1.1 Introduction: Motivation

Electric vehicles [EVs] offer great potential to reduce local air pollution, greenhouse gas emissions, and reduce oil use by the transportation sector[i]. Local governments and businesses need better tools to understand electric vehicle (EV) demand in their area to determine the placement of charging stations. Specifically, a geospatial view of the EV demand and current supply can help strategize. **Our project identifies the geographic areas where the ratio of available public charging stations to registered EVs is the least, areas most in need of additional charging infrastructure.**

1.2 Definition of the Problem

Given actual or estimated number of EVs and actual number of public charging stations in a US zip code, we segment the zip codes in a state and calculate the ratio of public charging stations to vehicles to find the areas where more public charging capacity is needed. We visualize the results as a choropleth state map of zip codes with five color gradations.

1.3 Literature Survey

In planning the public EV charging infrastructure, some papers proposed general models for locating stations in a given area[ii] or along with a given set of roads[iii], but the accurate placement of stations would need to consider local zoning laws, cost of real estate, and the local power grid. Other studies use a combination of recent EV adoption, measured by sales, demographic data, survey responses, and transportation patterns[iv],[v]. These studies limit estimation to a single state and may rely on surveys and non-public information. We did not find a model or method to use public data to measure and show geographic blocks where public EV charging stations are most needed.

Some models for predicting EV growth in the literature search are the logistic growth model[vi], choice modeling[vii], bass diffusion model[viii]. There are also several optimization models for selecting locations for charging stations. A paper by Metais et al.[ix], reviews the strengths and weaknesses of some models for optimizing EV charging stations. It identifies tour-based, also called activity-based models, work best for situation slower charging stations at a point of interest, rather than fast charging along highways. The paper offers a summary of models, but it does not offer any applications in a specific area and how changing EV adoption demand will impact the charging stations. Huang et al.[x] introduce genetic [CR1] [TM2] [CR3] algorithms or stochastic optimization tools to maximize profitability for station placement, from which our model can benefit. Although their work is limited to the Boston area and does not consider EV range limitations, our model can expand it.

Another literature source, “Optimal Siting and Sizing of Electric Vehicle Charging Stations”[xi] provides an optimization model for minimizing the cost of infrastructure to meet EV demand by locating charging stations. This paper was helpful for our project by using the optimization model to determine the best location of charging stations to reduce costs. The paper offers a model and a case study but does not provide a means of adjusting the EV demand rates. With an unclear projection of EV demand, a range of options will provide a more accurate report of costs.

In addition to EV growth models and location optimization, we found research that shows EV growth can be modeled spatially since geographic neighbors influence the EV adoption rate[xii]. Another research paper has shown EV demand exhibits spatial clustering[xiii] with characteristics such as education level, employment status, income level, population density, housing type, household size, car availability, and the presence of EVs being significant factors in explaining EV adoption rate. This emphasizes the importance of visualizing the EV demand on a map which was a feature of our project.

Our project provides several innovations for planning charging infrastructure. First, we use publicly available EV registration data to label EV adoption by zip code. For the states (a majority) that do not make registration data public, we estimate EV adoption using labeled zip codes with similar demographic and economic profiles. Second, we show areas where the charging supply/demand ratio is the smallest at the zip code level but covers the entire continental US. Third, the user can see the ratio for any zip-code in the Continental US.

1.4 Proposed Method

We first find clusters of labeled zip codes with similar predictors, then assign unlabeled zip codes to the cluster with the closest match of predictors. Finally, within each cluster, we create a regression model from the labeled data to estimate the number of EV registrations in the unlabeled zip codes.

