

# Analysis and Prediction of Electric Vehicle Ownership in Washington State

Stephen Dong

*Department of Electrical and Computer Engineering, University of California San Diego  
stdong@ucsd.edu (A69026970)*

## ABSTRACT

Over the last decade, electrical vehicles (EV) have rapidly gained popularity and demand in the United States as EV slowly become more efficient. This study explores the rising population of EVs on the roads to investigate trends and causes for this drastic shift in demand. Based on this analysis, a prediction of the future EV population will be made. The datasets utilized are limited to the availability of the data openly provided by the government. For this purpose, data exploration will be focused on the Washington (WA) state EV population datasets made publicly available by the WA State Department of Licensing (DOL) <sup>[1][2]</sup>.

This study will incorporate geographical and time series analysis of the EV population from 2017 to 2023. Based on the trend observed in the time series analysis, an appropriate regression model is trained to predict the amount of EVs that are likely to be registered between 2024 and 2026. The fitted forecast model will be evaluated for its accuracy using its Mean Absolute Percentage (MAPE) <sup>[3][4]</sup>. Inferences will be made on the outlook of EVs in the coming years using the final model. These observations can be useful for automakers in predicting future sales and making the appropriate decisions based on it.

The source code is available in the references.

## INTRODUCTION

For decades, developing fully electrical vehicles has been a daunting, yet highly desired goal by automakers. Beginning with its conceptions and initial renditions of steam powered EVs in the early 1900s, these vehicles were the center of attention until their decline after oil became a more efficient fuel source. The desire for EVs effectively perished until around the 1970s when the rise in gas prices urged automakers to seek alternatively fueled

vehicles. In tandem with the success of the first EV being the fully electric lunar-vehicle in 1971, the appeal for EVs was once again on the rise <sup>[5]</sup>.

Despite slow, but consistent research being put into EVs, it wasn't until the 2000s when efficient hybrid vehicles first hit the market. Immediately following this milestone, the first plug-in hybrid vehicles (PHEVs) and eventually complete EVs became commercially available. Despite this, PHEVs and EVs were still undesired compared to gas cars due to their relatively inefficient batteries and high prices. This is due to the cost of batteries alone, which likely also discouraged automakers from developing more efficient batteries. However, this quickly changed around 2014 when battery costs dropped, making EVs more affordable and desirable to the public <sup>[5]</sup>.

Fast-forward to 2023, EVs are now the center of attention, especially in the U.S. With the provision of Artificial Intelligence (AI) and federal rebate incentives, it is no wonder they are becoming more popular to this day <sup>[6]</sup>. Consequently, batteries have become drastically more efficient with a maximum of 80 miles in 2011 to 265 miles in 2015 and 520 miles in 2022 <sup>[7]</sup>.

This study seeks to explore how the popularity of EVs has changed in WA since 2017. With constant improvements in efficiency and declining prices, demand for EVs is likely to continue rising in the upcoming years. Analyzing this trend allows us to predict how many EVs we are likely to see hitting the road in the next few years. Simultaneously, geographical observations can help us pinpoint where we are likely to see the bulk of EV purchases. From this, we can also investigate the distribution of EV brands in WA.

For this study, hybrids, PHEVs, and EVs will all be referred to as EVs.

## THE DATASETS

For the last few years, WA's DOL has been actively collecting data on the vehicles that have been registered with the state. These datasets have been made publicly available on Data.gov <sup>[1][2]</sup> and contain information on all electrical vehicles currently registered with the state. Both datasets are assumed to be the population data for WA and are updated periodically throughout the year.

The first dataset, denoted "vehicle\_history", contains monthly vehicle registration counts the DOL has collected at the end of each month from January 2017 to November 2023. Each month contains basic information on newly registered EV, such as the type of EV, the battery and if it is a passenger vehicle. However, for this study, only the date and vehicle registration counts are relevant. There are a total of 19700 rows. Each row was either separated for its month or for its vehicle county-state information <sup>[1]</sup>.

