Investigating Electric Vehicle Adoption Using Correlation and Prediction Analyses

Fahad S. Alrasheedi and Hesham H. Ali
Department of Computer Science
University of Nebraska at Omaha
Omaha, USA
e-mail: {falrasheedi, hali}@unomaha.edu

***Abstract*—The transportation sector, dominated by gas-powered vehicles, is a major contributor to carbon dioxide emissions that pose significant threats to both environmental and public health. To address this issue, Electric Vehicles (EVs) have emerged as a promising alternative aimed at achieving zero-carbon emissions. However, EV adoption faces several challenges, including high costs, insufficient charging infrastructure, range anxiety, and other barriers. To promote EV adoption, authorities responsible for the management of EVs have implemented various incentives, such as tax reductions, credits, and support for charging infrastructure programs. Despite these targeted management efforts, the adoption of EVs remains a complex issue that requires extensive analysis to understand the factors driving increases or decreases in adoption rates. In this study, we employ a two-pronged approach to examine EV adoption growth rates across counties in six U.S. states. Our methodology integrates correlation network analysis and statistical prediction-based analysis. The primary finding of these analyzes highlights the critical role of geographical features and practices of local management of EVs in influencing similar patterns of EV adoption among counties. Additionally, we identify two clusters exhibiting declines in EV adoption, underscoring the need for further investigation into the management strategies and underlying causes of these decreases.**

***Keywords*- electric vehicle; charging stations; electric vehicle adoption; graph modeling, correlation networks.**

I. INTRODUCTION

The transportation sector is a major contributor to carbon dioxide (CO2) emissions, which pose a significant threat to life on Earth. For example, in the United States, 29% of CO2 emissions are caused by the transportation sector, which relies heavily on greenhouse gases such as gasoline. Light vehicles alone account for more than half of the transportation sector's emissions [1][2].

Electric Vehicles (EVs) are widely regarded as a replacement for gasoline-powered vehicles. However, EV adoption (represented by the number of EVs) faces several challenges, including high costs, insufficient charging infrastructure, range anxiety (i.e., the concern that the battery's remaining charge may not be sufficient to reach the next stop), and other barriers. Consequently, significant managment efforts have been made to transition the transportation sector toward electrification. For instance, U.S. authorities manage EV adoption by offering incentives such as tax reductions and credits for purchasing EVs and supporting various programs to enhance charging infrastructure.

Despite such management efforts to promote EV usage, the EV adoption remains a complex issue that requires in-depth investigation to provide insights into how adoption rates can be increased based on the characteristics of targeted populations.

In this study, we focus on counties in the U.S. We conduct our analysis at the county level rather than at the state or zip code level because states are too broad, while zip codes are too narrow to effectively capture differences in EV adoption behavior across regions. Therefore, an essential first step in addressing the complexity of EV adoption is to examine how different counties across various states in the U.S. are working to accelerate EV adoption.

We conducted two analyses as part of this effort: one using Graph Theory and the other employing statistical prediction analysis. Graph Theory has been applied in the EV domain as a method to optimize the distribution of charging stations [3]–[9]. On the other hand, statistical analyses have been used in studies to investigate the impact of charging stations and other factors on EV adoption [10]–[16]; however, these studies typically focus on one to three cities.

In our Graph Theory analysis, we leveraged a correlation network to build a network of counties and clustered them based on their correlations of EV growth rates. This approach identified several clusters of correlated counties. Counties within the same cluster exhibited similar EV adoption behaviors, opening avenues for future research to understand the reasons behind these shared behaviors.

The second analysis involved building various prediction models to forecast EV adoption in a county based on its demographic features. The best-performing model was selected, and further analyzed to identify significant features.

Our findings from the correlation network revealed that counties within the same cluster often belong to the same state and are geographically close to one another. This suggests that local managements and neighboring areas may play a significant role in EV adoption. Additionally, some clusters showed declines in EV growth rates, prompting the need for further studies to investigate the causes of these decreases.