1.4.1 Data Sources

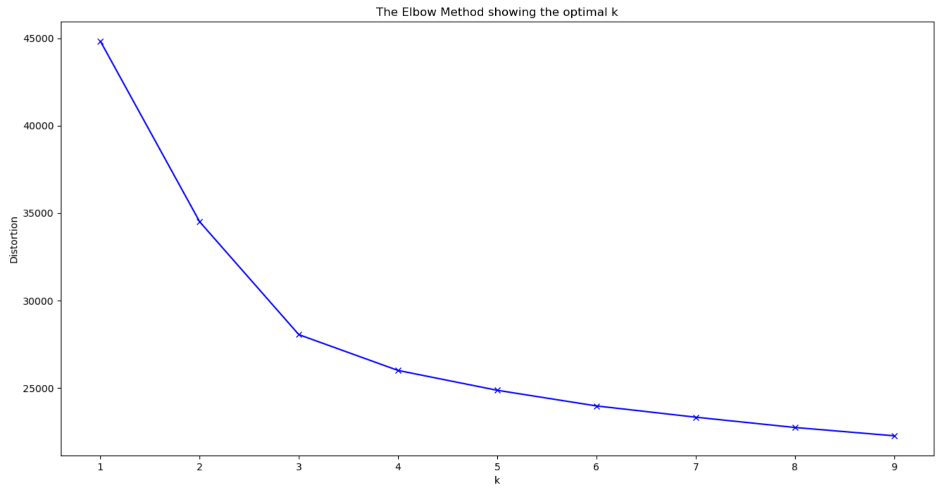
Several data sources were combined for this project using zip code as the primary key. Data availability had us using data for 2016 to provide uniform measures for comparisons. The American Community Survey (ACS) data by U.S. Census Bureau was the primary source of demographic and economic data. The location of EV charging stations was extracted from U.S. Department of Energy dataset. We are able to gather EV registration data for some states through Atlas EV HUB. Some states report EV registration by county, rather than by zip code, which we transformed using a cross-reference table from https://www.kaggle.com/datasets/danofer/zipcodes-county-fips-crosswalk which leverages US HUD and Census Bureau data sources. After combining and wrangling the data, we had a data file with over 39,000 rows (zip-codes) and close to 550 features. All data was publicly available for download.

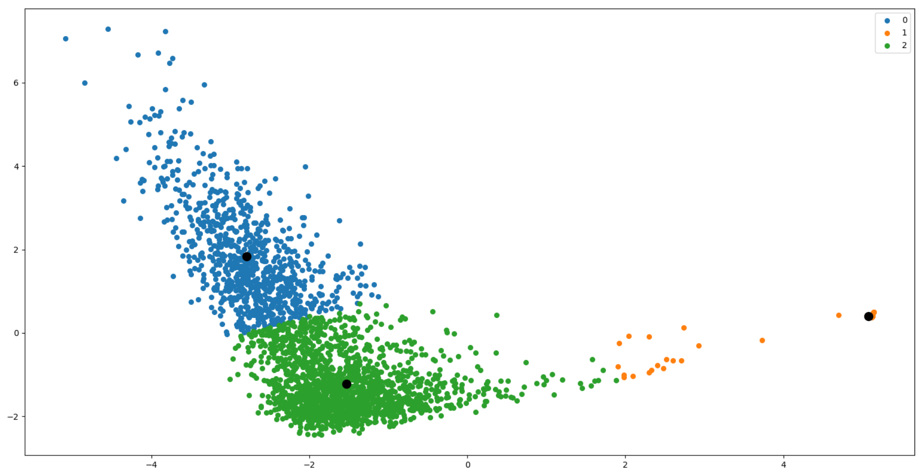
1.4.2 Models

1.4.2.1 Clustering of Zip-codes

The following steps below were performed to cluster the data points with known E.V registration numbers and then classify the data points with unknown registration numbers:

- Impute missing data for any feature with median of each column.
- Divide the data set into two. One data set with EV registration numbers and the other without EV registration numbers.
- Run the k-means clustering algorithm on the data set with registration data. Graphing the SSE vs. number of clusters showed three clusters capture most of the differences.
- Next, we ran the k-means clustering algorithm with k = 3 on the data set with EV registration data. We used the labels from this output to match the original data set. We then created another variable to identify which cluster the data points belongs (0, 1, and 2). The k-means clustering algorithm clustered the data points into 3 groups/clusters (0, 1, 2).
- Next, we used the k-nearest neighbor (KNN) classification algorithm to train the dataset from clustering model's output by using the new variable as the response variable. The mean accuracy on training data and labels is 98.15%.
- We then used the k-nearest neighbor model to predict the classification of the data points in the data set without EV registration numbers.





