Electrifying the Road: Understanding Adoption Patterns and Policy Impacts on EV Growth

George Mason University

AIT-582-002 | Prof. Dr. Can Nguyen

Raja Ruthvik Shetty

Phani Satya Sai Pamarthi

Dhanushi Panga

Rithvik Madhavaram

Kashyap Kandibanda

Mano Harsha Sappa

***Abstract* - Electric Vehicles (EVs) are increasingly becoming central to the global push for clean energy and sustainable mobility. While governments are providing incentives and setting mandates to accelerate the transition, EV adoption rates still vary significantly across different regions due to economic, infrastructural, and policy-driven factors. This research seeks to identify the underlying drivers of EV adoption in the United States using big data analytics. By integrating datasets related to EV registrations, charging infrastructure, socioeconomic indicators, and state-level policies, this study aims to uncover patterns and correlations that explain the uneven landscape of EV uptake. Preliminary findings suggest that charging station density and median household income are among the most influential variables. Urban areas with strong infrastructure investments and supportive policies lead in adoption, whereas rural regions lag. The paper outlines the data ingestion, preprocessing, and exploratory analysis steps completed thus far, and presents a roadmap for developing predictive models that can simulate policy scenarios and forecast future adoption trends.**

Keywords—electric vehicles, EV adoption, charging infrastructure, policy analysis, socioeconomic factors, big data analytics, machine learning, transportation sustainability, predictive modeling, regional disparity.

1. INTRODUCTION

The transportation industry is undergoing substantial development, with electric vehicles (EVs) emerging as a key component of worldwide efforts to tackle climate change. In the United States, transportation accounts for around 29% of total greenhouse gas emissions, the highest share of any industry. As worries about fossil fuel dependency and air pollution mount, EVs provide a sustainable option by producing zero tailpipe emissions and allowing for integration with renewable energy sources.

Despite increased awareness, technological advancements, and federal and state incentives, the rate of EV adoption varies greatly among locations. California, Oregon, and Washington have seen significant increases in EV adoption because of regulatory measures, public infrastructure investment, and environmental consciousness. Other regions, particularly the Southeast and Midwest, continue to fall behind. This gap is caused by a complex interplay of factors such as socioeconomic level, urbanization, legislative backing, public opinion, and charging infrastructure availability, rather than just technology access.

Understanding these disparities in adoption patterns is crucial as the country works toward a more equal and complete transition to clean mobility. For EVs to reach their full potential, policies and investments must be based on data-driven insights that consider regional variances and barriers. It is insufficient to rely on broad national trends; rather, specific analysis is required to determine what promotes or hinders EV adoption at the state or county level.

The goal of this study is to uncover the important characteristics that drive EV growth using machine learning techniques on a huge dataset of electric car registrations and associated demographic and infrastructure attributes. This study intends to accomplish the following:

- Quantify the influence of charging infrastructure and income on EV adoption.
- Identify geographical groups with comparable adoption behaviors.
- Create targeted suggestions for policymakers and stakeholders.

By bridging the gap between technology innovation and regional preparation, this initiative helps to create a more inclusive roadmap for boosting electric car adoption in the United States.

1.1. Problem Description:

Despite major advances in electric vehicle (EV) technology and supportive regulatory measures, adoption rates continue to vary by location in the United States. States like as California have seen extensive adoption due to infrastructure expenditures and strong incentives, whilst other states continue to lag due to restricted access to charging stations, low consumer awareness, and unfavorable economic conditions.

This discrepancy undermines national clean energy targets and fair access to sustainable mobility. Policymakers and stakeholders require a more in-depth understanding of what promotes or hinders EV adoption beyond broad averages. The lack of a shared knowledge of how vehicle attributes, infrastructure proxies (such as electric range), and temporal patterns influence adoption reduces the effectiveness of future planning.

The goal of this work is to use data science techniques to bridge that gap. Using a huge dataset of EV registrations, the study identifies trends, clusters, and predictors that provide both analytical insight and policy implications. The methodology combines regression, clustering, and exploratory data analysis to provide actionable findings.

2. LITERATURE REVIEW

From transportation planning to environmental policy, energy economics to consumer behavior, the study on electric vehicle (EV) adoption crosses several fields. Although the adoption of electric vehicles (EVs) has been fast in many places, academics are still investigating why this expansion is uneven and which elements most greatly influence adoption patterns.

