**VEHICLE CHARACTERISTICS AND CO2 EMISSIONS ANALYSIS**

**Table of Contents**

[1. Introduction 3](#_Toc186221768)

[1.1 Problem statement and research motivation 3](#_Toc186221769)

[1.2 Data set 3](#_Toc186221770)

[1.3 Research question 3](#_Toc186221771)

[1.4 Null hypothesis and alternative hypothesis 3](#_Toc186221772)

[2. Background Research 4](#_Toc186221773)

[2.1 Research papers 4](#_Toc186221774)

[2.2 Research gap 4](#_Toc186221775)

[3. Visualization 5](#_Toc186221776)

[3.1 Plot for the RQ output of an R script 5](#_Toc186221777)

[4. Analysis 7](#_Toc186221778)

[4.1 Statistical test used to test the hypotheses and output 7](#_Toc186221779)

[4.2 The null hypothesis rejected /not rejected 7](#_Toc186221780)

[5. Evaluation – Group’s Experience at 7COM1079 8](#_Toc186221781)

[5.1 What went well 8](#_Toc186221782)

[5.2 Points for Improvement 8](#_Toc186221783)

[5.3 Group’s Time Management 8](#_Toc186221784)

[5.4 Project’s Overall Judgement 8](#_Toc186221785)

[5.5 Comment on GitHub Log Output 8](#_Toc186221786)

[6. Conclusion 8](#_Toc186221787)

[6.1 Results Explained 9](#_Toc186221788)

[6.2 Interpretation of the result 9](#_Toc186221789)

[6.3 Future Work and Limitations of Your Study 9](#_Toc186221790)

[References 10](#_Toc186221791)

[Appendix 11](#_Toc186221792)

**List of Figure**

[Figure 1 : Correlation Heatmap 12](#_Toc186221813)

[Figure 2 : Linear Regression 12](#_Toc186221814)

[Figure 3 : Scatter plot: CO2 Rating Vs CO2 Emmision 13](#_Toc186221815)

[Figure 4 : Scatter plot: smog rating vs CO2 emissions 13](#_Toc186221816)

[Figure 5 : scatter plot : combined fuel consumption vs CO2 emmisions 14](#_Toc186221817)

[Figure 6 : Fuel consumption by vehicle class 14](#_Toc186221818)

[Figure 7 : Average fuel consumption by fuel type 15](#_Toc186221819)

[Figure 8 : Histogram ( Distribution of C02 emissions ) 15](#_Toc186221820)

[Figure 9 : Histogram (Distogram of engine size ) 16](#_Toc186221821)

[Figure 10 : Histogram ( Distribution of transmission types) 16](#_Toc186221822)

# 1. Introduction

## Problem statement and research motivation

The pressure on the automotive industry to deplete its environmental footprints has been rising recently. Manufacturing, fuel consumption, and related emissions have been a key source of this impact. Knowledge of the factors controlling fuel consumption and emissions enlightens the consumer on the best car to purchase and directs car manufacturers to create environmentally friendly automobiles (Wang and Zhang, 2020). This project identifies the correlation between vehicle properties, including engine capacity, fuel kind, transmission type, fuel consumption, and CO2 emissions.

## 1.2 Data set

The work employs the Fuel\_Consumption\_Ratings.csv data file that contains information about several brands of cars and their fuel efficiency ratings, CO2 emission levels, and other related parameters. This dataset has vehicle model year, make model, class, engine size, no of the cylinder, transmission, fuel, city L/100km, highway L/100km, combined L/100km, combined CO2 emission (mpg), CO2 emission (g/km), CO2 rating and smog rating. Following attrition, there are 1047 entries with 15 attributes in the dataset. However, due to the presence of null values in the dataset, we have excluded null values, and the dataset contains 1022 entries for its analysis.

## Research question

* What kind of vehicle characteristics (CO2 emission, smog index, engine displacement, fuel type, drive type, overall fuel efficiency) are connected with the emission of CO2, and how strong are those connections.?