In the statistical prediction-based analysis, the Gradient Boosting model emerged as the best-performing prediction model. Among the significant features identified in the best prediction model, the geographical feature 'Federal Information Processing Standards (FIPS)' stood out, aligning with the findings from the correlation network analysis. Hence, a local management's strategy for EV adoption may be influenced by both the characteristics of their own region and the strategies of neighboring regions in adopting EVs.

The remainder of this paper is organized as follows: Section II discusses our approach for employing Graph Theory to build the EV adoption correlation network and the development of

TABLE I. NUMBER OF COMPLETED COUNTIES BY STATE

No.	State	No. of Counties
1	Colorado	20
2	Minnesota	3
3	Montana	2
4	New York	48
6	Texas	30
7	Virginia	34
8	Total	137

prediction models. Section III discusses the results, followed by the conclusion and future work in Section IV.

II. METHODOLOGY

In this section, we describe the data collection process for this study, the application of Graph Theory in our analysis, and the development of prediction models.

A. Data Collection

1) EV Data: Atlas Hub [17] provides temporal data on EV registrations at the zip code level for several states in the U.S. For this study, we selected states that offered data from 2018 to 2023 and aggregated the data at the county level. We chose this time range based on data availability, as increasing the range results in a smaller number of states and counties, while decreasing the range shortens the time series and may negatively impact the analysis.

Consequently, we identified 137 counties from six states that provided a complete 12 months of EV registration data for each year within the study period. Table I presents the number of counties per state.

This study includes all EVs registered in each state, regardless of their usage purpose, such as personal or commercial, and whether they are light-duty or heavy-duty. The impact of usage purpose on EV adoption is worth further investigation in the future.

2) Charging Station Data: Charging station data is required as a predictor in the statistical prediction models. We collected the number of stations for each county of interest from the Alternative Fueling Station Locator [18]. Using the establishment dates for each station, we aggregated the number of stations established annually in each county. For the analysis, we used the number of stations as of 2022 to predict the number of EVs in 2023 (as explained in II-C), incorporating a one-year lag.

3) Demographic Data: This data was retrieved at the county level from the official Census Bureau of the United States [19]. The dataset, covering the period from 2017 to 2022, includes approximately 58 features categorized into the following groups: Population, Age and Sex, Race and Hispanic Origin, Population Characteristics, Housing, Families Living Arrangements, Computer and Internet Use, Education, Health, Economy, Transportation, Income Poverty, Business, and Geography.

B. The Correlation Network Method

First, we computed the month-to-month growth rates for each county in our study, resulting in 72 data points of growth rates per county. These growth rates were calculated using the equation:

$$\frac{Current\ Month - Previous\ Month}{Previous\ Month}$$

where *Current Month* means the cumulative number of EVs until the current month, and *Previous Month* means the the cumulative number of EVs until the previous month.

Next, since our data are not perfectly linear, we calculate the Spearman correlation [20] between counties, resulting in a 137 × 137 correlation matrix. Using this matrix, we created a correlation network where nodes represent counties and edges represent correlations that exceed a specified threshold. After testing several thresholds, we found that the optimal threshold for our case study was 0.72, which yielded clusters of correlated counties.

C. Statistical Prediction Analysis

Our second analysis leveraged the nature of our data, which includes 137 counties across multiple U.S. states, to build cross-sectional prediction models for estimating the number of EVs at the county level for a specific year. Specifically, we focused on predicting the number of EVs in 2023 using the following approach:

1) The target variable was the number of EVs in 2023.
2) The features included demographic data from the Census Bureau and the cumulative number of charging stations as of 2022, reflecting a one-year effect of charging stations on the number of EVs in 2023.
3) The statistical prediction models used included Linear Regression, Random Forest, Gradient Boosting, Decision Tree, Elastic Net, Lasso, and Ridge.

Finally, we identified the most significant features in the best-performing prediction model.

III. RESULTS AND DISCUSSION

In this section, we present the outcomes of our analysis, including the identification of clusters based on EV adoption patterns and the evaluation of our prediction models. We highlight the most significant features identified in our Gradient Boosting model and discuss their implications.

