FACTORS DRIVING EV ADOPTION & CHARGING STATION DEPLOYMENT

Team Eco-Warriors

MSA 8010: Data Programming

Lilly Parham, Pamela Alvarado-Zarate & Gracie Rehberg

December 10, 2024

OUTLINE

- Introduction
- Business Problem
- Solution Overview
- Dataset
- Machine Learning Model Solution

INTRODUCTION

- **Electric Vehicles (EVs):** Critical for EPA's Net-Zero targets (2030 – 2032)
  - **EPA Goal:** 35% to 56% of new car sales to be EVs
- **Current Challenge:** Only 6.9% of 2023 car sales were EVs, hindered by limited charging infrastructure
  - *Lack of chargers is a **major** barrier for potential buyers*
- **Strategic Solutions:** Data-driven strategies to optimize infrastructure investments and support climate objectives
- **Impact of Subsidies:** EPA-funded charging stations will boost EV adoption by overcoming key barriers

Environmental Protection Agency (EPA)

Goal

Electric Vehicle Adoption in the United States

Outcome

Net-Zero Emission (2030-2032)

Reducing Greenhouse Gas Emissions

Buyers in 2023

35-56%

Expected

6.9%

Actual

HIGH-LEVEL SOLUTION OVERVIEW

Data Sources

US Department of Energy

Census Data (Up to Previous Year)

Geographical Data

Data Preprocessing

Median Imputation

Label-Based Imputation ("Unknown")

Feature Selection/Extraction

Testing & Training Data

Split Ratio = 0.2

Cross Validation

ML Models

Random Forest

- Nonlinear relationships (predictive vs. target)

Ridge Regularized Linear Regression

- Address overfitting
- Provide interpretable coefficients.

Predict expected change in EV registrations when a new charging station is added to a given state (both using the same features).

THE DATASET

DATA DICTIONARY

Feature Name	Dataset	Type	Description
State	state, population, EV_registrations, fuel_stations	object	State name where the fueling station is located (e.g., “CA”)
Registration Count	EV_registration	int	Electric vehicle registrations by state in 2023
Fuel Station Count	“extracted” from fuel_station	float64	Number of fuel stations (for electric vehicles) in each state
Population	population	float64	Number of people at a certain age in each state in 2023
Median Age	“extracted" from population	float64	Average age (in years) of the population in each state
Sex Ratio	“extracted” from population	float64	Ratio of the number of males & number of females in each state
Completing College	education	object	Percentage of adults 25 years of age and older (in decimal form)
Mean Income	“extracted” from income	float64	Average income (in dollars) in each state
Area_km2		float64	Area (in km^2) of state
Pop_per_km2	“extracted” from population	float64	Number of people per km^2 by state
Fuel Stations per Capita	“extracted” from population & fuel_stations	float64	Number of fuel stations per person
EV Registrations per capita	“extracted” from population & EV_registration	float64	Number of electric vehicles per person

IN THE 50 STATES...