## 1.4 Null hypothesis and alternative hypothesis

**Hypothesis for the Correlation Analysis (CO2 Rating and CO2 Emissions):**

**Null Hypothesis (H0):** No significant difference exists between CO2 Rating and CO2 Emissions.

**Alternative Hypothesis (H1):** There is a significant difference between CO2 Rating and CO2 Emissions

**Hypothesis for the regression analysis**

**Null Hypothesis (H0):** The little change between the independent variables of (CO2.Rating, Smog. Rating, Fuel.Consumption.Combined..L.100km.) to the dependent variable of (CO2.Emissions..g.km.) support this inference.

**Alternative Hypothesis (H1):** The comparison of (CO2.Rating, Smog . Rating, and FuelHillary Rating) indicated a statistically significant difference with a representative p less than 0.05. Consumption.Combined. .L.100km G.102 and (CO2.Emissions..g.km.).

# 2. Background Research

## 2.1 Research papers

Huang et al. (2019) assessed the real drive cycles of automobiles with internal combustion engines (ICEs) and hybrid electric vehicles (HEVs). Despite using 23% to 49% less gasoline, HEVs actually increased CO emissions rather than reducing HC emissions. Although the "hybridization" process reduced fuel use, the strategy did not achieve the desired improvement in urban air quality.

The short-term effects of eco-driving on fuel consumption and driver stress are examined by Lois et al. (2019). Their analysis of 1156 trips with 24 drivers shows that while driving behaviors and outside factors significantly affect fuel consumption, eco-driving has no discernible effect on stress levels.

Mubarak et al., (2020) researched the comparative analysis of HEVs and conventional vehicles considering fuel economy and emissions based on Iraqi drive cycles. Studying performance differences between HEVs and traditional cars, especially during real-world dynamic tests, is the focus of research made by the authors in their article published in the Journal of Mechanical Engineering Research and Developments.

## 2.2 Research gap

The research question is important given that it fits within a significant subfield of environmental sustainability. Whereas earlier works focus on specific aspects of fuel usage and emissions separately, providing a detailed correlation between vehicle parameters and CO2 emissions is critical to define further opportunities for efficiency improvement in car manufacturing. The present research extends knowledge on the specified subject by combining perspectives concurrently with a vast database, which may help consumer behavior and the atmosphere to evolve more sustainably.

# 3. Visualization

## 3.1 Plot for the RQ output of an R script

# Loading necessary libraries

library(ggplot2)

library(dplyr)

# Checking correlation between CO2 Rating and CO2 Emissions

correlation\_CO2 <- cor(FuelData$CO2.Rating, FuelData$CO2.Emissions..g.km., use = "complete.obs")

print(paste("Correlation between CO2 Rating and CO2 Emissions:", round(correlation\_CO2, 2)))

# Correlation matrix for selected columns

cor\_matrix <- cor(FuelData %>%

select(CO2.Rating, CO2.Emissions..g.km., Smog.Rating, Fuel.Consumption.Combined..L.100km.),

use = "complete.obs")

print("Correlation Matrix:")

print(round(cor\_matrix, 2)) # Rounded for clarity

# Building a linear regression model

model <- lm(CO2.Emissions..g.km. ~ CO2.Rating + Smog.Rating + Fuel.Consumption.Combined..L.100km., data = FuelData)

print(summary(model)) # Displaying the model summary

# Plotting regression line for CO2 Rating vs CO2 Emissions

ggplot(FuelData, aes(x = CO2.Rating, y = CO2.Emissions..g.km.)) +

geom\_point(color = "blue", alpha = 0.6) +

geom\_smooth(method = "lm", color = "red", se = FALSE) +

labs(

title = "Regression: CO2 Emissions vs CO2 Rating",

x = "CO2 Rating",

y = "CO2 Emissions (g/km)"

) +

theme\_minimal()

# Scatter plot for CO2 Rating vs CO2 Emissions

ggplot(FuelData, aes(x = CO2.Rating, y = CO2.Emissions..g.km.)) +

geom\_point(color = "red", alpha = 0.7) +

labs(

title = "Scatter Plot: CO2 Rating vs CO2 Emissions",

x = "CO2 Rating",

y = "CO2 Emissions (g/km)"

) +

theme\_minimal()