A. Graph Theory Based Clustering

First, the number of counties meeting our correlation threshold is 40 out of 137 counties. Among the correlations between these counties, we identified four main clusters, as shown in the correlation network in Figure. 1. Table III shows the number of counties and their corresponding states for each cluster in the resulting correlation network.

We observed that the correlated counties in each cluster belong to a single state. For instance, the counties in clusters 1, 2, 3, and 4 are from Colorado, New York, Texas, and Minnesota, respectively. Hence, our primary finding in this analysis is

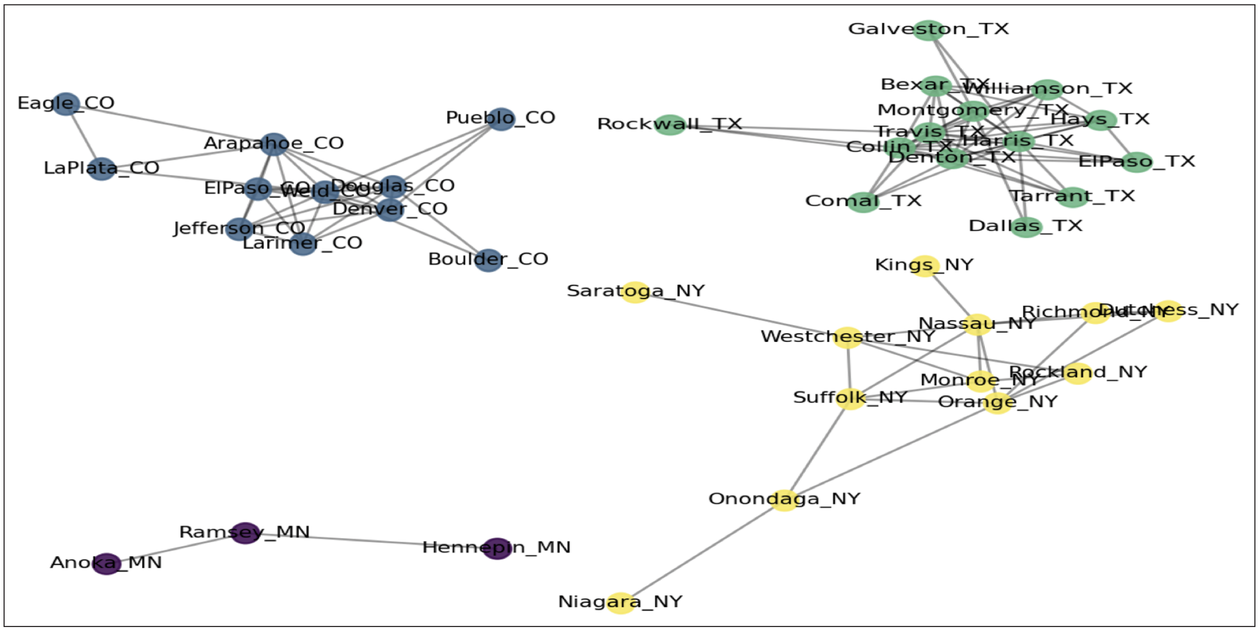

Figure 1. Correlation Network: Nodes represent counties, with labels indicating the county name appended with the state abbreviation, where colors distinguish different states (e.g., Saratoga_NY represents Saratoga County in New York). Edges correspond to correlations exceeding 0.72.

