CO2 Emissions - Analysis and Solutions

Ajay Dhamanda, Jordan Philippe, Frank Olomu, Ravjot Singh, Mayokun Bankole

York University

CSDA1110: Group Project

Instructor : Hemant Sangwan

23 April, 2023

Executive Summary

The purpose of this report is to provide recommendations to our client on reducing their carbon footprint and making data-informed decisions in their business. The following questions are aimed to be explored:

1. Top-performing vehicles in terms of emissions?
2. Variations in top-performing cars by class, engine, cylinders and fuel consumption?
3. Key patterns in data to influence decision making changes in their product lines along with demographic trends.

The analysis was conducted using a Multiple Linear Regression (MLR) model, which included number of cylinders, engine capacity in liters, fuel consumption in city and highways as independent variables while emission was the dependent variable. The dataset was publicly available & the outliers were included for an accurate representation of the data.

Key recommendations include:

1. Launch new vehicles with attributes aligned towards reducing emissions - electric cars, smaller engines, to reduce their carbon footprint and avail government subsidies related to green technology.
2. Improve the performance of existing models by making changes to improve fuel efficiency, weight of the vehicle along with eliminating underselling vehicles to save cost.
3. Focus on research and development for progress in areas such as fuel consumption and/or efficiency, cylinders, engine size and hybrid cars.

Problem Formulation

Purpose and Scope

The main goal of our project is to analyze data related to variables associated with CO_2 emissions in cars and provide insights that can help in important decision-making in the field of marketing for non-electric cars. These are some of the questions we aim to answer:

1. What are the top-performing companies in terms of CO_2 emissions by region?
2. How do the top-performing cars vary along key dimensions such as vehicle classes, engine sizes, number of cylinders, type of fuel, and fuel consumption?
3. Can we group or segment products based on meaningful dimensions such as high or low-CO_2 emitting cars, engine sizes, and fuel types?
4. What key patterns in the data can help us understand and influence decision-making in areas such as launching a new car, eliminating underperforming cars and, understanding demographic trends that suggest changes in specific features?

External Environment

An overview of the external environment:

1. **Government regulations and social responsibility**: Governments around the world are imposing increasingly stringent regulations on CO_2 emissions in order to combat climate change. We must also consider the World Economic Forum's (WEF) initiatives to tackle climate change and their impact on the industry. In this context, electric vehicles, in comparison to conventional vehicles, are a viable alternative.

2. **Technological advancements:** As the industry shifts towards electric and hybrid vehicles, advances in battery technology and charging infrastructure will be important drivers of this transition, eventually reducing overall CO2 emissions from vehicles.
    3. **Economic factors:** The automotive industry is susceptible to economic fluctuations and consumer trends. The impending global recession, the COVID-19 pandemic, and shortages in the manufacturing of high-end chips could have a profound impact on the industry.
    4. **Competition and innovation:** The automotive industry is highly competitive, and innovation is necessary to remain competitive. Innovations relating to features, efficiency, and ease of use for customers will be important drivers of demand.
    5. **Demographic and cultural trends:** Changes in demographics and cultural trends can affect the types of vehicles consumers prefer. For instance, the growing popularity of SUVs and crossover vehicles has implications for the CO2 emissions of the automotive industry. Additionally, some small towns might only purchase cars that are more efficient on highways, while customers in urban areas would prefer cars that are more efficient in city-type situations and areas.

Data Requirements

Vehicle Data- Data on different types of vehicles such as SUVs or sedans, along with attributes such as make, model, year of manufacture, fuel type, engine size, transmission, and emission standards.

Geographic Data: Data on geographic location of vehicles, including information on urban, rural, or highway driving patterns, road conditions, and topography, which can affect emissions.

Fuel Data- Data on different types of fuels used in vehicles, such as gasoline, diesel, and biofuels and emission characteristics.

Consumer Data*:* Data on consumer behavior and preferences such as vehicle purchasing patterns, cars sold per vehicle class and other trends such as increase or decrease in certain types of car sales.

Based on these parameters, we can collect and analyze the data to gain meaningful insights into the factors influencing CO2 emissions in vehicles and develop strategies to reduce emissions.