# Scatter plot for Smog Rating vs CO2 Emissions

ggplot(FuelData, aes(x = Smog.Rating, y = CO2.Emissions..g.km.)) +

geom\_point(color = "green", alpha = 0.7) +

labs(

title = "Scatter Plot: Smog Rating vs CO2 Emissions",

x = "Smog Rating",

y = "CO2 Emissions (g/km)"

) +

theme\_minimal()

# Scatter plot for Fuel Consumption vs CO2 Emissions

ggplot(FuelData, aes(x = Fuel.Consumption.Combined..L.100km., y = CO2.Emissions..g.km.)) +

geom\_point(color = "purple", alpha = 0.7) +

labs(

title = "Scatter Plot: Fuel Consumption vs CO2 Emissions",

x = "Fuel Consumption (L/100km)",

y = "CO2 Emissions (g/km)"

) +

theme\_minimal()

# 4. Analysis

## 4.1 Statistical test used to test the hypotheses and output

* Correlation
* Regression

## 4.2 The null hypothesis rejected /not rejected

* It states that the slope from the line equals the correlation coefficient for CO2Rating and CO2Emissions, respectively. Hence, the overall null hypothesis is rejected
* Fails to reject the null hypothesis; it implies that the overall regression model has some worth in predicting the amount of variability in CO2 emissions using the selected independent variables.

# 5. Evaluation – Group’s Experience at 7COM1079

## 5.1 What went well

Handling environmental issues, the research incorporates appropriate data and statistical tools such as regression and correlation (Rosero et al., 2021). The key hypotheses are clear, the presented visualizations are understandable, and the results satisfactory; the authors are also honest about the limitations of their work.

## 5.2 Points for Improvement

The project’s findings might be better if hypothesis testing results were elaborated, more advanced techniques were applied to the regression model, assumptions were discussed, other variables were investigated, different statistical approaches were used, or heatmaps were improved.

## 5.3 Group’s Time Management

They could manage their study time by ensuring they had a thorough group research. Nevertheless, more focus on additional model refinement, the nature of the data, and the discussion of statistical significance could improve the study.

## 5.4 Project’s Overall Judgement

It is possible to note that the project demonstrates the use of statistical methods to study vehicle characteristics and CO2 emissions (Chong et al., 2020). This research is sound, has a good background for subsequent research, and meets basic requirements for improvement with specific guidance.

## 5.5 Comment on GitHub Log Output

The lack of the GitHub log precludes analyzing participation, code branching, and management; it is also unclear how and by whom the code was updated, added to, and improved.

# 6. Conclusion

## 6.1 Results Explained

The results reveal that with the rise in fuel economy, CO2 emission also rises, and emissions are reduced with the increase in CO2 and smog levels. High levels of linear dependences between the data on the size of engines, fuel consumption, and CO2 emissions underline their influence. These attributes need to be optimized to lessen vehicle emissions and reduce their environmental impact.

## 6.2 Interpretation of the result

The result suggests that the vehicles with higher EPU and emissions levels release lower CO2 emissions than their counterparts. This underscores the fact that improvements to efficiency in the production of fuel and reduction of emissions are very efficient (Sun et al., 2019). However, fuel consumption is an equal factor that triggers actual emissions production. The analysis, therefore, indicates that there is parity in consumers' behavior in their choice of vehicle and its influence on CO2 emissions.

## 6.3 Future Work and Limitations of Your Study

**Future work**

Subsequent studies are also planned to apply more complex regression models to investigate potential nonlinear associations between the variables (Butt and Singh, 2023). A predictive model of CO2 emissions will again be created, and its performance will be tested against other approaches. Including more inputs, different vehicles, and different driving environments for the tested cars will enhance the reliability of the findings and the understanding of the emissions behavior they provide.

**Limitations**

The study is also subject and incomplete or limited because it overlooked other parameters such as driving behavior or environmental condition of the automobile or transport vehicles, which can be only inferred from the given dataset attributes (Hien and Kor, 2022). They could have used a newer dataset since the one used in the study is rather old. Finally, low representation of some vehicle models may reduce the generalization of the results concerning all automobiles.