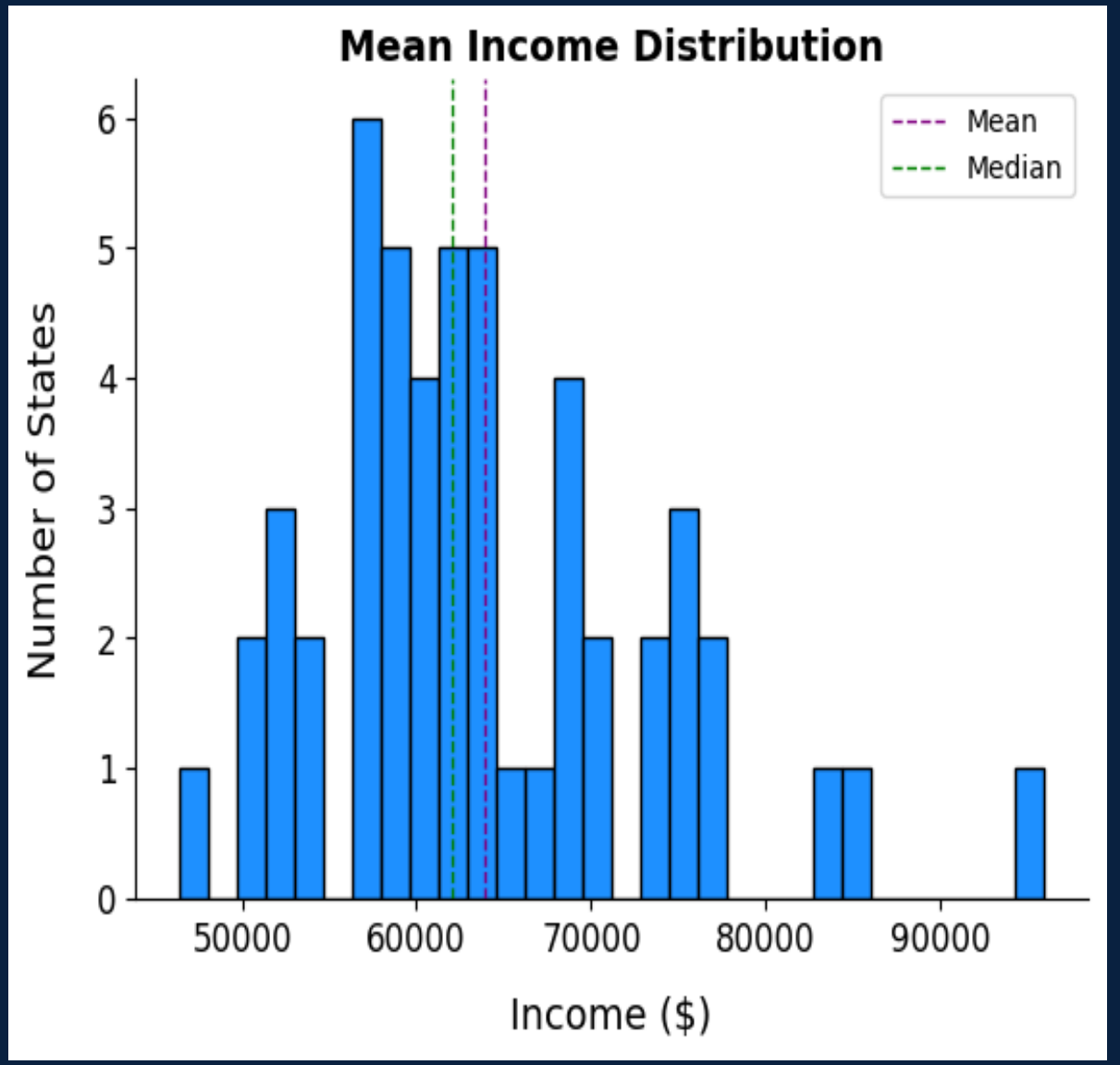
Median Age Range: 31- 44 years old