The second dataset, denoted "vehicle\_data", contains all 163005 EV vehicle information currently registered as of 2023 by the DOL. The dataset contains each vehicle's 1-10 VIN, county, vehicle brand, and geographical coordinate. This dataset will only use the geographical data of the county, coordinates, and brand <sup>[2]</sup>.

## DATA PREPERATION

### Pre-Processing

Both datasets must be cleaned before performing any analysis. Rows of both datasets containing any missing values will be removed. For vehicle\_history, missing data entails either the date, county or EV count was missing. Unfortunately, we do not have any features that could confidently fill the NaN values. Similarly, vehicle\_data would entail missing geographical or automaker information. Although the make, model, and model year columns can be used to predict each other, observation of the data reveals that these features are either completely present or completely missing, making estimative filling improbable.

The dates for vehicle\_history were converted into pd.datetime objects for easier extraction of the parts of the date. Doing this allows us to quickly transform the original categorical type date values to numerical date values that will assist with our predictive model.

The coordinates of vehicle\_data were originally stored as Point objects, originally being used via shapely module. These values were transformed to list formatted tuples of [x, y] for mapping usage by the folium package.

### Geographical Preparation

Two supplementary GeoJSON datafiles <sup>[8]</sup> were included to provide the coordinates that will be used to map out WA and its counties. The states coordinates are separated from county coordinates. Each of the datafiles contain 500K coordinates of each state in the U.S. and the counties for each of the states separately. The county coordinates must be extracted for our geographical analysis; however, the datafile categorizes them using the state ID. Extracting the state ID from the states datafile provides us with id=53, which is used to extract the dictionary of coordinates for each county in WA. With all state and county information, the coordinate outlines for WA and its counties were extracted from the JSON file.

## ANALYSIS

### Understanding Vehicle History

The vehicle history data contains a count of the total number of EVs and the total number of vehicles registered each month as described in the data information provided on Data.gov <sup>[1]</sup>.

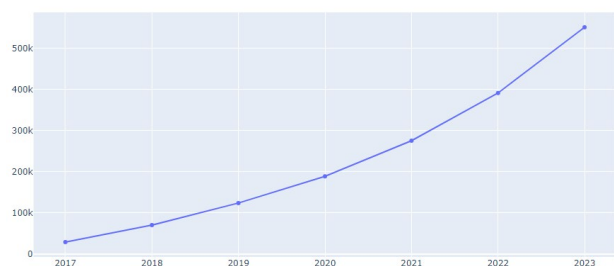


Figure 1: Cumulative # of EVs Registered with WA Up to 2023



Figure 2: # of EVs Actively Registered as of Each Month Between 2017 to 2023

Assuming this is true, we can compute the total count of new EVs for a given year. Calculating the cumulative sum of each year produces the trend in Figure 1. This trend matches the expected observed rise in EV popularity throughout the last few years. However, this implies that about 275 thousand of the approximately 7 million residents of WA (3.9% of population) own an EV in 2021.

As exciting as this sounds, 275 thousand is about 16.2% of the total EVs in operation in the U.S. in 2021, with California, Texas, and Florida taking the lead [9]. However, California alone holds 76.5% of all the EVs in the U.S. If Texas and Florida are above WA, our computed sum becomes illogical.

This insight implies that the provided description was incorrect. Knowing this, we can adopt another description that instead `vehicle_history` counts the total number of EVs that were *actively* registered with the DOL at a given month, i.e. each month contains the *net cumulative sum* of all EVs registered in the WA.

### Vehicle History Analysis

Figure 2 provides the cumulative sum of EVs registered with the DOL as of a given month. The trend confirms our initial understanding of the rise of EVs. We can also observe an exponential increase in EV ownership with each progressing month. In 2021, WA owns is approximately 86

thousand EVs, or about 1.2% of its total population and 5.1% of U.S.'s total EVs in operation. It is immediately more consistent with the data of active EVs in the U.S. mentioned earlier. Unfortunately, it is also unveiled that December 2023 currently has no data (as of December 18, 2023).

