Project 0: Deadline: October 13, 2024, 11:59pm Eastern Time.

Group Members:
Lilly Parham - lparham2@student.gsu.edu
Dorothy (Gracie) Rehberg - drehberg1@student.gsu.edu
Pamela Alvarado-Zarate - palvaradozarate1@student.gsu.edu

The International Energy Association has published the following findings and estimates:

- "Electric vehicles are the key technology to decarbonize road transport, a sector that accounts for over 15% of global energy-related emissions."
- "18% of new cars sold in 2023 will be electric"
- The use of EVs displaced around 0.9 Mb/d (2 EJ) of oil in 2023.
- EVs would need to displace around 8.2 Mb/d (18 EJ) of oil in 2030 to be in step with the Net-Zero Emissions Scenario.

In furtherance of pursuing a Net-Zero Emissions Scenario, many countries have set targets for electric vehicles to make up a larger percentage of all newly manufactured vehicles in the coming decades.

In the United States in 2023, electric vehicles made up 6.9% of new vehicle purchases. The US Environmental Protection Agency set a target range of 35% to 56% of all new vehicles sold between 2030 and 2032 to be all electric.

The rate of electric vehicles manufactured is driven partly by consumer demand and partly by manufacturer choices. Mainly, the EPA tries to reach its target by publishing stricter regulations for manufacturers of combustion engine vehicles to push them towards manufacturing electric vehicles. But, the EPA is exploring ways to reach its electric vehicle targets by influencing consumer demand.

In pursuit of achieving its 2030-2032 target, the EPA has approached **your data analytics team** to search for **opportunities to boost consumer demand for electric vehicles** in the United States. Your team suspects that many regions of the US have consumers who would choose to buy an electric vehicle if the density of charging stations near their home were higher.

The EPA wants to create an economic stimulus program to subsidize the **construction of new electric vehicle charging stations**. They want to choose geographic areas in which an additional new charging station will have the highest impact on electric vehicle purchases.

Preliminarily, your task is as follows:

1. Identify which demographic features are highly predictive of preference for electric vehicle purchases via an appropriate regression model to predict an appropriately formulated target variable: the strength of preference for electric vehicle in an individual's next vehicle purchase.
2. Use the methods taught in class for cross-validation, hyperparameter tuning, and model selection (amount the various available regression models) to create the best regression

model that optimizes R^2 to most strongly predict the target variable in a validation (holdout) dataset.
3. Use the chosen regression model to assess geographic areas (could be census tracts, counties, or core-based statistical areas) based on a metric you will create to predict electric vehicle demand based on demographic features of that area acquired from either the census or American Community Survey / Current Population Survey.
4. Use the electric vehicle fuel stations geographic dataset to measure geographic areas by their number of charging stations per capita
    a. Alternate: U.S. department of energy data download of electric vehicle fuel stations
5. Identify the geographic areas with the lowest ratio of (fuel stations per capita) to (your prediction of EV demand)
6. *To the extent possible*, create a cost-benefit analysis of the cost of subsidizing construction of a new charging stations vs. the benefit of emissions reduction of constructing a new station in a given geographic area

Your team must act as the investigators to determine which datasets might be helpful in finding which demographic features are predictive of electric vehicle preference. Datasets that have the potential to provide helpful insights are as follows:

- The National Household Travel Survey published by the Federal Highway Administration
- The City and County Vehicle Inventories dataset from the Department of Energy
- The Residential Energy Consumption Survey making use of the "ELECVEH" microdata variable which corresponds to "Household owns or leases an electric vehicle"
- The Global Automotive Demand Tracker - EV from S&P Mobility
- The National Renewable Energy Laboratory data catalog search for "electric vehicle"

Ideally, the regression model that you make to predict electric vehicle preference rates should be built on microdata responses to surveys, conducive to building a model on multiple features, rather than data in reports that have already been aggregated by demographic characteristics.

Ideally, your models and your proposals should rely on regularly-published data from government agencies or survey agencies, so that your workflow can be reproduced on new data published by the same agencies 1 year, 2 years, 5 years, or N years from now.

It will not be necessary to worry about processing data in batches, because at most the model will only need to generate a prediction of a demand metric on a relatively small number of 85,158 United States census tracts (at most, if that is the appropriate geographic scale, or fewer than that if a different geographic scale is more appropriate). The model will be used to create a report of recommendations for the EPA, not used in a production environment to power some app.

(Note: The Department of Transportation already subsidizes the construction of new electric vehicle charging stations through discretionary grants).