Vehicle CO2 Emissions – EDA

*By William Gray Renton, Isbelis Castro, Gunel Garayeva, Eugenia Chien –Group 06. ​*

Alexander Booth (Instructor)*​ and* Reed Schelitzche. – (TA)*​*

## A diagram of a car Description automatically generatedIntroduction

*Note: Co2 emission characterist…..*

## Approach

  For the data analysis were used secondary data from Kaggle.

* Vehicle Emissions Data Set -used secondary data. ​
* The base date has information about 6,756 records.​
* Programming language: Python ​ with libraries as panda, numpy, matplotlib, plotly, seaborn, Scipy, sklearn.

## Data cleaning

The dataset consisted of 6,756 cars. During cleaning we noticed that there was a subset that had null for emissions data. Upon further investigation, we found it that those were the Electric cars that did not produce any CO2 emissions. So for the purposes of this project, we used a "drop" function to remove all "Electric" cars from the "transmission\_type" column. That removed a little over 200 records and the final dataset consisted of 6,542 records.​

A close-up of a button

Description automatically generated*Note: the figure shows the data cleaning process by owner group.*

## Dataset

### Fields list

The dataset has 11 fields, the following table:

**A screenshot of a computer

Description automatically generated**

1. A **technical writeup** (40pts) expanding on your presentation. Code, screenshots, appendix, any extra details - go nuts. The writeup will be organized the same as your presentation - think of it as a formalized script from your presentation. The goal here is not to start from scratch, but to simply expand your skeleton and slide deck into a written report. Feel free to expand as much as you want.
2. This is what most people will read when reviewing this project in your portfolio btw
3. Complete sentences, grammar, etc. This is a professional writeup.
4. Add in as much extra detail that you couldn't fit into a 15 min presentation. You can share as many visuals as you would like here.