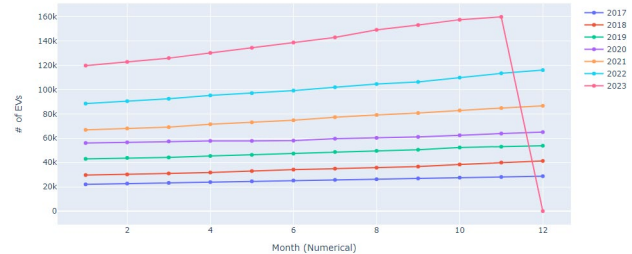


Figure 3: # of EVs Registered Each Month (Separated)

Figure 3 provides a side-by-side view of each month throughout the years from which we can observe a steady trend for each month from 2017 to 2022. The slope between each point on a given line provides the net change in EVs that were actively registered between each month. With each year plotted side by side, a consistent rate of change can also be observed between two months across each year. By computing the mean change of each year, we can obtain an approximate estimation of the number of EVs for December 2023.

The missing information can also be obtained by utilizing a regression model; however, a decision

was made to not use this method. Each month may have different factors, such as holidays, annual refreshes, and others, contributing to their change. Assuming a common trend for the entire year is likely to produce an imprecise estimate. It is more probable to observe the same effects of a month across the years, allowing the consequent change to be a more precise feature to estimate with.

The following equation can be used to produce a mean estimator for December:

$$\bar{X} = \frac{\sum \mu_i}{n}$$

This provides us with an average rate of increase of 2.1%, or the equivalent of a net increase 3383 EVs from November 2023. Using this estimation, an approximate total of 163222 actively registered EVs in December 2023 is obtained.

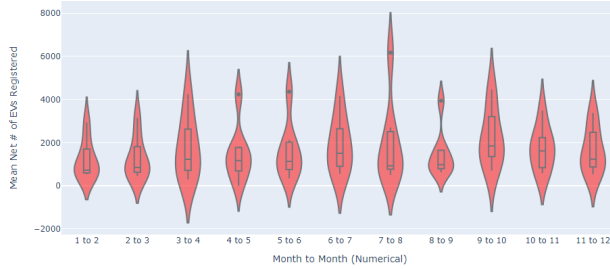


Figure 4: Distribution of the Mean Change Between Each Month (across 2017 - 2023)

The resulting distribution of net change of EVs between two months is observed in Figure 4 using violin plots. Most change rates fall within the same range between 400 and 3100 EVs. Likewise, there is a relatively large sum of outliers that also appear to affect their respective distributions. Despite this, each month sees a relatively normally distributed gain in EVs throughout 2017 to 2023.

Some months also exhibit a positively skewed distribution. This along with the outlier would have affected the prediction for December's gain value in a regression model. Although the outliers could be removed to help improve the precision skewedness, there are not enough years in the dataset to permit sacrificing an entire month's worth of data <sup>[10]</sup>.

Fortunately, we can focus on only December 2023 since it is portrayed to have a more centered normal distribution. This supports the earlier decision to utilize the mean rather than a predictor to estimate the value for December 2023. In fact, the reliability of this estimate is reinforced by vehicle\_data, which suggests that there are 163005 EVs registered (before pre-processing), as of December 16, 2023.



Figure 5: # of EVs Actively Registered as of Each Month Between 2017 to 2023 (Fixed)

## Pre-Selecting a Prediction Model

The new cumulative number of actively registered EVs is shown in Figure 5, which contains the estimated count for December 2023. The trend shown suggests a regression model would be beneficial in predicting the larger future of EVs, at least for the next 3 years.