# References

Butt, M.H. and Singh, J.G., 2023. Factors affecting electric vehicle acceptance, energy demand, and CO2 emissions in Pakistan. *Green Energy and Intelligent Transportation*, *2*(3), p.100081.

Chong, H.S., Kwon, S., Lim, Y. and Lee, J., 2020. Real-world fuel consumption, gaseous pollutants, and CO2 emission of light-duty diesel vehicles. Sustainable Cities and Society, 53, p.101925.

Hien, N.L.H. and Kor, A.L., 2022. Analysis and prediction model of light-duty vehicles' fuel consumption and carbon dioxide emissions. *Applied Sciences*, *12*(2), p.803.

Huang, Y., Surawski, N.C., Organ, B., Zhou, J.L., Tang, O.H. and Chan, E.F., 2019. Fuel consumption and emissions performance under actual driving: Comparison between hybrid and conventional vehicles. Science of the Total Environment, 659, pp.275-282.

Lois, D., Wang, Y., Boggio-Marzet, A. and Monzon, A., 2019. Multivariate analysis of fuel consumption related to eco-driving: Interaction of driving patterns and external factors. Transportation Research Part D: Transport and Environment, 72, pp.232-242.

Mubarak, L.M., Al-Samari, A., Alazawi, D.A. and Fadel, M., 2020. Comparison study of fuel consumption and emissions of HEVs and conventional vehicles in Iraq using real-world cycle. Journal of Mechanical Engineering Research and Developments, 43(5), pp.185-203.

Rosero, F., Fonseca, N., López, J.M. and Casanova, J., 2021. Effects of passenger load, road grade, and congestion level on real-world fuel consumption and emissions from compressed natural gas and diesel urban buses. Applied Energy, 282, p.116195.

Sun, H., Li, M. and Xue, Y., 2019. Examining the factors influencing transport sector CO2 emissions and their efficiency in central China. *Sustainability*, *11*(17), p.4712.

Wang, H. and Zhang, X., 2020. The spatial heterogeneity of factors influencing transportation CO2 emissions in Chinese cities is based on the geographically weighted regression model. *Air Quality, Atmosphere & Health*, *13*(8), pp.977-989.

Yaacob, N.F.F., Mat Yazid, M.R., Abdul Maulud, K.N. and Ahmad Basri, N.E., 2020. A review of the measurement method, analysis, and implementation policy of carbon dioxide emission from transportation. Sustainability, 12(14), p.5873.