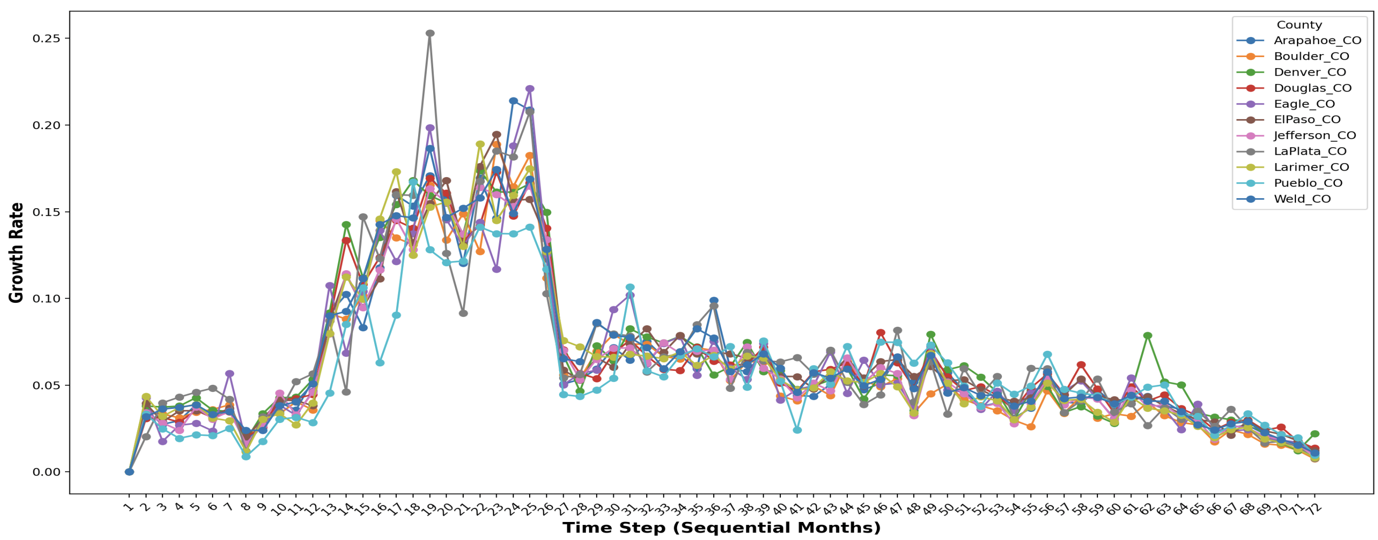

Figure 2. Growth rates of counties in cluster 1 (Colorado). The X-axis represents 72 months, from January 2018 to December 2023.

TABLE II. THE SEVEN MOST SIGNFICANT FEATURES IN THE GRADIENT BOOSTING MODEL

Feature	Group	Importance
Nonminority-owned employer firms, Reference year 2017	Business	5.26e-01
Living in same house 1 year ago, percent of persons age 1 year+, 2018-2022	Families & Living Arrangements	2.02e-01
Station Counts	Station data	4.96e-2
Total annual payroll	Business	4.75e-2
Men-owned employer firms, Reference year 2017	Business	4.60e-2
Women-owned employer firms, Reference year 2017	Business	1.66e-2
FIPS Code	Geography	1.05e-02