Some of the cities classified as 0: Pensacola, FL; Beaumont, CA; Duluth, GA; Bessemer, AL; Norcross, GA; Calabasas, CA. Some of the cities classified as 1: Willow Springs, MO; Colony, KS; Hurricane, UT; Oakville, IA; Braselton, GA. Some of the cities classified as 2: Dallas, TX; Los Angeles, CA; Orlando, FL; Minneapolis, MN; Oakland, CA; Jacksonville, FL

Some of the similarities among the zip codes in each cluster are education level, population over 25 years old. Most of the cities in cluster 0 are suburbs and most of the cities in cluster 2 are urban.

1.4.2.2 Regression Model

Since we know the actual EV registration for some zip codes we used this dataset to train a regression model and impute EV registrations for zip codes without the this information with predicted values. The steps for building the regression models are outlines below.

- Start with clustered data with registration information available
- For each cluster (0,1,2), perform features selection to identify top 10 best scoring predictors using Univariate feature selection

- Start with over 500 predictors which included granular demographic data (Education, Age, income, transportation type, gender, household characteristics, etc) along with number of charging stations, gas prices, electricity rates.
- We want to build a regression model with smaller number of features and avoid overfitting.

Several other feature selection methods were considered such as manual selection of features based on literature survey, reducing dimensionality with PCA components, and Recursive feature elimination (RFE) method. These did not yield predictions that matched actual values (R^2 values were < 0.01, even negative for some).

- For each cluster (0,1,2), train a regression model with 70% of the data and test with remaining 30%.

	Cluster 0	Cluster 1	Cluster2
MSE (train)	16713	16713	16713
MSE (test)	11449	171	1443
R.SQ (train)	0.18707	0.18707	0.18707
R.SQ (test)	0.18679	-1.05960	0.45594

The features selected for each cluster align with the characteristics of each cluster as stated. For cluster 0 with more suburbs the EV registration was correlated to education, home value, mortgage (proxy for income) and transportation. For cluster 1, with more rural areas the EV registration was correlated with level of income, unemployment, household characteristics, transportation, and rent. For cluster 2, with more urban areas the correlation was with home values, mortgage (proxy for income) and transportation type.

Once three regression models were created for each cluster label, the same models were used to predict EV registration for zip codes were the data was not available. Depending on the classification label (0,1,2), the corresponding regression model was used to yield a predicted EV registration. The predictions were clipped at a minimum of 0 and values were rounded to nearest integer.

Few observations on the predictions:

- Areas with high number of charging stations show higher EV registration
- Areas with no charging station tended to have 0 for EV registrations
- Top 10 states with highest EV reg were : CA, NY, NJ, TX, FL, PA, IL, MA, MN, and OR
- Top 10 States with most charging station: CA, NY, FL, TX, MA, WA, CO, GA, MD, VA
- This indicates some states like NJ, IL, MN, and OR might have high demand for charging stations
- Top 10 States with lowest EV reg: DE, WY, RI, AK, HI, ID, VT, ND, DC, SD
- Top 10 states with lowest # of chargers: AK, SD, ND, WY, MT, WV, MS, ID, DE, LA
- These are some of the rural or smaller states. This implies some states have larger counts perhaps because of their geographic size.
- If we focus on zipcodes, following states are still on top of the list: NJ, OR, VA, CA, WA, OR, MI, TX
- Similarly, AL, AR, AZ are in lower EV Reg list based on zipcodes.

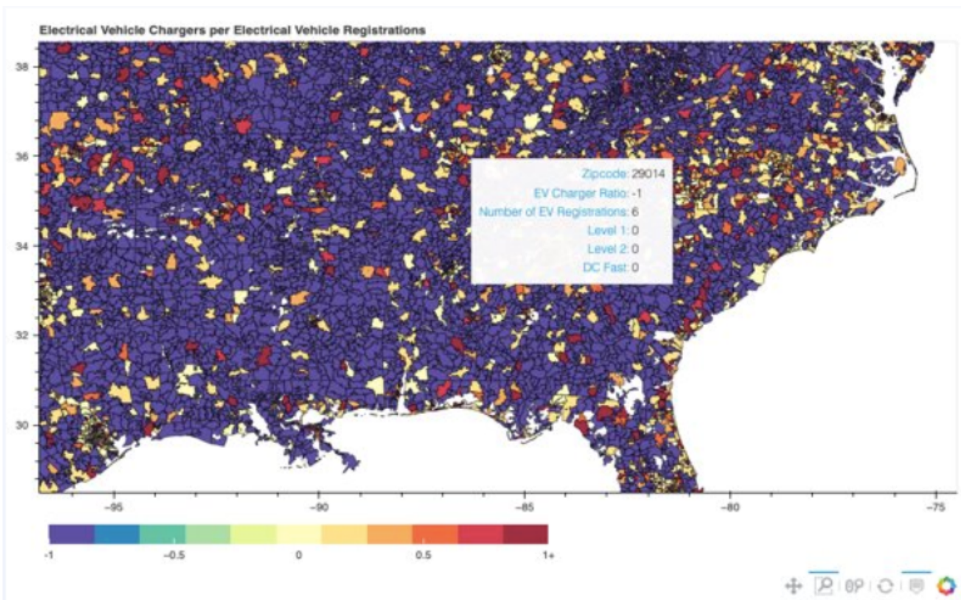
We used the OLS regression model since Herling 2019[xiv] showed it to be an effective way to model EV adoption. That paper also provides significant factors of the OLS regression model that affect the rate of