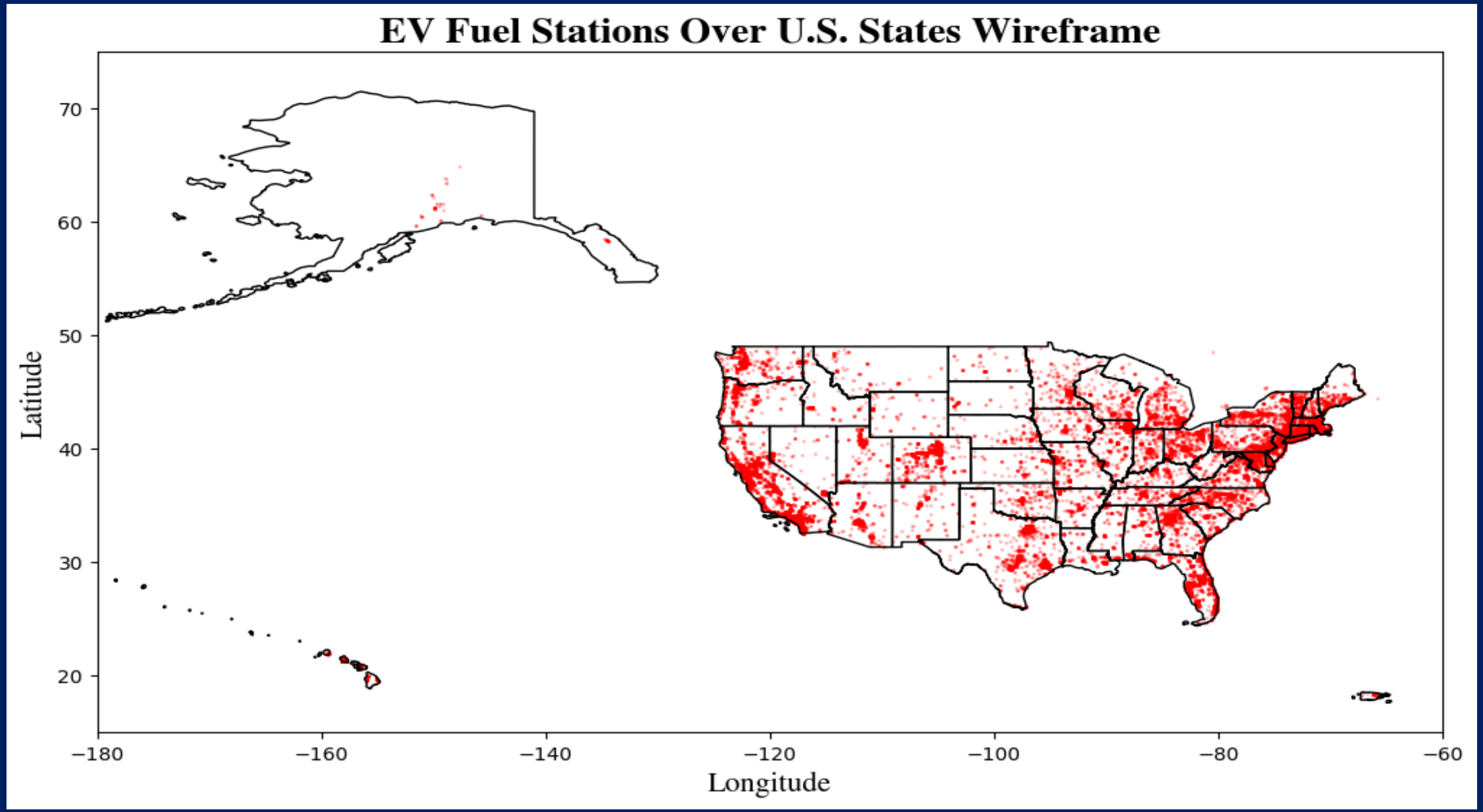
Average % of College Completion: 34%

Average Individual Income: $63,862