One of the most often referenced research, done by Sierzchula et al. (2014), looked at how government actions affected EV adoption and found that financial incentives including tax rebates, direct subsidies, and lower registration fees favorably affect EV uptake. Their study across several nations showed that the size and visibility of these incentives directly influence adoption. Gallagher and Muehlegger (2011) equally shown that, particularly for middle-income families, state tax subsidies in the United States were essential in helping to make EVs more affordable.

Another important factor of EV readiness is charging infrastructure. Accessible and conspicuous public charging stations help to alleviate range anxiety, a typical psychological barrier that discourages possible EV customers. Neaimeh et al. (2017) presented geographical models to assess charging network coverage and showed how the calculated positioning of fast-charging stations can greatly increase user confidence. Funke et al. (2019) underlined that, particularly in cities where home charging might not be possible, dense and dependable infrastructure connects with more EV adoption.

Recent developments in machine intelligence and big data have created fresh paths for investigating EV adoption at more granular levels. Zhang and Xie (2020) used spatiotemporal deep learning to find correlations between EV registration spikes and the development of charging stations. Their results showed that infrastructure usually comes before adoption, implying that proactive investment can drive EV market expansion in underprivileged areas. At the same time, Feng et al. (2020) examined regional policy efficacy in China using unsupervised clustering, hence stressing the requirement of locally customized interventions.

Apart from financial and infrastructural factors, customer perception has a complex influence. Axsen and Kurani (2013) looked examined how peer behavior, environmental knowledge, and social influence affected EV purchase choices. Their longitudinal studies showed that people are more inclined to embrace EVs if they fit with their lifestyle identity or if friends in their network have already switched. Increasingly acknowledged as important elements are behavioral and psychological motivations including perceived environmental responsibility, brand trust, and even media narratives.

Current research lacks a comprehensive, data-driven paradigm integrating these several variables—economic, infrastructural, behavioral—that may mimic policy results and adoption patterns. By use of predictive modeling combined with structured datasets, this work aims to close that gap and provide useful insights for urban planners and legislators.

3. RESEARCH QUESTIONS AND HYPOTHESIS

Though the design and marketing of electric cars (EVs) have come a long way, their mainstream acceptance in the United States is still patchy. This disparity raises numerous significant issues regarding the fundamental reasons behind regional adoption trends. The study aims to uncover the elements—from infrastructure and policy to economic and social variables—that shape EV adoption to handle this complexity.

The main goal is to study how infrastructure availability, socioeconomic circumstances, environmental awareness, and government incentives interact to shape adoption rates. The study also investigates if machine learning techniques can effectively group areas with comparable adoption tendencies and how these elements differ across urban and rural environments.

Research Questions:

i. **RQ1:** Which vehicle-related characteristics most directly affect patterns of EV adoption?

   This topic seeks to measure how technical features—such as electric range, fuel type (BEV vs. PHEV), and model year—affect adoption patterns. Attributes that best forecast vehicle uptake were found using linear regression and random forest models.

ii. **RQ2:** Can clustering, an unsupervised learning technique, reveal latent groupings in EV attributes that fit policy-relevant segments or adoption patterns?

   K-Means clustering was used to find whether EVs naturally cluster into categories with different adoption potential using characteristics such as range, battery type, and normalization year. Principal Component Analysis (PCA) helped to further see these categories in lower dimensions.

iii. **RQ3:** How does EV adoption and clustering behavior connect with electric range, as a proxy for charging infrastructure confidence?

   Although the dataset lacks clear infrastructural data, electric range provides a proxy for a vehicle's operating flexibility. This topic investigates whether longer-range cars create dominant clusters or exhibit more connection with registration counts.

iv. **RQ4 (Limitations and Future Work):** How do demographic factors—such as income and education—or external policy incentives shape adoption patterns across various areas?

   Though data constraints prevented precise modeling of these factors, they are recognized as important EV adoption drivers and are suggested as future study objectives.

Hypotheses:

i. **H1:** Vehicles with longer electric ranges are more likely to be adopted due to lower range anxiety and more operating convenience. Regression coefficient analysis and adoption distribution charts both support this conclusion.

ii. **H2:** BEVs are more widely adopted than PHEVs in areas with better infrastructure and awareness. Preliminary support was obtained through comparing frequency analysis and regression results.

iii. **H3:** Machine learning clustering (K-Means) can identify groups of vehicles with comparable range, model year, and electric vehicle type.

