DATA VISUALIZATION PROJECT PROPOSAL

Project Title: Advancing Sustainable Mobility: A Data Visualization Analysis of U.S. EPA Automotive Trends (1975 - preliminary 2024) for Fuel Economy, CO_2 Emissions Reduction, and Cleaner Technologies in Alignment with SDGs 7 and 13

Group name : Tres Sigmas - 4CSE (DataViz-E 1)

- Inventado, Charles Fredric G.
- Rodelas, John Vincent B.
- Valles, James Vincent V.

Date : September 29, 2025

Project Overview :

The Tres Sigmas group from 4CSE will explore the detailed **2024 EPA Automotive Trends Report** taken from the **United States Environmental Protection Agency (EPA)**, with the goal to be attributed to improving the two of the **United Nations Sustainable Development Goals (SDG)**. The group will brainstorm forming SDG-aligned questions and address these by utilizing data visualization techniques to explore the historical trends, patterns, and progression on fuel economy, CO_2 emissions reduction, and adopting cleaner vehicle technologies over time.

Project Background

The study tackles global issues prevalent in the transport sector as one of the largest contributors of greenhouse gas emissions affecting the air quality from urban areas and catalyzes the impacts of climate change. Considering that mobility today is on demand and heavily reliant on traditional fossil fuels, we will face long-term consequences regarding sustainability as resources may go scarce and be limited in the future.

The project aims to contribute on **SDG 7 (Affordable and Clean Energy)** by examining the **fuel economy** to have a better insight into how vehicles advance toward clean and reliable modern energy, and **SDG 13 (Climate Action)** by analyzing the **Carbon Dioxide (CO_2) emissions** data and the adoption of other **fuel-efficient** and **cleaner vehicle technologies** to promote sustainable consumption and to take urgent action on CO_2 emission reduction to and take combat climate change.

Statement of the Problem

The sector of transportation is a primary cause of global carbon dioxide (CO_2) emissions, contributing for a substantial portion of consumption of fossil fuels, climate change, air pollution, and insecurity of energy. In the U.S., many fleets of vehicles still continue to rely on internal combustion engines despite the regulatory advancements since 1975. These results in persistent challenges to achieving much higher fuel economy standards and widespread adoption of cleaner technologies and also slows down the progress toward United Nations Sustainable Development Goals, specifically the SDGs 7 (Affordable and Clean Energy) and 13 (Climate Action), and other commitments involving international climate agreements such as the

Paris Accord. The transition to a sustainable and low-emission mode of transport remains suboptimal without having any deeper insights into historical trends, patterns, and barriers, increasing environmental degradation and delaying global policy objectives for emission reductions

Objectives

The project will utilize data visualization techniques to discover important trends and occurrences and discover long-term patterns in vehicle fuel and economy efficiency and emissions. Addressing the key challenges in the transport sector will help us uncover relevant insights and strategies that could be possibly done aligned with global sustainability efforts, thus, we could successfully contribute to the target SDGs by committing to these objectives:

- Examine relevant variables and be able to visualize historical trends in fuel economy, CO_2 emissions, and the adoption of fuel-efficient and cleaner vehicle technologies using the data available spanning from 1975 to preliminary 2024 data.
- IIdentify key patterns, progressions, and significant factors in the U.S. vehicle sector aligned with SDG 7 to sustain clean energy, and SDG 13 to prevent the severity of impacts of climate change.
- Brainstorm to form and be able to address SDG-aligned questions by having data-driven analysis, exploring the economic, environmental, and technological implications of emission reductions.
- Provide recommendations that can be improved by policymakers, industry stakeholders, and researchers to promote sustainable consumption, accelerate technology adoption, and support global development goals and commitments in the transportation sector.

For the scope of the study, first, the vehicles fleets and regulations were based only from the country the dataset originated from, the United States (U.S.), thus excluding any other countries than the stated one. Second, the data covers only from 1975 until the preliminary of 2024, which may not include the latest EV technologies and innovations made recently of late 2024 and the rest of 2025. Lastly, regards to contributing pollution features are focused only on the CO_2 emissions and not other contributing factors such as Nitrogen Oxide and Sulfur Oxide.

Regarding the limitation of the study, first, the dataset is containing missing values on columns requiring imputation of data on some records such as when a vehicle will not have complete value on all of powertrain columns and the number of gears, there are also missing years on manufacturers such as Tesla's data were available since only 2011. Second, percentages or fraction values are to describe technological adoption such as hybrid systems and transmission type present on the dataset, which needs the team to carefully interpret to use the appropriate visualization or model. Third limitation is that vehicle type values were aggregated averages, that not a specific car model is represented thus limiting the possibility of a more detailed analysis.

