SPDS - Group 2 - Project Proposal

Investigating the Relationship Between Vehicle Type and CO2 Emissions

Team Name: Data Wizards

Team Members:
- Alexander Rabmer - K11906487
- Fathy Shalaby - K11701175
- Manuel Vujakovic - K11903999
- Philipp Bauer - K11941210
- Burak Demirdegmez - K12137646

Dataset:

This dataset (taken from Kaggle[1]) captures the details of how CO_2 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 a total of 7385 rows and 12 columns. There are few abbreviations that have been used to describe the features.

Data:

The data for this project will be sourced from the United States Environmental Protection Agency (EPA) and will include information on the CO_2 emissions of various vehicle types. The dataset will include variables such as vehicle type, engine size, fuel type, and CO_2 emissions.

"Make"	"Model"	"Vehicle.Class"	"Engine.Size.L."	"Cylinders"	"Transmission"	"Fuel.Type"	"Fuel.Consumption..."	"Fuel..."	"Fuel..."	"Fuel..."	"CO2.Emissions.g.km."
"ACURA"	"ILX"	"COMPACT"	2.0	4	"AS5"	"Z"	9.9	6.7	8.5	33	196.0
"ACURA"	"ILX"	"COMPACT"	2.4	4	"M6"	"Z"	11.2	7.7	9.6	29	221.0
"ACURA"	"ILX HYBRID"	"COMPACT"	1.5	4	"AV7"	"Z"	6.0	5.8	5.9	48	136.0
"ACURA"	"MDX 4WD"	"SUV - SMALL"	3.5	6	"AS6"	"Z"	12.7	9.1	11.1	25	255.0
"ACURA"	"RDX AWD"	"SUV - SMALL"	3.5	6	"AS6"	"Z"	12.1	8.7	10.6	27	244.0
"ACURA"	"RLX"	"MID-SIZE"	3.5	6	"AS6"	"Z"	11.9	7.7	10.0	28	230.0

The variables themselves are self explanatory, some of the abbreviations for the values of these variables are listed here:

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

[1] https://www.kaggle.com/datasets/debajyotipodder/co2-emission-by-vehicles

Problem:
The goal of this project is to investigate the relationship between vehicle type and CO2 emissions. We will analyze data mentioned above to determine whether certain vehicle types tend to have higher or lower CO2 emissions. We will also investigate potential factors that may contribute to these differences.

There are several motivations for conducting this project. Firstly, reducing vehicle emissions is a critical step in addressing climate change. As the transportation sector is one of the largest sources of greenhouse gas emissions, it is important to understand which vehicle types are the most polluting in order to target efforts to reduce emissions. Additionally, reducing vehicle emissions can also improve air quality, which has numerous health benefits.

Another motivation for this study is to provide valuable information to consumers who are interested in purchasing low-emitting vehicles. By identifying which vehicle types tend to have lower CO2 emissions, we can help consumers make more informed decisions and potentially encourage them to choose vehicles that are better for the environment.

Furthermore, this study has the potential to inform policy and regulatory decisions related to vehicle emissions. By providing a better understanding of the relationship between vehicle type and CO2 emissions, we can help policymakers develop effective strategies for reducing emissions from the transportation sector.

Overall, our goal is to contribute to a better understanding of the relationship between vehicle type and CO2 emissions, and to provide valuable information that can be used to reduce emissions and address climate change.

Methodology:
We will use regression analysis to explore the relationship between vehicle type and CO2 emissions. This will involve fitting a regression model to the data and analyzing the model coefficients to identify which vehicle characteristics are most strongly associated with CO2 emissions. We will also try and see if we can use classification in a meaningful way.

Timeline and Plan:
Our team will meet once or twice per week to discuss our progress and plan our next steps. At first we will be cleaning and preparing the data. Afterwards, we will work together to carry out the statistical analysis with the methods described above. Last but not least, our group will interpret the results together and write the final report. We plan to complete this project within the next 6 weeks with a short break during the "core" holiday season.