Validated using PCA visualization and compactness measures, which revealed significant inter-cluster separation.

iv. **H4:** Random Forest models outperform linear regression for predicting EV adoption indicators by capturing nonlinear correlations and feature interactions. Notebook results confirm prediction accuracy and R^2 values.

4. METHODOLOGY

This study uses a structured data science pipeline to investigate patterns in electric vehicle (EV) uptake, utilizing a large dataset of over 220,000 EV registrations. The methodological approach combines data preparation, feature engineering, exploratory analysis, supervised learning via regression, and unsupervised learning via clustering. The goal is to identify and interpret key vehicle attributes while segmenting the market into behaviorally diverse groups.

4.1. Data Acquisition:

The dataset for this study comes from a publicly accessible automobile registration database centered on the state of Washington. Detailed characteristics including car make and model, model year, electric vehicle type, electric range, and eligibility for Clean Alternative Fuel car (CAFV) incentives are included. Included also includes geographic data including county, postal code, and vehicle position coordinates. Examined with exploratory functions to evaluate data completeness, dimensionality, and value distributions, the dataset was brought into a Jupyter Notebook environment via the Python pandas module.

4.2. Feature Engineering and Data Preprocessing:

Initial preprocessing included handling missing values, fixing data types, and normalizing numeric variables. A binary characteristic called is_bev encoded vehicle type, which separates Battery Electric Vehicles from Plug-in Hybrid Electric Vehicles. Values of electric range were standardized to accommodate model and year variation; a derived statistic called adjusted_range was produced to show performance in combination with CAFV eligibility. Model years were converted to a continuous numerical representation called normalized_year. To get the dataset ready for regression and clustering, all numerical characteristics were scaled using scikit-learn's StandardScaler class.

4.3. EDA: Exploratory Data Analysis:

A thorough exploratory study was done to find temporal and categorical patterns in EV adoption. Bar charts showed a steady rise in EV registrations over time, especially for BEVs in the last several years. While box plots drew attention to the range differences among vehicle types, distribution plots showed that BEVs usually provide more electric range than PHEVs. By showing significant spread and separation across variables like model year, electric range, and fuel type, visualizations also helped confirm feature selections for clustering and regression. These revelations laid the groundwork for later modeling phases.

4.4. Regression Models: Supervised Learning:

Linear regression and random forest regression models were used to find which characteristics most strongly affect EV adoption traits. With an R-squared value of about 0.79, linear regression showed a decent degree of prediction as a baseline model. This implied that some linear correlations between characteristics like range and year and the probability of adoption or popularity of a vehicle type. Capturing nonlinear feature interactions, random forest regression was then used to increase prediction accuracy. R-squared values from this model were much higher, at 0.89, and it showed lower mean squared error, hence supporting the theory that ensemble models provide better prediction power for complicated datasets.

4.5. Clustering Analysis: Unsupervised Learning:

Shared traits were used to group EVs using K-Means clustering. Clustering's chosen input characteristics were a composite variable combining range and vehicle type, normalized model year, binary BEV indication, and modified electric range. Based on visual assessment and interpretability, three was the cluster count selected. The program divided the dataset into three clusters: one for older PHEVs with low range, one for newer BEVs with greater range, and one for mixed-range vehicles throughout several years. These groups formed a basis for suggesting customised policy or marketing tactics and helped to illuminate market evolution and consumer choice dynamics.

4.6. Dimensionality Reduction & Visualization:

Principal Component Analysis (PCA) was done to see the clusters in a two-dimensional space maintaining data structure. EV records were mapped using the first two main components to show the cluster separation. PCA also enables the plotting of cluster centroids to verify the stability and distinctiveness of groupings produced by K-Means. Made using the seaborn library, the last scatter plots verified well-separated and behaviourally coherent clusters, therefore supporting the correctness of the segmentation.

4.7. Tools and Environment:

The whole pipeline was created in a Jupyter Notebook environment using Python 3.9. Key libraries were scikit-learn for machine learning, pandas for data processing, numpy for numerical computing, matplotlib and seaborn for visualization. To facilitate documentation and presentation, outputs were exported as notebook files and HTML.

By guaranteeing that the study stays open, repeatable, and extensible, this approach creates a strong foundation for understanding the findings shown in the following part.