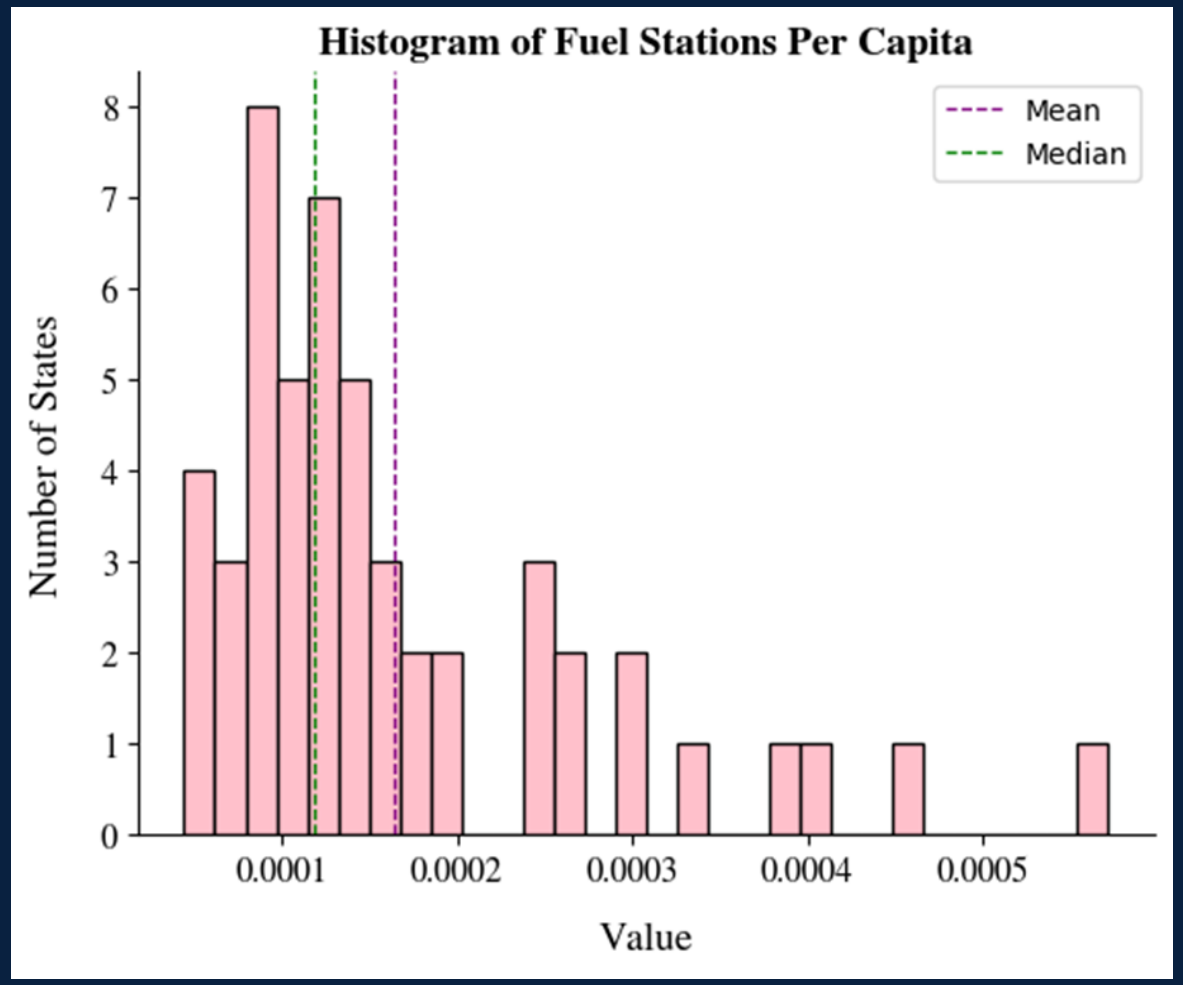
Majority of Male or Female Buyers:

