**Capstone Two - Project Proposal**

**Problem statement formation**

How can government agencies (federal and state) promote the adoption of EV to constitute 50% of the automotive industry in the next 5 years by identifying the important factor(s) or group of factor(s) that lead to the adoption in the US?

**Context**

Based on the recent announcement by President Biden on his goal towards the reduction of greenhouse emission by 50-52% by the year 2030, there will be a push for the government agencies at the state levels to attain these targets. One way of achieving this goal would be to promote the usage of electric vehicles across different use cases. While electric vehicles still require electricity to function, this power can be provided by cleaner alternate sources of energy. Electric Vehicles, as such, do not emit greenhouse gases in the immediate vicinity where they run as compared to an internal combustion engine (ICE) vehicle.

Based on the results of the analysis, government agencies can focus on the factors that were deemed to be important in the adoption of EV vehicles and focus more on developing policies to promote the adoption of EV for these cases. These policies can include things such as tax rebates, electricity rebates, and lower auto loan interest rates for the identified factors.

**Criteria for success**

The metric of success for government agencies will be lining up policies to achieve a 50% market share of automotive industry in the next 5 years as EV.

**Scope of solution space**

The focus of the project will be to identify the important factors that lead to the adoption of EV.

**Constraints**

EV support infrastructure is an important component of EV adoption. However, there is lack of information regarding EV support infrastructure, as it is a relatively new area.

**Stakeholders**

State agencies, governors, and other government agencies.

**Data sources**

The dataset that will be used for this project will be the [2017 National Household Travel Survey (NHTS)](https://nhts.ornl.gov/downloads). This survey is conducted by the federal administration (FHWA). It includes data that allows one to analyze trends in personal and household travel. It contains information on (1) daily travel linked to (2) individual personal and household characteristics, (3) socio-economic characteristics, (4) vehicle ownership and vehicle attributes.

**Problem approach**

1. Perform Data Wrangling
   1. Combine all four datasets.
   2. Prepare the dataset for processing and analysis.
   3. Make variables either numeric or categorical.
   4. Rename variables if required.
   5. Drop irrelevant variables if required.
   6. Create new variables based on existing variables.
   7. Check reliability of data.
2. Exploratory Data Analysis
   1. Identify patterns for variables and develop hypothesis.
   2. Visualize explanatory variables with output variable (whether EV vehicles were adopted).
   3. Visualize explanatory variables against each other.
   4. Build an intuition about which variables are related to whether EV vehicles were adopted or not.
3. Pre-processing
   1. Encode categorical variables.
   2. Remove or impute missing values.
   3. Scale the data.
   4. Develop the training and test data.
   5. Find a solution to the imbalanced data issue.
4. Modelling
   1. Perform classification using Logistic Regression, Random Forest etc.
   2. Find the model that has the lowest test error.
   3. Find the important variables that contribute to EV adoption.

**Deliverables**

* A GitHub repo containing the work completed for each step of the project.
  1. A slide deck
  2. A project report