# Appendix

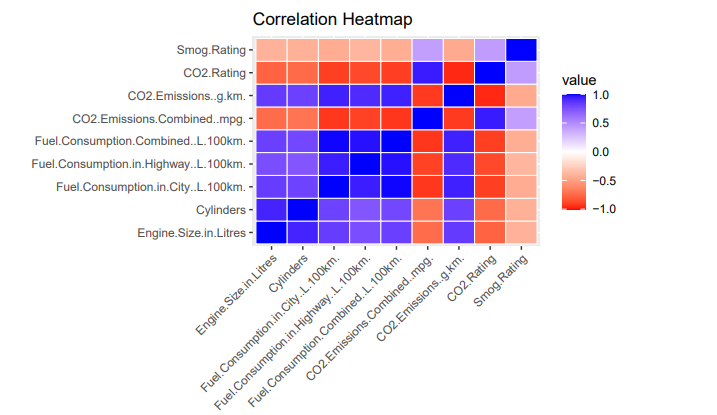


Figure : Correlation Heatmap

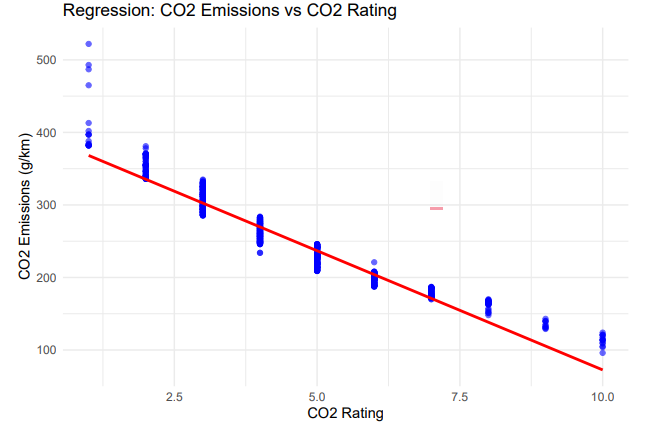


Figure : Linear Regression

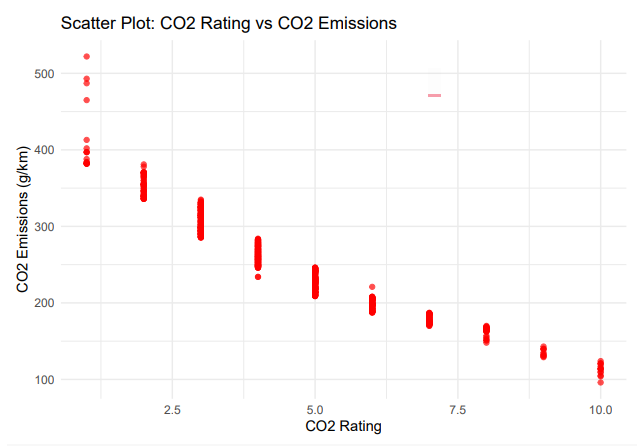


Figure : Scatter plot: CO2 Rating Vs CO2 Emmision

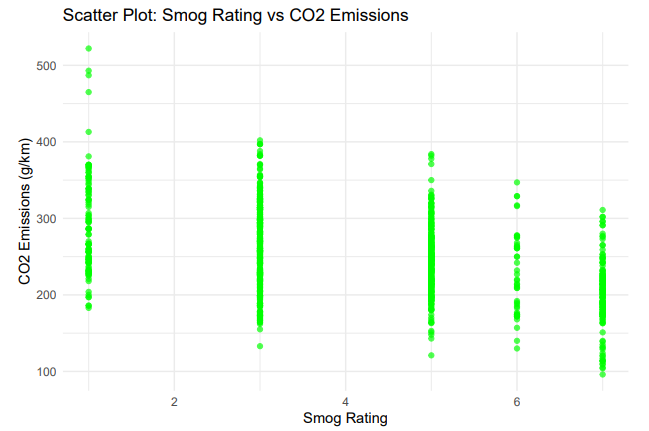


Figure : Scatter plot: smog rating vs CO2 emissions

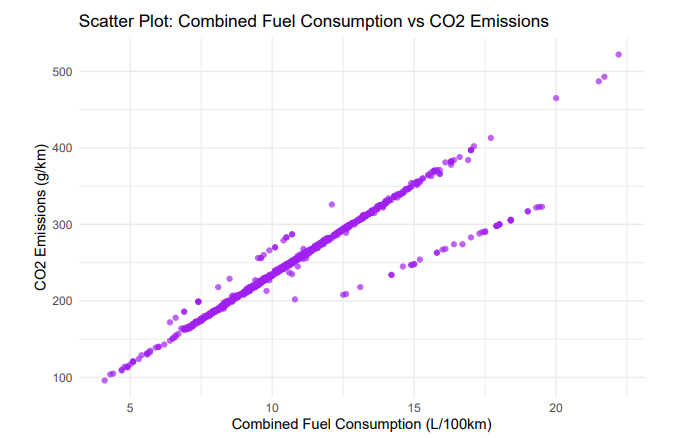


Figure : scatter plot: combined fuel consumption vs CO2 emissions