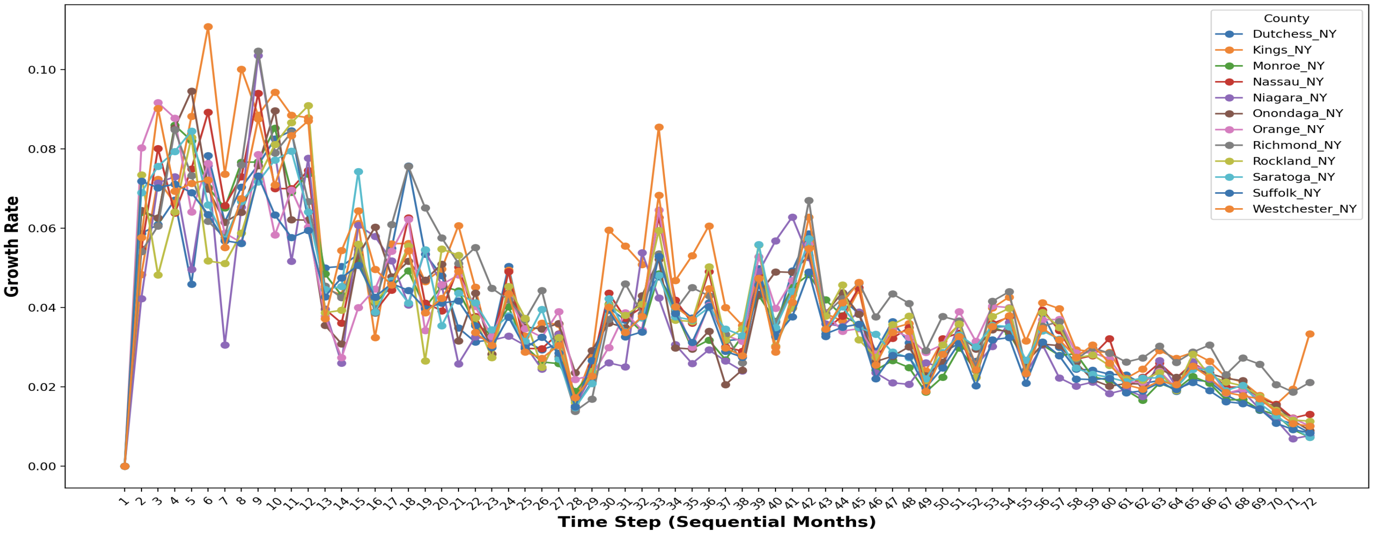

Figure 3. Growth rates of counties in cluster 2 (New York). The X-axis represents 72 months, from January 2018 to December 2023.

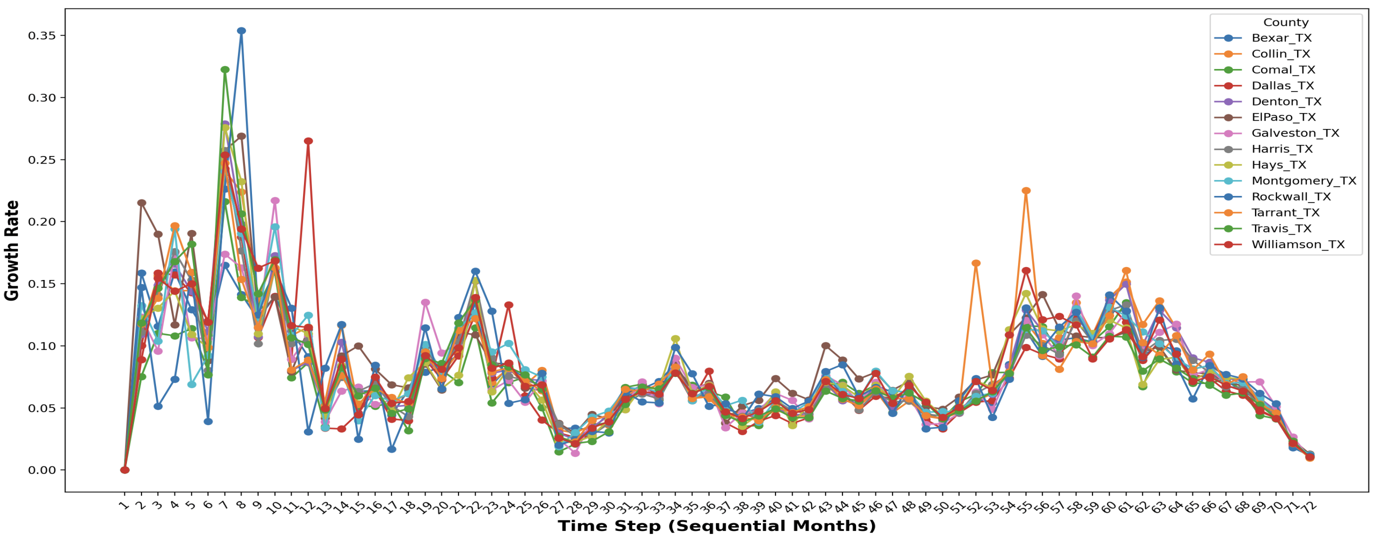

Figure 4. Growth rates of counties in cluster 3 (Texas). The X-axis represents 72 months, from January 2018 to December 2023.

that correlated counties tend to cluster geographically within individual states. Furthermore, beyond manual investigations, these correlated counties often appear to be neighbors within the same state. This suggests that the management strategies of neighboring regions and the geographical characteristics of counties may play a significant role in driving EV adoption.