Background on the Dataset

Name : A. Detailed Real-World Fuel Economy, CO2 Emissions, and Vehicle Attribute and Technology Data
Source : EPA United States Environmental Protection Agency
https://www.epa.gov/automotive-trends/explore-automotive-trends-data#SummaryData

Regarding coverage, notice from the Model Year column, the time span of the records are usually from 1975 to 2024. The records were also further organized into the 15 (including "all" on the count) manufacturers of the automobile, the corresponding regulatory class (All, Car, and Truck), and the vehicle classification. (All, All Car, All Truck, Car SUV, Minivan/Van, Pickup, Sedan/Wagon, Truck SUV)

Aside from the stated four column identifiers stated above, from the original dataset gathered, there are over 56 indicators that could be themed by production (ex: Production (000), Production Share) , fuel efficiency performance (ex : 2-Cycle MPG, Real-World MPG, etc…), Carbon Emissions (ex : Real-World CO2 (g/mi), etc…), Vehicle specifications (ex : Weight (lbs), Footprint (sq. ft.), Engine Displacement, Horsepower (HP), etc…), Transmission / Drivetrain Features (ex : Cylinders in Gasoline ICE Vehicles, Drivetrain - Front, Transmission - Manual, 4 or Fewer Gears, etc…) , and Fuel Delivery Type and Powertrain Types (ex : Fuel Delivery - Diesel, Powertrain - Diesel, etc…)

The data can be accessed through the hyperlink given on the source through **CSV download**. There were no secondary datasets used, the primary dataset focused thoroughly on the vehicle's characteristics, vehicle performance, fuel efficiency , and carbon emissions.

Literature review

The **Energy Policy, 146, 111783 (Greene et al., 2020)** study offers a historical evaluation of the U.S. Corporate Average Fuel Economy (CAFE) and Greenhouse Gas (GHG) standards since 1975, utilizing EPA data to quantify impacts, attributing a significant portion of fuel economy gains to regulations. Its longitudinal trend analysis directly inspires the project's use of the EPA data to visualize historical patterns in fuel economy and CO2 emissions, aligning with Sustainable Development Goal (SDG) 7 (clean energy transition) and SDG 13 (emission avoidance). The study's limitations include a lack of interactivity, multi-faceted overlays (e.g., combining CO2 with technology adoption timelines), and exclusion of explicit SDG framing or global policy comparisons. The project will address these by incorporating SDG-aligned questions and advanced, dynamic visualizations up to preliminary 2024 data.

The **Proceedings of the National Academy of Sciences, 120(14). (DiStefano and Zeitler, 2023)** case study assesses the influence of National Academies reports (2002–2021) on EPA/NHTSA regulations, using EPA Automotive Trends data for fleet-average fuel economy timelines and showing implementation rates via static bar charts and tables. The study's recommendation-tracking framework inspires the project's development of SDG-aligned policy questions, using EPA data visualizations to evaluate progress toward cleaner technologies (SDG 7) and urgent climate action (SDG 13). The primary gaps are static visualizations, no integration of recent 2024 data, and the absence of SDG metrics. The project will mitigate these by extending the analysis to preliminary 2024 figures and creating dynamic, SDG-explicit visualizations.

The **Harvard Law School** (**Greenstone et al., 2020)** paper critiques the inefficiencies of the 2012 National Program, proposing a cap-and-trade system and using extensive EPA compliance data to depict line graphs of projected versus actual Miles Per Gallon (MPG)/CO2, scatter plots of lifetime consumption, and market share trends. The study's detailed EPA-derived visualizations of projections and loopholes inspire data exploration techniques for plotting CO2 reductions against technology adoption, supporting SDG 13 targets. The limitations are descriptive and non-interactive figures, a focus on policy critique without SDG integration, and a lack of global context for trading visualizations. The project will utilize the full 1975-preliminary 2024 dataset for interactive visualizations and incorporate SDG-specific brainstorming.

The **Intergovernmental Panel on Climate Change (IPCC, 2022)** reviewed links to SDG 7 and SDG 13, using EPA-aligned U.S. data within broader analyses for visualizations like bar charts comparing modes and scenario projections. The chapter's use of SDG-metric tables and emission comparisons inspires the project's objective to use EPA data to quantify U.S. progress on SDG 7 and SDG 13 urgency within a global context. The gaps include global aggregation of EPA data with a lack of granular historical trends or interactive tools, and insufficient detail on regulatory specifics. The project will address this by focusing on the EPA's 50-year dataset to deliver detailed, U.S.-centric visualizations with SDG-explicit overlays.