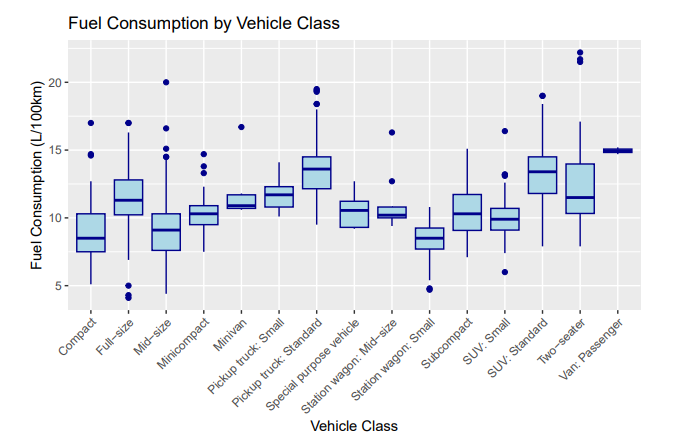


Figure : Fuel consumption by vehicle class

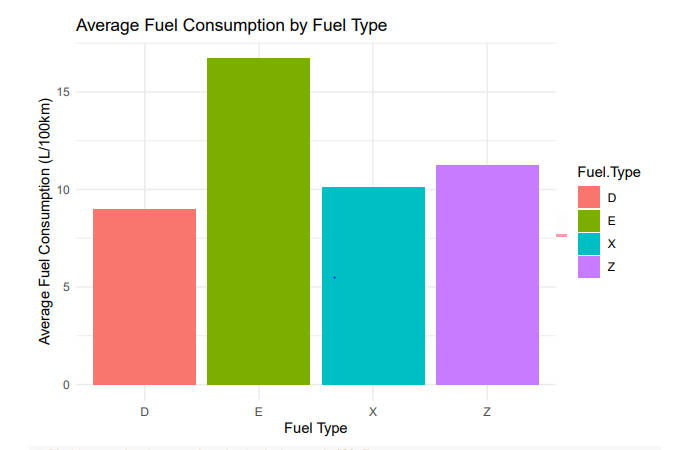


Figure : Average fuel consumption by fuel type

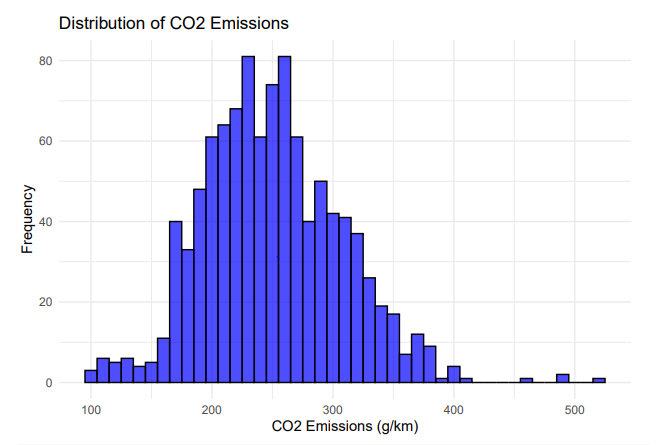


Figure : Histogram ( Distribution of C02 emissions )

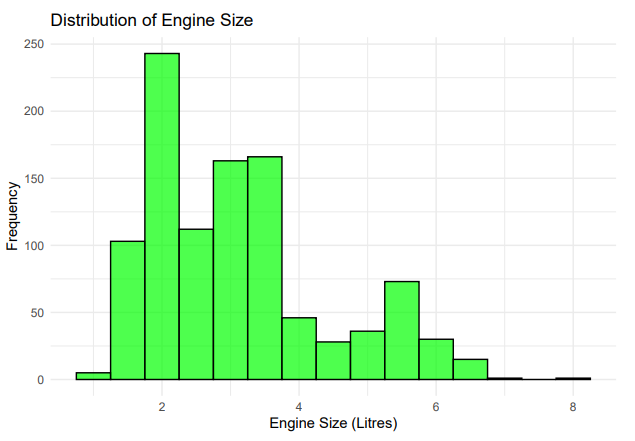


Figure : Histogram (Distogram of engine size )

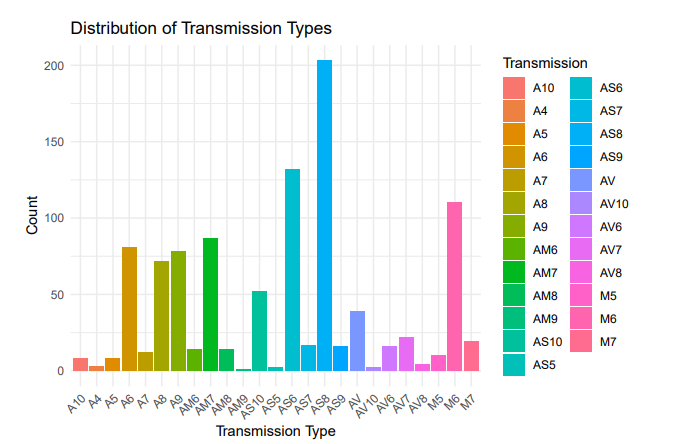


Figure : Histogram ( Distribution of transmission types)