Furthermore, we visualized the growth rates of the counties in Colorado cluster, Texas cluster, and New York cluster in Figures. 2, 4, and 3, respectively (we ignored the Minnesota cluster since it only contained three counties). These visualizations reveal the strength of correlations within each cluster. Interestingly, the growth rates in Colorado and New York tend to decline, highlighting the need for further investigation to understand the underlying causes in these counties. Such insights could help local authorities manage and address this decline in EV adoption more effectively.

TABLE III. THE FOUR CLUSTERS FOUND IN THE CORRELATION NETWORK, HOW MANY COUNTIES IN EACH CLUSTER, AND THE STATES OF THESE COUNTIES

Cluster Code	No. of Counties	States
Cluster 1	11	Colorado
Cluster 2	12	New York
Cluster 3	14	Texas
Cluster 4	3	Minnesota

TABLE IV. COMPARISON OF SEVERAL ML MODELS IN PREDICTING EV ADOPTION

Model	MSRE	R-Squared
Linear Regression	101850027.5567	0.5377
Random Forest	69206289.8	0.6859
Gradient Boosting	**58108466.13**	**0.7362**
Decision Tree	141672157.654	0.357
Elastic Net	122403797.77	0.444
Lasso	98257697.207	0.554
Ridge	96591486.7415	0.5616

B. Prediction models

We applied statistical prediction models to predict the number of EVs at the county level. These models were evaluated using metrics such as mean squared regression error (MSRE) and adjusted R-squared. The models tested include Linear Regression, Random Forest, Gradient Boosting, Decision Tree, Elastic Net, Lasso, and Ridge Regression. Table IV compares the performance of these models, with Gradient Boosting emerging as the best performer. It achieved MSRE of 58108466.13, and adjusted R-squared of 0.7362, explaining 73.62% of the variability in EV numbers.

Finally, we prioritized features based on their importance in the Gradient Boosting model and identified the top seven features, as shown in Table II. Among these, the FIPS feature emerged as one of the most significant predictors of EV adoption at the county level. The FIPS feature, being geographical in nature, aligns with our findings in the correlation network, where counties from the same state tend to cluster together. This highlights the influence of local authorities and geographic location on EV adoption behavior.

IV. Conclusion and Future Directions

We presented a two-pronged analysis of EV adoption in counties across six U.S. states. The first approach utilized a correlation network from Graph Theory, where nodes represent counties and edges indicate correlations in their EV growth rates. We then clustered the counties based on these correlations. The second approach involved developing various statistical prediction models to forecast EV adoption in 2023 using demographic and charging station data as predictors. The best-performing model was selected and further analyzed to identify significant features.

Our key finding is that the geographical characteristics of counties, such as the state in which a county is located and its neighboring counties, play a significant role in EV adoption. This is evident in the correlation network, where counties within the same state exhibit similar EV growth rate patterns, and in the prediction model, where the FIPS feature (a geographical identifier) emerges as one of the most significant predictors in the best-performing model.

Additionally, we identified two clusters with declining EV growth rates, highlighting the need for further investigation into their underlying causes. Future research could enhance prediction models by incorporating political, environmental, and climatic factors while also expanding the dataset to cover more counties across states. More specifically, an in-depth exploration of how gas prices interact with EV adoption remains a promising area of study. Lastly, distinguishing between different types of EVs in future adoption analyses may yield valuable insights.

References

[1] United States Environmental Agency (Epa), *Inventory of U.s. greenhouse gas emissions and sinks*, en. North Charleston, SC: Createspace Independent Publishing Platform, Jul. 2018.

[2] O. Us Epa, "Fast facts on transportation greenhouse gas emissions", en, Aug. 2015.