EV adoption by zip code in Massachusetts. We can use a similar method for the rest of the country. Another study by Vergis et al.[xv] describes potential explanatory variables for EV adoption, including existing charging stations and gas prices. This paper also lists data sources that can be useful when gathering data for this project.

1.4.3 Visualization

The range of ratios for chargers per registered EV is shown on a map of the lower 48 states. Each zip code is displayed with a color associated with a color gradient indicating the ratio of public charging stations to vehicles. A ratio of -1 indicates no public chargers in the zip code.

The Python library Bokeh was used to generate a US map to display the location of electric vehicle charging stations. The map provides box zoom, wheel zoom, scroll, and reset user commands. Hovering over a zip code shows a tool tip that displays the zip code, number of EVs registered, number and type of public charging stations, and the ratio of public charging stations to registered EVs. Bokeh provided the display flexibility and features we needed with less programming complexity than D3.



1.4.4 Project Plan

All team members contributed equally in time and effort to the project.

1.4.5 Testing, Scoring, and Evaluation

Our estimation of EV registration numbers depends on the assumption that zip codes with similar demographic and economic profiles have similar EV adoption rates. This assumption may be weak for similar zip code profiles in different regions of the US. For example, similar zip codes in Washington, Oregon, and California may have similar EV adoption rates, but that EV adoption rate may be lower in a Midwestern state like Wisconsin or Tennessee. We tested our assumptions about similar zip code profiles by building a classification model using a training data set that holds out one state and checking the classification accuracy using the withheld state zip codes as test data. Our regional adjustments were incomplete, with no labeled state from the Deep South – GA, NC, SC, AL, MS, LA, AR.

Our different census and geographic data sources are not all available for a recent (2020 or later) year, while the actual EV registration records are updated at least annually. Given the recent growth in EV sales, we would like to use the most current registration data. We will test if demographic and census values in a zip code for an earlier year (between 2016 and 2020) are stable enough over the period to be used with recent EV registration numbers. We will compare the model accuracy for a test/train split of older census data and older registration data (filter for vehicle registration date) against accuracy for older census data and recent registration data. Similar accuracy scores would show that census data for a zip code is stable enough to use with recent or current registration data.

1.4.6 Observations

Large portion of the country showed no chargers with at least 1 EV registered. These maybe using in-home chargers. Obtaining data for number of in-home charging stations and adding that to the map might provide more insight. It may be that a lack of public charging stations is an indicator and partial cause of slow EV adoption and that a more interesting study would be to compare EV adoption rates where the supply of charging stations seems to be leading the number of registered vehicles.

In Pennsylvania, a state familiar to some on our team, we note some zip codes in rural areas that have a larger number of charging stations to registered EVs than expected. These areas include colleges, ski and golf resorts, and a noted recreation area. Since the regression models had low coefficient of determination (less than 0.5) and perhaps businesses and attractions may be as good or better predictors for EV adoption than local demographics. Other non-demographic factors may lead to better estimation of EV adoption; such as solar panels, local environmental laws/initiatives, LEED-certified buildings however collecting data for these factors may be challenging.

It was noted that most of the charging stations are Level-2, which are much faster than Level-1. There may be a preference for faster charging stations, especially if they are located at a business or public parking. Obtaining data lower than zip-code level might help understand usage of in-home chargers vs private/public chargers.