**Code**# Loading necessary libraries

library(ggplot2)

library(dplyr)

library(reshape2)

# Loading the dataset

FuelData <- read.csv("Fuel\_Consumption\_Ratings.csv", stringsAsFactors = FALSE)

# Printing the first few rows of the data

head(FuelData)

# Printing the last few rows of the data

tail(FuelData)

# Printing the structure of the data

str(FuelData)

# Checking the null values in the data

sum(is.na(FuelData))

# Removing the null values from the data

FuelData <- na.omit(FuelData)

# Printing the first few rows of the cleaned data

head(FuelData)

# Printing the summary statistics of the data

summary(FuelData)

# Printing the column names of the data

colnames(FuelData)

# Calculating average consumption

AvgFuelConsumption <- FuelData %>%

group\_by(`Fuel.Type`) %>%

summarise(Average\_Consumption = mean(`Fuel.Consumption.Combined..L.100km.`, na.rm = TRUE))

# Printing the summary of the Average fuel consumption

summary(AvgFuelConsumption)

# Descriptive Statistics for the Fuel Data

# Summary statistics for numerical columns

numerical\_summary <- FuelData %>%

select(CO2.Rating, CO2.Emissions..g.km., Smog.Rating, Fuel.Consumption.Combined..L.100km.) %>%

summarise\_all(list(

min = ~min(., na.rm = TRUE),

max = ~max(., na.rm = TRUE),

mean = ~mean(., na.rm = TRUE),

median = ~median(., na.rm = TRUE),

sd = ~sd(., na.rm = TRUE),

var = ~var(., na.rm = TRUE)

))

print("Descriptive Statistics for Numerical Variables:")

print(round(numerical\_summary, 2))

# Calculate Correlation between CO2 Rating and CO2 Emissions (substitute for Price)

correlation <- cor(FuelData$CO2.Rating, FuelData$CO2.Emissions..g.km., use = "complete.obs")

print(paste("Correlation between CO2 Rating and CO2 Emissions:", round(correlation, 2)))

# Calculate Correlation Matrix for CO2 Rating, CO2 Emissions, Smog Rating, and Combined Fuel Consumption

cor\_matrix <- cor(FuelData %>%

select(CO2.Rating, CO2.Emissions..g.km., Smog.Rating, Fuel.Consumption.Combined..L.100km.), use = "complete.obs")

print("Correlation Matrix for CO2 Rating, CO2 Emissions, Smog Rating, and Combined Fuel Consumption:")

print(round(cor\_matrix, 2)) # Rounded for clarity

# Selecting numeric columns for correlation

FuelDataNumericalColumns <- FuelData %>% select\_if(is.numeric)

# Calculating correlation matrix

FuelDataCorrelationMatrix <- cor(FuelDataNumericalColumns, use = "complete.obs")

# Melting the correlation matrix

FuelDataMeltedCorrelations <- melt(FuelDataCorrelationMatrix)

# Plotting the correlation matrix

ggplot(FuelDataMeltedCorrelations, aes(x = Var1, y = Var2, fill = value)) +

geom\_tile(color = "white") +

scale\_fill\_gradient2(low = "red", high = "blue", mid = "white", midpoint = 0, limit = c(-1, 1)) +

labs(

title = "Correlation Heatmap",

x = "",

y = ""

) +

theme(axis.text.x = element\_text(angle = 45, hjust = 1))

# Build Linear Regression Model for CO2 Emissions Prediction

model <- lm(CO2.Emissions..g.km. ~ CO2.Rating + Smog.Rating + Fuel.Consumption.Combined..L.100km., data = FuelData)

print(summary(model)) # Display model summary

# Plot Regression Line for CO2 Rating vs CO2 Emissions

ggplot(FuelData, aes(x = CO2.Rating, y = CO2.Emissions..g.km.)) +

geom\_point(color = "blue", alpha = 0.6) +

geom\_smooth(method = "lm", color = "red", se = FALSE) +

labs(

title = "Regression: CO2 Emissions vs CO2 Rating",

x = "CO2 Rating",

y = "CO2 Emissions (g/km)"