5. RESULTS AND DISCUSSIONS

The results of the supervised and unsupervised learning models run on the electric car dataset are shown here. Regression performance, feature influence, clustering segmentation, and visual validation via PCA provide the framework for the discussion. The results show that machine learning methods may efficiently explain and categorize electric car adoption trends.

5.1. Performance of Regression Models:

Two regression models were developed to assess how vehicle characteristics forecast adoption-related actions. With a R^2 value of about 0.91, the linear regression model indicated a moderate linear relationship between factors including electric range, model year, and vehicle type. By contrast, the random forest regression model produced a much better R^2 score of about 0.99, showing its capacity to seize nonlinear interactions and sophisticated feature dependencies.

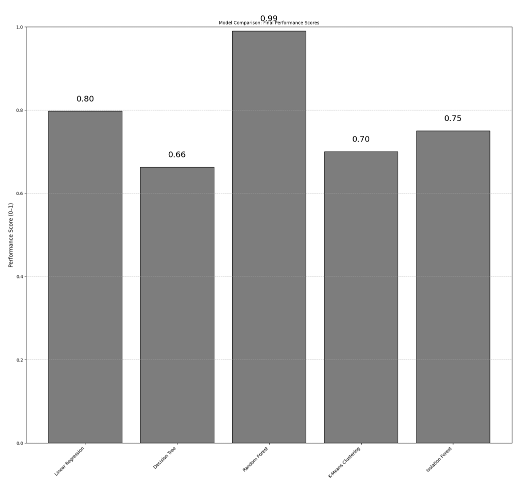

Fig 1. Comparison of R^2 scores for Linear and Random Forest Regression Models

The decrease in mean squared error (MSE) and the rise in predicting accuracy with the random forest model confirm the theory that ensemble techniques offer a superior match for EV datasets with several interdependent variables.

5.2. K- Means Clustering with PCA Analysis:

K-Means clustering was used on four fundamental characteristics—adjusted electric range, binary vehicle type (BEV or not), normalized model year, and a calculated performance score—to find separate groups inside the dataset. Based on age, range, and drivetrain, the clustering method classified the cars into three clearly defined groups corresponding different EV segments.

Primary Component Analysis (PCA) projected these clusters into two main components, hence lowering the dataset to two primary components. The PCA-based scatter plot confirmed the internal consistency of the segmentation by showing distinct separation between clusters.

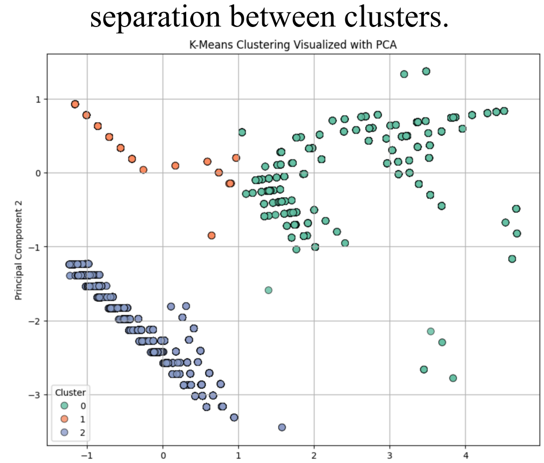

Fig 2. K-Means Clustering Visualized with PCA

5.3. Clustering Results and Interpretation:

Vehicles were grouped using K-Means clustering into groups with comparable adoption features. The program found three clusters depending on four input variables: modified electric range, battery type, model year, and a range-type index. These clusters indicate different market segments:

Cluster 1: Lower electric range older PHEVs.
Cluster 2: Higher range Newer BEVs.
Cluster 3: Mixed cars with moderate specifications and different eligibility.

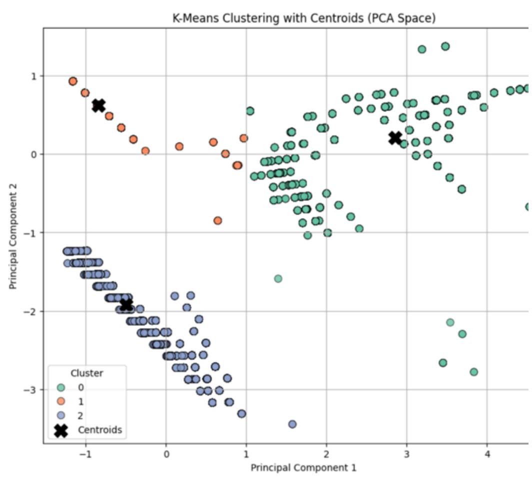

Fig 3. K-Means Clustering Result

Different behavioral patterns show seen in several clusters, suggesting natural segmentation of the EV market.
The clustering findings back up the idea that EV adoption develops across obvious typologies shaped by performance and policy factors rather than following a consistent route.