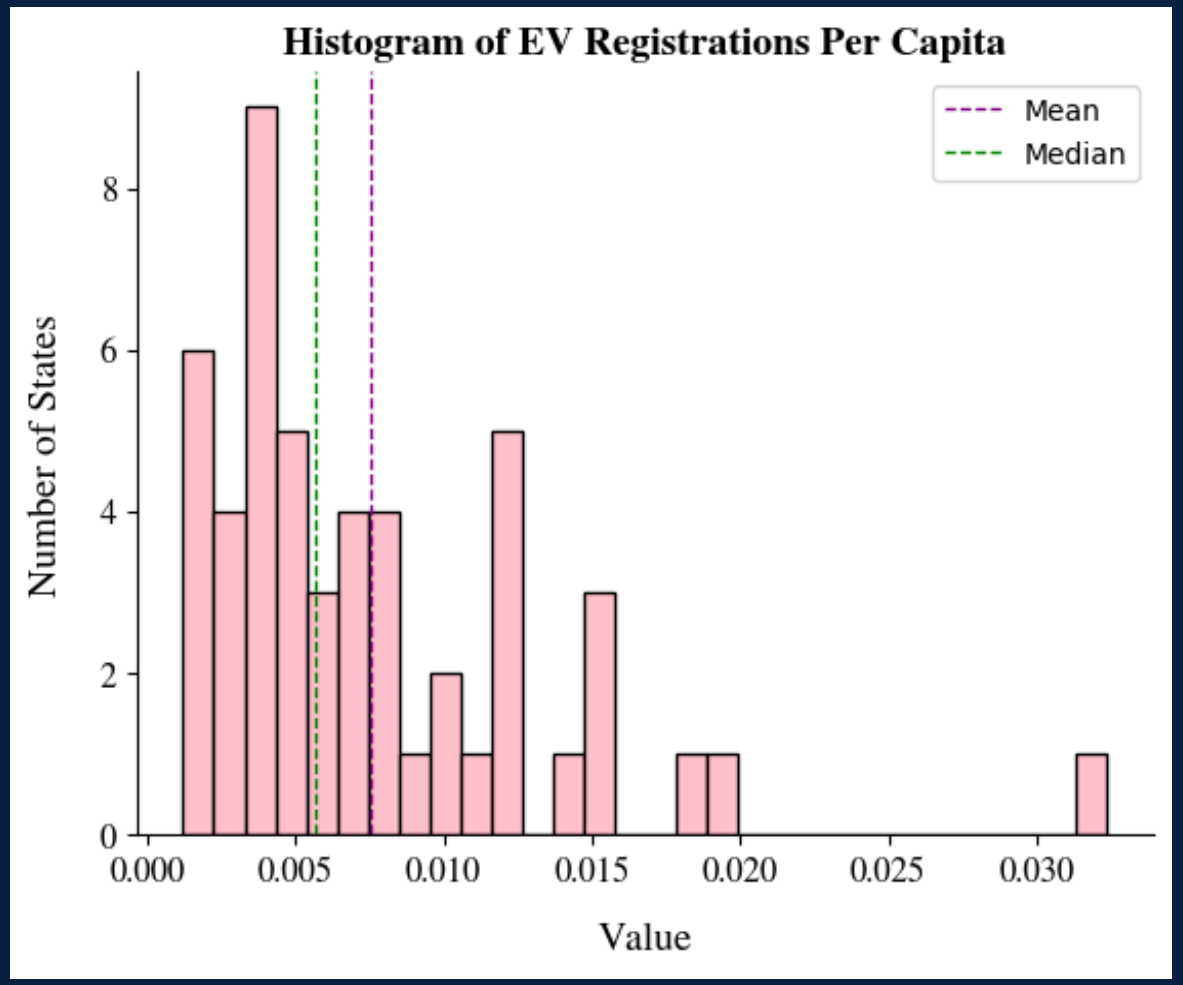
Equally (~0.98 ratio)




EV Fuel Stations Over U.S. States Wireframe
Latitude
70
60
50
40
30
20
−180
−160
−140
−120
−100
−80
−60
Longitude