[3] M. A. Moni and P. Liò, "Network-based analysis of comorbidities risk during an infection: Sars and hiv case studies", *BMC bioinformatics*, vol. 15, pp. 1–23, 2014.

[4] A. Gagarin and P. Corcoran, "Multiple domination models for placement of electric vehicle charging stations in road networks", *Computers & Operations Research*, vol. 96, pp. 69–79, 2018.

[5] E. M. Arkin, P. Carmi, M. J. Katz, J. S. Mitchell, and M. Segal, "Locating battery charging stations to facilitate almost shortest paths", *Discrete Applied Mathematics*, vol. 254, pp. 10–16, 2019.

[6] S. Funke, A. Nusser, and S. Storandt, "Placement of loading stations for electric vehicles: No detours necessary!", *Journal of Artificial Intelligence Research*, vol. 53, pp. 633–658, 2015.

[7] S. Funke, A. Nusser, and S. Storandt, "Placement of loading stations for electric vehicles: Allowing small detours", in *Proceedings of the international conference on automated planning and scheduling*, vol. 26, 2016, pp. 131–139.

[8] P. K. Agarwal, J. Pan, and W. Victor, "An efficient algorithm for placing electric vehicle charging stations", in *27th International Symposium on Algorithms and Computation (ISAAC 2016)*, Schloss Dagstuhl–Leibniz-Zentrum für Informatik, 2016, pp. 7–1.

[9] T. G. Altundogan, A. Yildiz, and E. Karakose, "Genetic algorithm approach based on graph theory for location optimization of electric vehicle charging stations", in *2021 Innovations in intelligent systems and applications conference (ASYU)*, IEEE, 2021, pp. 1–5.

[10] C. Ledna, M. Muratori, A. Brooker, E. Wood, and D. Greene, "How to support ev adoption: Tradeoffs between charging infrastructure investments and vehicle subsidies in california", *Energy Policy*, vol. 165, p. 112 931, 2022.

[11] R. Chen, Q. Meng, and J. J. Yu, "Optimal government incentives to improve the new technology adoption: Subsidizing infrastructure investment or usage?", *Omega*, vol. 114, p. 102 740, 2023.

[12] L. V. White, A. L. Carrel, W. Shi, and N. D. Sintov, "Why are charging stations associated with electric vehicle adoption? untangling effects in three united states metropolitan areas", *Energy Research & Social Science*, vol. 89, p. 102 663, 2022.

[13] H. A. U. Khan, S. Price, C. Avraam, and Y. Dvorkin, "Inequitable access to ev charging infrastructure", *The Electricity Journal*, vol. 35, no. 3, p. 107 096, 2022.

[14] G. H. Broadbent, C. I. Allen, T. Wiedmann, and G. I. Metternicht, "Accelerating electric vehicle uptake: Modelling public policy options on prices and infrastructure", *Transportation Research Part A: Policy and Practice*, vol. 162, pp. 155–174, 2022.

[15] U. Illmann and J. Kluge, "Public charging infrastructure and the market diffusion of electric vehicles", *Transportation Research Part D: Transport and Environment*, vol. 86, p. 102 413, 2020.

[16] B. Haidar and M. T. A. Rojas, "The relationship between public charging infrastructure deployment and other socio-economic factors and electric vehicle adoption in france", *Research in Transportation Economics*, vol. 95, p. 101 208, 2022.

[17] N. Nigro, *State EV registration data*, en, https://www.atlasevhub.com/materials/state-ev-registration-data/, Accessed: 2024-5-6.

[18] *Alternative fueling station locator*, en, https://afdc.energy.gov/stations, Accessed: 2024-5-6.

[19] United States Census Bureau, "United states census bureau: Data and statistics", 2024, [Online]. Available: https://www.census.gov (visited on 06/13/2024).

[20] C. Spearman, "The proof and measurement of association between two things", *The American Journal of Psychology*, vol. 15, no. 1, pp. 72–101, 1904.