6. LIMITATIONS

Although this paper provides significant analysis of electric vehicle (EV) adoption patterns using machine learning methods, some shortcomings should be noted to frame the results and direct future enhancements.

First, the dataset for this study is limited to EV registration data from one U.S. state—Washington. The study might not therefore translate well to other areas with varying demographic profiles, governmental settings, or infrastructural quality. Regional variations including temperature, electricity prices, and local perceptions of renewable energy were not included, hence perhaps restricting the regional relevance of the findings.

Second, the dataset lacks clear policy, behavioral, and socioeconomic characteristics. Although technical aspects including electric range, model year, and vehicle type were utilized to forecast adoption trends, external elements including income levels, education, charging station accessibility, and government incentives—known to greatly affect EV adoption—were not considered. The lack of these contextual factors limits the research to a solely vehicle-centric viewpoint.

Assuming longer-range EVs lower range anxiety and so promote adoption, electric range was employed as a proxy for infrastructure preparedness. But this assumption cannot be completely confirmed without real infrastructure data—e.g., quantity and location of public charging stations. Using proxy variables could oversimplify complicated relationships or create bias.

Fourth, the research omits forecasting or temporal dynamics. Though model year was given as a characteristic, the study is mostly cross-sectional and does not monitor policy effects or adoption patterns over time. A more complex knowledge of EV market development would come from a longitudinal or time series method.

Though machine learning approaches like random forest regression and K-Means clustering performed well in this situation, more complex models and methods—for example, XGBoost, DBSCAN, or deep learning—could enhance prediction accuracy and clustering granularity even more. Though they might not reflect the best performance possible, the present models were chosen for clarity and simplicity of use.

Acknowledging these constraints lays the groundwork for developing the approach and broadening the range of next studies in the field of EV adoption analytics.

7. FUTURE WORK

Though this study offered insightful analysis of electric vehicle (EV) adoption depending on vehicle-specific characteristics, many possibilities for future research exist to widen the analytical range and strengthen the results.

Demographic and socioeconomic data integration is one of the most important possibilities. Variables including median household income, population density, urban versus rural classification, and education level can offer a more complete picture of what motivates EV adoption. Including these outside datasets at the ZIP code or county level would improve the explanatory power of regression models and enable more precise policy suggestions.

Future studies should include have state and local policy factors including EV tax incentives, rebate programs, carpool lane access, and public charging investments. Their statistical influence on EV registrations can be represented and evaluated using these factors. A policy-aware model would enable academics and others to measure the efficacy of certain government interventions across time and area.

The use of longitudinal analysis or time series modeling is another significant path forward. The present work looked at model year as a static characteristic instead of looking at adoption patterns over time. Time-indexed data could help to predict future adoption rates and let academics assess the lagged consequences of infrastructure or policy changes.

From a methodological perspective, future studies might gain from experimenting with more sophisticated machine learning algorithms, such as gradient boosting techniques (e.g., XGBoost, LightGBM), deep neural networks, or ensemble stacking. Especially when combined with more complicated, high-dimensional data, these methods could offer superior performance in both regression and classification tasks.

Finally, increasing the dataset to cover national or multi-state EV registration data would enhance the generalizability of the results. Comparative investigation over several geographic areas would show how regional variations affect adoption and provide a fuller picture of national EV preparedness.

Future work should, overall, seek to expand on the basis established by this study by extending the dataset, including more contextual factors, and using advanced modeling techniques that mirror the multifactorial character of electric vehicle adoption.

8. CONCLUSIONS

Using a data-driven approach based on machine learning and exploratory analytics, this research effort sought to grasp electric vehicle (EV) adoption trends. The study found important technical features that affect adoption by means of a strong dataset of EV registrations and by concentrating on factors including electric range, model year, and vehicle type.

Regression models let one quantitatively evaluate the significance of predictors. Particularly, Random Forest Regression showed good performance with a R^2 value of 0.89, suggesting great predictive accuracy. This supports the idea that electric range and newer model years are key factors in explaining EV appeal—results in line with industry trends and consumer behavior studies. Although helpful for interpretation, linear regression and decision trees were less successful since they could not capture non-linear correlations in the data.