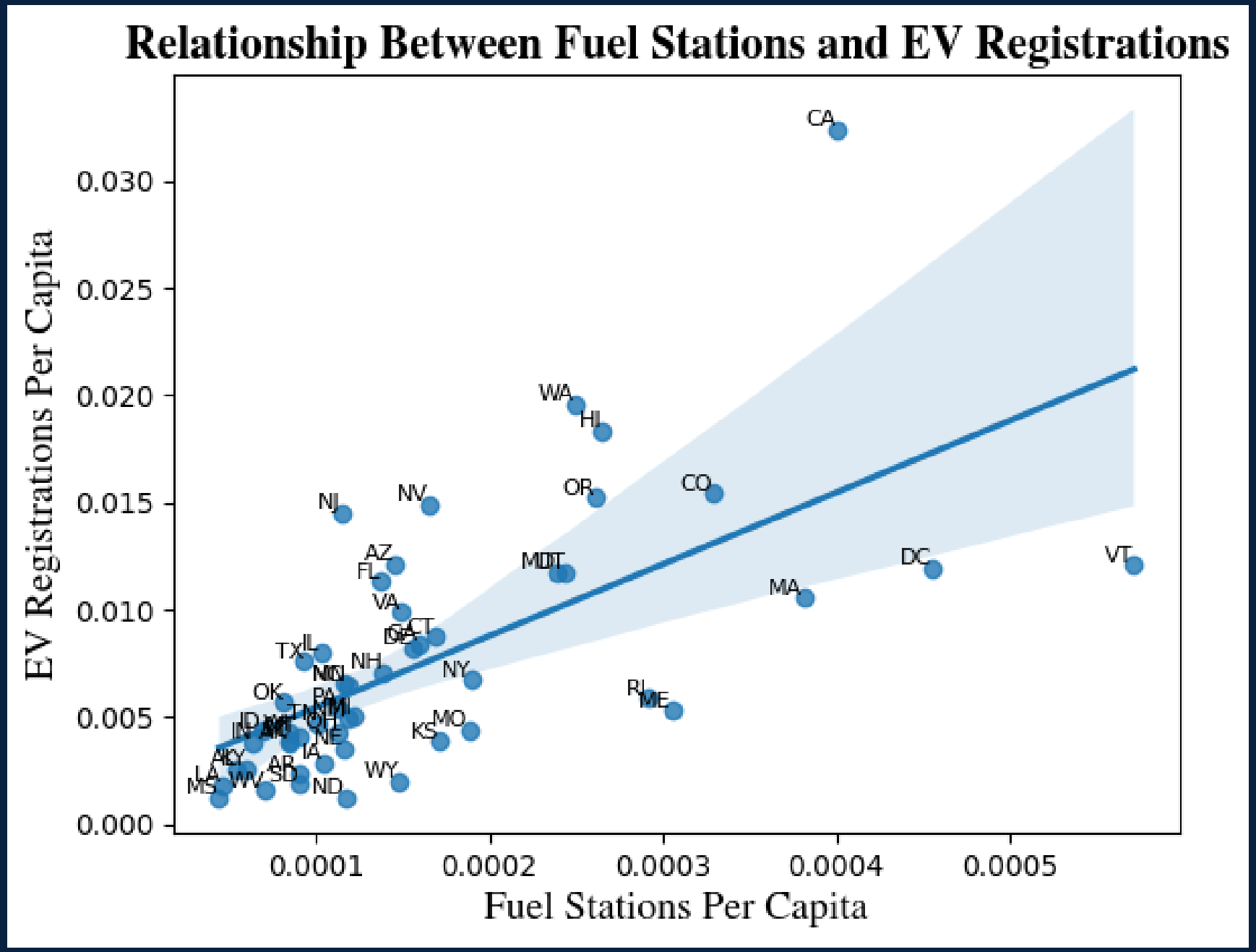
DISTRIBUTIONS





Average: 0.000164 fuel stations/capita

Average: 0.007568 fuel stations/capita



Linear Relationship
(Positive Correlation)

As the # of electric-charging stations become available within an area, the # of electric vehicle registrations increase within that area.

Leading State: CA

MACHINE LEARNING MODEL SOLUTION

BUILDING & EXPERIMENTING

Random Forest

- Multiple **(7)** hyperparameters
- Testing **96 combinations** of hyperparameters using **5-fold cross-validation.**

Ridge Regression

- **Alpha –** Defined by parameter grid for regularization variability
- Testing **(5) alphas** using 5-fold cross validation.

- Using **GridSearchCV** for hyperparameter tuning.
- **Cross-Validation MSE -** Estimate of model's average error during training.
- **Test MSE -** Measurement of model’s generalization to unseen data