) +

theme\_minimal()

# Scatter Plot for CO2 Rating vs CO2 Emissions

ggplot(FuelData, aes(x = CO2.Rating, y = CO2.Emissions..g.km.)) +

geom\_point(color = "red", alpha = 0.7) +

labs(

title = "Scatter Plot: CO2 Rating vs CO2 Emissions",

x = "CO2 Rating",

y = "CO2 Emissions (g/km)"

) +

theme\_minimal()

# Scatter Plot for Smog Rating vs CO2 Emissions

ggplot(FuelData, aes(x = Smog.Rating, y = CO2.Emissions..g.km.)) +

geom\_point(color = "green", alpha = 0.7) +

labs(

title = "Scatter Plot: Smog Rating vs CO2 Emissions",

x = "Smog Rating",

y = "CO2 Emissions (g/km)"

) +

theme\_minimal()

# Scatter Plot for Combined Fuel Consumption vs CO2 Emissions

ggplot(FuelData, aes(x = Fuel.Consumption.Combined..L.100km., y = CO2.Emissions..g.km.)) +

geom\_point(color = "purple", alpha = 0.7) +

labs(

title = "Scatter Plot: Combined Fuel Consumption vs CO2 Emissions",

x = "Combined Fuel Consumption (L/100km)",

y = "CO2 Emissions (g/km)"

) +

theme\_minimal()

# Plotting a Boxplot of Fuel Consumption by Vehicle Class

ggplot(FuelData, aes(x = `Vehicle.Class`, y = `Fuel.Consumption.Combined..L.100km.`)) +

geom\_boxplot(fill = "lightblue", color = "darkblue") +

labs(

title = "Fuel Consumption by Vehicle Class",

x = "Vehicle Class",

y = "Fuel Consumption (L/100km)"

) +

theme(axis.text.x = element\_text(angle = 45, hjust = 1))

# Plotting a Bar chart for the distribution of Average Fuel Consumption by Fuel Type

ggplot(AvgFuelConsumption, aes(x = `Fuel.Type`, y = Average\_Consumption, fill = `Fuel.Type`)) +

geom\_bar(stat = "identity") +

labs(

title = "Average Fuel Consumption by Fuel Type",

x = "Fuel Type",

y = "Average Fuel Consumption (L/100km)"

) +

theme\_minimal()

# Plotting a histogram for distribution of CO2 Emissions

ggplot(FuelData, aes(x = `CO2.Emissions..g.km.`)) +

geom\_histogram(binwidth = 10, fill = "blue", color = "black", alpha = 0.7) +

labs(

title = "Distribution of CO2 Emissions",

x = "CO2 Emissions (g/km)",

y = "Frequency"

) +

theme\_minimal()

# Plotting a histogram for distribution of Engine Size

ggplot(FuelData, aes(x = `Engine.Size.in.Litres`)) +

geom\_histogram(binwidth = 0.5, fill = "green", color = "black", alpha = 0.7) +

labs(

title = "Distribution of Engine Size",

x = "Engine Size (Litres)",

y = "Frequency"

) +

theme\_minimal()

# Plotting a histogram for distribution of Transmission Types

ggplot(FuelData, aes(x = `Transmission`, fill = `Transmission`)) +

geom\_bar() +

labs(

title = "Distribution of Transmission Types",

x = "Transmission Type",

y = "Count"

) +

theme\_minimal() +

theme(axis.text.x = element\_text(angle = 45, hjust = 1))

# Plotting a histogram for distribution of Fuel Types

ggplot(FuelData, aes(x = `Fuel.Type`, fill = `Fuel.Type`)) +

geom\_bar() +

labs(

title = "Distribution of Fuel Types",

x = "Fuel Type",

y = "Count"

) +

theme\_minimal() +

theme(axis.text.x = element\_text(angle = 45, hjust = 1))