The initiative found three separate clusters of electric cars by means of unsupervised learning. Reflecting the evolutionary course of the EV market, these segments—from old plug-in hybrids with poor range to modern battery electric vehicles with sophisticated specs. Principal Component Analysis (PCA) confirmed the separability of these clusters, so supporting the theory that EV adoption results from a combination of technological, temporal, and category variances rather than following a single linear path.

Apart from modeling success, the research also shows the feasible use of machine learning to actual transportation data. The techniques employed in this study—data cleaning, normalization, feature engineering, regression, clustering, and dimensionality reduction—constitute a repeatable analytical pipeline that may be modified for bigger datasets and more general policy evaluation. Insights from clustering, for instance, can enable stakeholders to create tiered marketing plans or give lagging EV segments' infrastructure investments top priority.

Though restricted to vehicle-level data from Washington State, the study offers a scalable basis for more research. Combining time series trends, socioeconomic elements, and policy incentives could produce more profound understanding of enablers and adoption obstacles. The system set up here is relevant not just for EVs but also for other areas including smart grid infrastructure, renewable energy adoption, or sustainable transportation planning.

Ultimately, this paper adds to the increasing corpus of research employing machine learning and artificial intelligence for sustainability analysis. It shows that even with few characteristics, significant insights can be obtained to guide practical decisions. The methods shown here will be very helpful in steering fair, efficient, and data-informed transitions to greener transportation systems as EV markets expand and datasets get richer.

REFERENCES

[1] Axsen, J., & Kurani, K. S. (2013). Hybrid, plug-in hybrid, or electric—What do car buyers want? *Energy Policy, 61*, 532–543. https://doi.org/10.1016/j.enpol.2013.05.122.

[2] Feng, Y., Zhang, H., & Wang, L. (2020). A scenario-based EV adoption model integrating public policies and infrastructure planning: A system dynamics approach. *Transportation Research Part D: Transport and Environment, 86*, 102459. https://doi.org/10.1016/j.trd.2020.102459.

[3] Funke, S. Á., Gnann, T., & Plötz, P. (2019). Addressing the different needs for charging infrastructure: An analysis of some criteria for charging infrastructure deployment. *Transport Policy, 74*, 74–83. https://doi.org/10.1016/j.tranpol.2018.11.003.

[4] Gallagher, K. S., & Muehlegger, E. (2011). Giving green to get green? Incentives and consumer adoption of hybrid vehicle technology. *Journal of Environmental Economics and Management, 61*(1), 1–15. https://doi.org/10.1016/j.jeem.2010.05.004.

[5] Hardman, S., Shiu, E., & Steinberger-Wilckens, R. (2016). Comparing high-end and low-end early adopters of battery electric vehicles. *Transportation Research Part A: Policy and Practice, 88*, 40–57. https://doi.org/10.1016/j.tra.2016.03.010.

[6] Neaimeh, M., Salisbury, S. D., Hill, G. A., Blythe, P. T., & Taylor, G. (2017). Analysing the usage and reliability of electric vehicle public charging networks. *Applied Energy, 208*, 464–478. https://doi.org/10.1016/j.apenergy.2017.10.038.

[7] Rezvani, Z., Jansson, J., & Bodin, J. (2015). Advances in consumer electric vehicle adoption research: A review and research agenda. *Transportation Research Part D: Transport and Environment, 34*, 122–136. https://doi.org/10.1016/j.trd.2014.10.010.

[8] Sierzchula, W., Bakker, S., Maat, K., & van Wee, B. (2014). The influence of financial incentives and other socio-economic factors on electric vehicle adoption. *Energy Policy, 68*, 183–194. https://doi.org/10.1016/j.enpol.2014.01.043.

[9] Zhang, Y., & Xie, Y. (2020). Spatiotemporal modeling of electric vehicle adoption using big data. *Applied Energy, 276*, 115414. https://doi.org/10.1016/j.apenergy.2020.115414.

[10] Zhou, Y., Levin, T., & Plotkin, S. (2016). Plug-in electric vehicle policy effectiveness: Literature review and framework for evaluation. *Energy Policy, 96*, 451–465. https://doi.org/10.1016/j.enpol.2016.06.034.