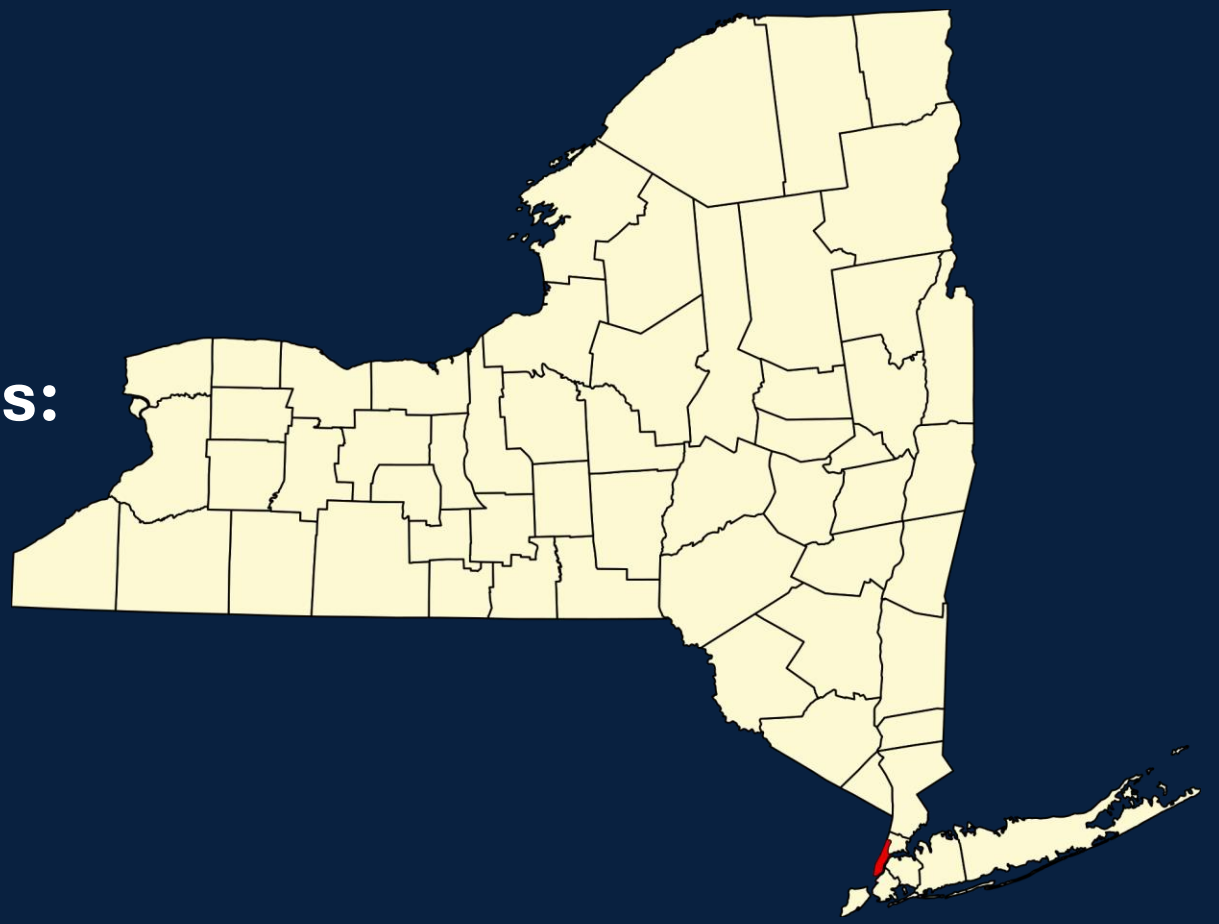
SIMULATING SCENARIO

INCREMENTING BY 1 TO CHANGE OF ELECTRIC-CHARGING STATION IN A STATE

OUTCOMES – Random Forest Regression Model

Recommendation: *Build an EV Charging station in New York County, NY*

- **Predicted # of New EV Registrations:** 2781
- 6,220 EVs registered currently
- 398 EV charging stations currently



OUTCOMES – Ridge Regression Model

Recommendation: *Build an EV Charging station in Los Angeles County, CA*

- **Predicted # New EV Registrations:** 81
- R^2 = 0.8676
- 320,110 EVs registered currently
- 3,738 EV charging stations currently

EVALUATION OF MODELS

Random Forest

- Handles complex & nonlinear relationships.
- Control of tree depth & data sampling influences model performance.
- Generalizes predictive data (T < CV)

- Overly specialized to split data, leading to potential sensitivity across dataset

Ridge Regression

- Handles multicollinearity among features
- **Regularization -** Improves model performance & prevents overfitting
- Minimizes MSE as much as possible with imputed options.

- Data variability or overestimation of training performance

MODEL MAINTENANCE

Random Forest

- Regular hyperparameter updating
- GridSearch – Optimizes model even with changes to data
- Flexible & adaptable **BUT** computationally expensive

Ridge Regression

- Rechecking alpha values (hyperparameter tuning)
- Requires feature adjustments if changes to data
- Faster & simpler **BUT** limited & rigid

- Re-training & re-testing data for performance

TAKEAWAYS

High Population

+

Strong Income Levels

+

Existing (Limited) EV Infrastructure

Greatest responsiveness to additional charging stations