Analytical Plan

The data (Natural Resources Canada, 2023) was acquired from the Canadian Government's website and was part of publicly available data. In the dataset, there were not many missing values so we did not require any cleaning at this stage. Upon inspecting the variables, we found a number of variables that, in theory, are directly correlated with CO2 emissions, such as number of cylinders and fuel consumption per 100km. Therefore, there was no need to manipulate and/or combine data to create new variables. Using a scatter plot (app) we were able to identify certain outliers in nearly all of the integer-type columns. These outliers were included in our analysis as it was more helpful to consider these, for example - extremely high fuel consumption cars, rather than exclude them as excluding them would fail to represent the average of all cars. The medians and modes were not affected.

Basic Analysis: The measures of central tendencies can be found in Appendix 1. The following statement incorporates all of the measures of central tendencies -

"On average, each car has an engine capacity of 3.4L while having 6 cylinders. In addition, each car, on average, consumes about 13L of fuel per 100km in the city & consumes 8.8L of fuel per 100km on highways, while emitting 248 grams of CO_2 per km traveled."

Modeling and Analysis: At this stage, a Multiple Linear Regression model was used. For our purposes we used four independent variables being - number of cylinders, engine capacity in liters along with fuel consumption in city and highways. Two variables related to fuel consumption were used as it was hypothesized that those two variables might be correlated with each other but might have a unique relationship with emissions i.e - it is possible that while some vehicles consume less fuel on highways, the emissions would not necessarily be lesser.

Conclusion

In conclusion, the recommendations for the client are that they focus on launching new product lines with specific attributes that are aligned towards the aim of reducing CO_2 emissions, as this will not only help them market as environmentally friendly and sustainable but also avail the government subsidies as a consequence. These attributes should include electric and smaller engine vehicles. Additionally, the manufacturer should work towards improving the performance of its existing vehicles by making changes in addition to eliminating underperforming vehicles. The variables that impact CO_2 emissions the most such as fuel consumption, number of cylinders and engine size should be considered when focusing on research and development. Other areas of focus would be reducing the weight of the vehicle, improving fuel efficiency, improving air dynamics or even using hybrid technology that leverages the power of electric cars while being able to operate on either type of fuel. By leveraging these insights gained from the MLR analysis, the manufacturer can make data-driven decisions that could lead to fulfillment of their business objectives such as increasing sales for specific cars, improving their brand image or even increasing market share as a result of better placed products in the market.

References

Natural Resources Canada. (2023, April 4). *Fuel Consumption Ratings*. Open Government Portal. Retrieved April 23, 2023, from https://open.canada.ca/data/en/dataset/98f1a129-f628-4ce4-b24d-6f16bf24dd64

Appendices

Appendix 1

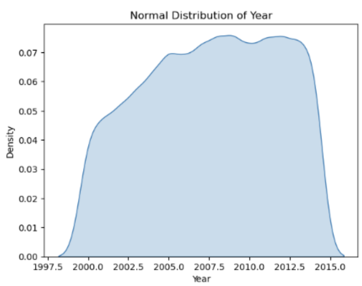
Variable	Count (n)	Mean	Std	25%	50%	75%	Max
Engine (L)	14253	3.47785	1.311102	2.4	3.4	4.4	8.4
Cylinders	14253	5.99123	1.774153	4	6	8	16
FC - City (L/100km)	14253	12.934456	3.541439	10.6	12.4	14.9	30.6
FC - Highway (L/100km)	14253	8.831348	2.338169	7.2	8.3	10.2	19
Combined (L/100km)	14253	11.089637	2.969484	9.1	10.5	12.8	24.8
Emissions	14253	248.790149	59.07719	209	239	285	570

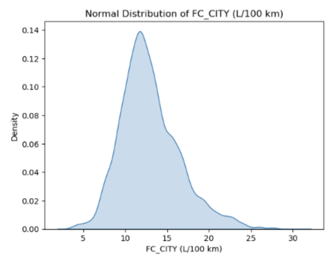
Appendix 2 - Before and After - Data Manipulation