5. Methodology (Plan)

a. **World Development Indicators (WDI) to Use:**

- Average CO2 Emissions per Mile (g/mi) : *[Real-World CO2 (g/mi)]*
- Average Vehicle Fuel Economy (MPG) : *[Real-World MPG]*
- Average Vehicle Weight (lbs) : *[Weight (lbs)]*
- Average Engine Horsepower (HP) : *[Horsepower (HP)]*
- Annual Change in Fleet Emissions (%) : *[Real-World CO2 (g/mi) across diff. Model Year]*
- Urban vs. Highway Emissions Disparity : *[Real-World CO2_City (g/mi) and Real-World CO2_Hwy (g/mi)]*

Other important columns for visualization:

- Manufacturer, Model Year, Regulatory Class and Vehicle Type, Production (000), Powertrain Types, Fuel Delivery Types, Footprint Class

b. **Data Preparation:**

1.) **Fix data formatting -** On Model Year, Replace the term “Prelim. 2024” to “2024” on column “Model Year”
2.) **Replace missing or null values -** Check for missing or null values, replace the values of these cells with a simple imputation technique by getting the respective column’s mean value on numerical data while mode for the possible categorical data.
3.) **Check for Duplicates -** using tableau prep builder as well, check and remove records that are detected as duplicates.
4.) **Removal of columns not significant to be used on the study -** filter to only include relevant columns on the study, listed above on indicators we are planning to use.
5.) **Monitoring outliers -** using Box and Whisker Plot, check the outliers on columns; although we will not pursue replacing nor removing the value on these upper (90% quartile) or lower (10% quartile) extremities.
6.) **Scale the values for standardization -** normalize the values of columns using StandardScaler.
7.) **Proper column labels -** rename the columns appropriately and ease to understand, we can rename columns especially with technical terms or difficult to understand abbreviations.

c. **Exploratory Data Analysis (EDA) :**

1.) **Descriptive Statistics** - computing for the columns statistical values such as mean, median, mode, mind, max, and standard deviation for numerical data and make frequency counts for categorical data.
2.) **Visualize outliers of Columns** - using box and whisker plots (under data preparation too), we can visualize onto what extent values at lower and upper extremities affect the affected statistical values we have.
3.) **Correlation of Columns** - use scatter plot to show the relationship between two numerical columns and determine whether the pair has a strong correlation or not.
4.) **Distribution Analysis** - on numerical data, use histograms to check for distribution such as skewness, and bar charts for categorical columns to visualize class frequency.

d. **Planned Visualizations:**

1.) **Bump Chart:** Shows fuel economy rank of vehicle types over time. Reveals efficiency leaders and laggards. Assesses SDG 7 progress in market segments.
2.) **Treemap Chart:** Displays manufacturer and class contribution to the fleet (size) and their CO2 intensity (color). Aggregates high-volume producers and their CO2 cleanliness.
3.) **Stacked Bar Chart:** Tracks the rate and mix of cleaner technology adoption over time. Visualizes technological transition. Evaluates SDG 7 and SDG 13 commitment.
4.) **Dual Axis Line Chart:** tracks CO2 reduction (SDG 13) against vehicle mass/size trends. Identifies technology masking effects, where heavier vehicles necessitate aggressive CO2 cuts.

References

DiStefano, J. G., & Zeitler, E. L. (2023). Implementation of National Academies' study recommendations: A fuel economy case study. Proceedings of the National Academy of Sciences, 120(14). https://doi.org/10.1073/pnas.2303242120

Greene, D. L., Greenwald, J. M., & Ciez, R. E. (2020). U.S. fuel economy and greenhouse gas standards: What have they achieved and what have we learned? Energy Policy, 146, 111783. https://doi.org/10.1016/j.enpol.2020.111783

Greenstone, M., Sunstein, C. R., & Ori, S. (2020). Fuel economy 2.0 (Harvard Public Law Working Paper No. 17-27). Harvard Law School. https://journals.law.harvard.edu/elr/wp-content/uploads/sites/79/2020/04/44.1-Greenstone-Sunstein-Ori.pdf

IPCC. (2022). Chapter 10: Transport. Www.ipcc.ch. https://www.ipcc.ch/report/ar6/wg3/chapter/chapter-10/