

Summary of what I've learned and had to do to complete the project for DSC540 Data Preparation class

What I learned in this Course and what was the ultimate goal of the project?

This course is one of the steppingstone of data science journey and I got to learn and brush up lots of new skills and techniques as part of this course. One of the key things I got to learn from this course is web scraping and API requests to load data. Data Preparation is fundamental step in data analysis and this course helped me learn how data need to be transformed, cleaned, and prepared based on data types and how to visualize the data for any outliers or bad data or missing values and how to handle them.

During this project I wanted to analysis electric car battery disposal pollution and the lifetime emissions of gasoline-powered cars vs. electric cars. Unfortunately, I couldn't get any data for the electric car batteries disposal pollution, However I focused on Fuel, City MPG, Hwy MPG, and Cmb MPG for Gasoline, Gasoline/Electricity, Electricity, and Hydrogen vehicles from collected all datasets to find out the Greenhouse Gas Score and Air Pollution Score for mentioned cares. Below is summary analysis of the project.

During this 5-milestone analysis I ended up with air pollution score and greenhouse gas score based on vehicles fuel. Any issues such as smog, haze and other problems caused by carbon emissions are all part of air pollution. Vehicles are scored on a scale of 1 to 10, which 1 being dirtiest and 10 being the cleanest. When cars burn gasoline, they emit pollutions. Gasoline fumes escape into the air even when we pump gasoline into our fuel tanks. In the last visualization portion of the project (Visualization 5: Lineplots using Pandas DFs) we clearly see the differences air pollution scores and greenhouse gas score between gasoline, gasoline/electricity, and electricity vehicles. Below is the graph for Air Pollution Score & Greenhouse Gas Score.

Air Pollution Score & Greenhouse Gas Score