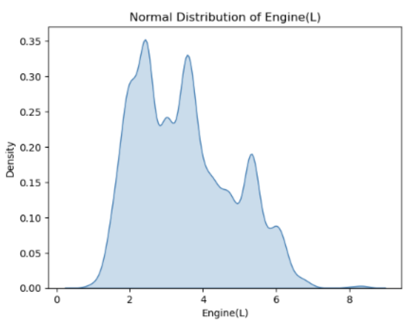
	A	B	C	D	E	F	G	H	I	J	K	L	M
1	MODEL	MAKE	MODEL	VEHICLE CLASS	ENGINE SIZE	CYLINDERS	TRANSMISSION	FUEL	FUEL CONSUMPTION				CO2 EMISSIONS
2	YEAR		# = high output engine		(L)			TYPE	CITY (L/100 km)	HWY (L/100 km)	COMB (L/100 km)	COMB (mpg)	(g/km)
3	2000	ACURA	1.6EL	COMPACT	1.6	4	A4	X	9.2	6.7	8.1	35	186
4	2000	ACURA	1.6EL	COMPACT	1.6	4	M5	X	8.5	6.5	7.6	37	175
5	2000	ACURA	3.2TL	MID-SIZE	3.2	6	AS5	Z	12.2	7.4	10	28	230
6	2000	ACURA	3.5RL	MID-SIZE	3.5	6	A4	Z	13.4	9.2	11.5	25	264
7	2000	ACURA	INTEGRA	SUBCOMPACT	1.8	4	A4	X	10	7	8.6	33	198
8	2000	ACURA	INTEGRA	SUBCOMPACT	1.8	4	M5	X	9.3	6.8	8.2	34	189
9	2000	ACURA	INTEGRA GSR/TYPE R	SUBCOMPACT	1.8	4	M5	Z	9.4	7	8.3	34	191
10	2000	ACURA	NSX	SUBCOMPACT	3	6	AS4	Z	13.6	9.2	11.6	24	267
11	2000	ACURA	NSX	SUBCOMPACT	3.2	6	M6	Z	13.8	9.1	11.7	24	269
12	2000	AUDI	A4	COMPACT	1.8	4	A5	Z	11.4	7.2	9.5	30	218
13	2000	AUDI	A4	COMPACT	1.8	4	M5	Z	9.7	6.8	8.4	34	193
14	2000	AUDI	A4	COMPACT	2.8	6	A5	Z	13	8.2	10.8	26	248

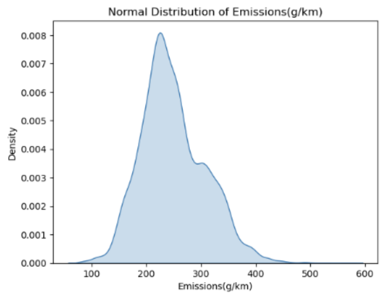
	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Year	Make	Model(#)	Vehicle Class	Engine (L)	Cylinders	Transmission	FuelType	FC_CITY (L/100 km)	FC_HWY (L/100 km)	Combined (L/100 km)	COMB (mpg)	Emissions(g/km)
2	2000	ACURA	1.6EL	COMPACT	1.6	4	A4	X	9.2	6.7	8.1	35	186
3	2000	ACURA	1.6EL	COMPACT	1.6	4	M5	X	8.5	6.5	7.6	37	175
4	2000	ACURA	3.2TL	MID-SIZE	3.2	6	AS5	Z	12.2	7.4	10	28	230
5	2000	ACURA	3.5RL	MID-SIZE	3.5	6	A4	Z	13.4	9.2	11.5	25	264
6	2000	ACURA	INTEGRA	SUBCOMPACT	1.8	4	A4	X	10	7	8.6	33	198
7	2000	ACURA	INTEGRA	SUBCOMPACT	1.8	4	M5	X	9.3	6.8	8.2	34	189
8	2000	ACURA	INTEGRA GSR/TYPE R	SUBCOMPACT	1.8	4	M5	Z	9.4	7	8.3	34	191
9	2000	ACURA	NSX	SUBCOMPACT	3	6	AS4	Z	13.6	9.2	11.6	24	267
10	2000	ACURA	NSX	SUBCOMPACT	3.2	6	M6	Z	13.8	9.1	11.7	24	269
11	2000	AUDI	A4	COMPACT	1.8	4	A5	Z	11.4	7.2	9.5	30	218
12	2000	AUDI	A4	COMPACT	1.8	4	M5	Z	9.7	6.8	8.4	34	193
13	2000	AUDI	A4	COMPACT	2.8	6	A5	Z	13	8.2	10.8	26	248
14	2000	AUDI	A4	COMPACT	2.8	6	M5	Z	11.7	7.5	9.8	29	225
15	2000	AUDI	A4 QUATTRO	COMPACT	1.8	4	A5	Z	12.1	7.7	10.1	28	232
16	2000	AUDI	A4 QUATTRO	COMPACT	1.8	4	M5	Z	10.7	7.5	9.3	30	214
17	2000	AUDI	A4 QUATTRO	COMPACT	2.8	6	A5	Z	13.3	8.5	11.1	25	255
18	2000	AUDI	A4 QUATTRO	COMPACT	2.8	6	M5	Z	12.7	8.7	10.9	26	251

