison of Model 1 & Model 2

The Root Mean Squared Error (RMSE) for Model 1 with all features is 0.00012186278387644087, indicating a very small error in prediction.

For Model 2, which includes only 'Engine Size(L)', 'Cylinders', and 'Fuel Consumption Comb (L/100 km)', the Training R^2 Score is 0.911 and the Testing R^2 Score is 0.908.

Comparing the two models, Model 2 achieves slightly lower performance in terms of R^2 scores compared to Model 1, suggesting that the additional features in Model 1 contribute to better predictive accuracy. However, Model 2 still performs remarkably well with a high R^2 score, indicating that these three features alone can explain a significant portion of the variance in CO2 emissions.

Recommendation

Choose cars with smaller engines: When buying a car, consider one with a smaller engine size and fewer cylinders. These cars usually use less fuel and produce fewer emissions.

Use alternative fuels: Look for options like ethanol, natural gas, or electric cars. These fuels are cleaner and release fewer harmful gases into the air compared to traditional gasoline or diesel.

Use public transportation: Whenever possible, use buses, trains, or other forms of public transportation instead of driving alone. This reduces the number of cars on the road and lowers overall emissions.

Support stricter rules: Support laws that make car manufacturers produce vehicles with lower emissions. This encourages the development of cleaner technologies and reduces pollution.

Drive responsibly: Practice eco-friendly driving habits like keeping your tires properly inflated, avoiding unnecessary idling, and driving smoothly. These habits help save fuel and reduce emissions.

By following these suggestions, we can all play a part in reducing CO2 emissions and protecting our environment.

Conclusion

In conclusion, our analysis of CO2 emissions from vehicles has provided valuable insights into the factors influencing emissions and potential mitigation strategies. We found that engine size, cylinder count, and fuel consumption are significant determinants of CO2 emissions, with larger engines and higher fuel consumption contributing to greater emissions.

Furthermore, our examination of different transmission types revealed varying levels of CO2 emissions, highlighting the importance of considering transmission technology in efforts to reduce emissions. We also explored the impact of different fuel types on emissions, with premium gasoline and diesel fuel showing higher emission levels compared to alternatives like ethanol and natural gas.

Overall, our findings underscore the urgent need for action to address vehicle emissions and combat climate change. By promoting fuel-efficient technologies, supporting alternative fuels, and advocating for stricter emissions regulations, we can work towards a cleaner and more sustainable future for generations to come. It is imperative that individuals, businesses, and policymakers alike prioritize environmental stewardship and take proactive steps to mitigate the adverse effects of CO2 emissions on our planet.

Through collective effort and commitment to sustainable practices, we can make significant strides in reducing CO2 emissions and protecting the health of our planet for current and future generations.