Last, it will be interesting to see how demand and chargers change after Ford introduces the Lightning F-150 electric pick-up truck

1.5 Suggested Extensions

Some ways that the project could be extended:

- Use data on businesses and attractions (stadiums, theaters, recreational areas) to augment the demographic data in estimating EV registration for unlabeled zip-code areas
- Add a filter to set a floor on EV registrations when showing relative need for charging stations
- Allow a user to select specific zip codes, perhaps by bounding box, to augment zoom capability
- Allow user to enter a zip-code and highlight that area plus surrounding
- Compare growth of EV adoption and charger availability over blocks of time (2-year increments)
- The EV registration data included the vehicle year, make, and model. Given the rapid change in battery and optimization of “fuel” efficiency, the model could make more detailed estimates of charging needs based on vehicle characteristics.

References

[i] Hall, D., & Lutsey, N. (2017). Emerging best practices for electric vehicle charging infrastructure. *Washington, DC: The International Council on Clean Transportation (ICCT)*.
[ii] Cen, X., Lo, H. K., Li, L., & Lee, E. (2018). Modeling electric vehicles adoption for urban commute trips. *Transportation Research Part B: Methodological*, *117*, 431-454.
[iii] Metais, M. O., Jouini, O., Perez, Y., Berrada, J., & Suomalainen, E. (2022). Too much or not enough? Planning electric vehicle charging infrastructure: A review of modeling options. *Renewable and Sustainable Energy Reviews*, *153*, 111719.
[iv] Davis, A. W., & Tal, G. (2021). Investigating the Sensitivity of Electric Vehicle Out-of-Home Charging Demand to Changes in Light-Duty Vehicle Fleet Makeup and Usage: A Case Study for California 2030. *Transportation Research Record*, *2675*(10), 1384-1395.
[v] Egbue, O., Long, S., & Samaranayake, V. A. (2017). Mass deployment of sustainable transportation: evaluation of factors that influence electric vehicle adoption. *Clean Technologies and Environmental Policy*, *19*(7), 1927-1939.
[vi] Rietmann, N., Hügler, B., & Lieven, T. (2020). Forecasting the trajectory of electric vehicle sales and the consequences for worldwide CO2 emissions. *Journal of cleaner production*, *261*, 121038.
[vii] Bolduc, D., Boucher, N., & Alvarez-Daziano, R. (2008). Hybrid choice modeling of new technologies for car choice in Canada. *Transportation Research Record*, *2082*(1), 63-71.
[viii] Lee, J. H., Hardman, S. J., & Tal, G. (2019). Who is buying electric vehicles in California? Characterising early adopter heterogeneity and forecasting market diffusion. *Energy Research & Social Science*, *55*, 218-226.
[ix] Metais, M. O., Jouini, O., Perez, Y., Berrada, J., & Suomalainen, E. (2022). Too much or not enough? Planning electric vehicle charging infrastructure: A review of modeling options. *Renewable and Sustainable Energy Reviews*, *153*, 111719.
[x] Huang, Y., & Kockelman, K. M. (2020). Electric vehicle charging station locations: Elastic demand, station congestion, and network equilibrium. *Transportation Research Part D: Transport and Environment*, *78*, 102179.
[xi] Jia, L., Hu, Z., Song, Y., & Luo, Z. (2012, March). Optimal siting and sizing of electric vehicle charging stations. In *2012 IEEE International Electric Vehicle Conference* (pp. 1-6). IEEE.
[xii] Liu, X., Roberts, M. C., & Sioshansi, R. (2017). Spatial effects on hybrid electric vehicle adoption. *Transportation Research Part D: Transport and Environment*, *52*, 85-97.
[xiii] Morton, C., Anable, J., Yeboah, G., & Cottrill, C. (2018). The spatial pattern of demand in the early market for electric vehicles: Evidence from the United Kingdom. *Journal of Transport Geography*, *72*, 119-130.
[xiv] Herling, S. (2019). *Modeling electric vehicle adoption in Massachusetts: Problems, opportunities, and implications for municipalities and utilities* (Order No. 22588487).
[xv] Vergis, S., & Chen, B. (2015). Comparison of plug-in electric vehicle adoption in the United States: A state by state approach. *Research in Transportation Economics*, *52*, 56-64.