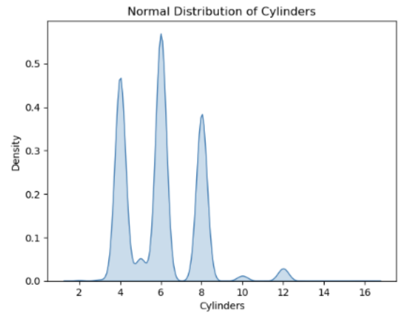
Appendix 3 - Distributions

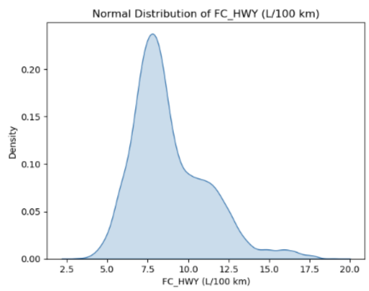




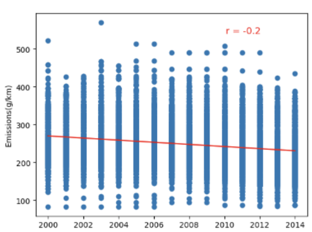


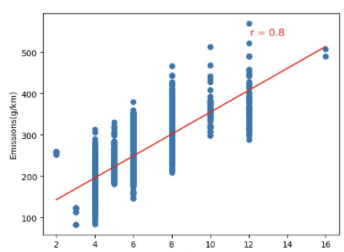


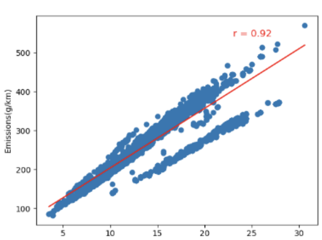


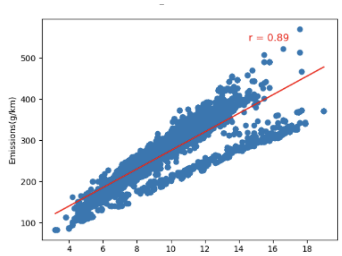


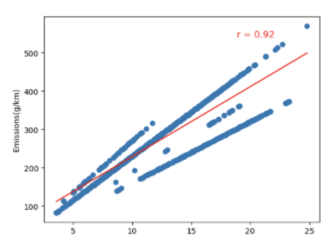
Appendix 4 - Correlations



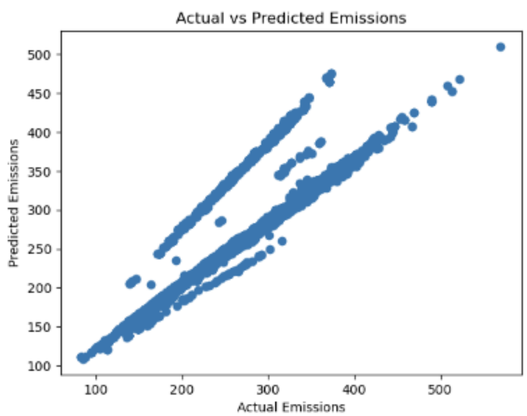








Appendix 5 - Model



Model summary:
R-squared: 0.84
Mean Squared Error: 546.45

Appendix 6 - Top 5 and Bottm 5

Top 5 makes:

	Make	FC_CITY (L/100 km)	FC_HWY (L/100 km)	Combined (L/100 km)
6	BUGATTI	26.566667	15.500000	21.566667
13	FERRARI	21.683562	13.202740	17.869863
4	BENTLEY	20.708065	12.553226	17.043548
25	LAMBORGHINI	20.572549	12.500000	16.929412
29	MASERATI	18.106522	11.158696	15.169565

Bottom 5 makes:

	Make	FC_CITY (L/100 km)	FC_HWY (L/100 km)	Combined (L/100 km)
44	SMART	5.635294	4.588235	5.152941
14	FIAT	7.210000	5.590000	6.495000
32	MINI	8.001531	5.883163	7.055102
43	SCION	8.393548	6.270968	7.425806
1	ALFA ROMEO	9.100000	6.100000	7.700000

Appendix 7 - MLR Model

Coefficients:
FC_CITY (L/100 km): 7.262109307990461
FC_HWY (L/100 km): 6.938470943732328
Cylinders: 4.989094071545951
Engine(L): 5.623708653494453
Intercept: 44.13344046089662
R-squared: 0.87
Mean Squared Error: 468.32