We had originally concluded that a regression model was imprecise for predicting an *individual month*. However, in the perspective of predicting many years in the future, it's better to look at the overall trend to provide a rough estimate.

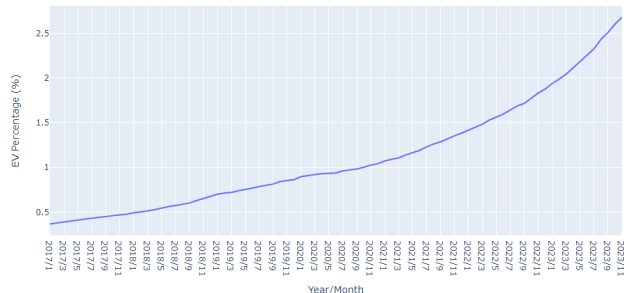


Figure 6: Rate of EV Increase Each Month (2017-2023)

Both Figure 6 and Figure 3, through its growing gap between the years, suggests that the change rate is growing exponentially. Similarly, Figure 4's outliers

and the inclusion of the December estimate are beginning to skew the average change rate distribution. Eventually the change rate distribution will become heavily skewed, such that simply using the normal distribution and mean to estimate will become less accurate than a regression model.

## Vehicle Ownership Geospatial Analysis

The vehicle\_data dataset in tandem with the GeoJSON coordinates can provide an insight into the geographical distribution of EV ownership across WA. Plotting the count of registered EVs in each individual county produces the choropleth in Figure 6.

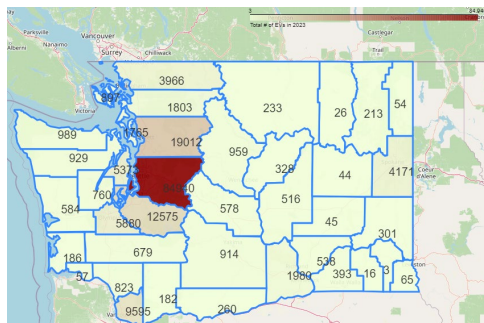


Figure 7: Choropleth of EV Ownership in WA

The majority of WA is demonstrated to have relatively little, but consistent numbers of EV ownership. Five counties are shown to have a higher EV ownership than every other county. However, King County (crimson red) is portrayed to own more EVs (about 52% of WA's total) than all other counties combined.

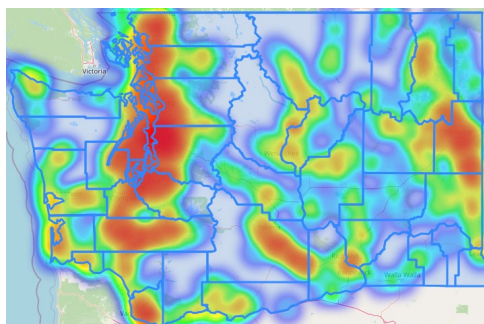


Figure 8: Heatmap of EV Ownership in WA

Figure 7 plots each EV registration's exact coordinate locations throughout WA. This provides a deeper understanding of where the bulk of EV

ownership is located. Counties with higher EV ownerships tend to reside in densely populated cities; however, this aligns with expectations.

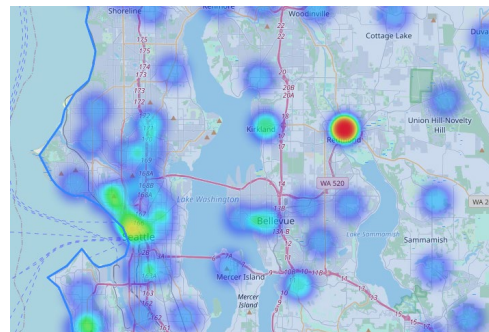


Figure 9: Heatmap of Part of King County

Focusing on the individual hotspots, we can observe the cities and the concentration of EVs they own. Figure 9 zooms in on the denser regions of King County. Redmond is seen to have the majority of the EVs in King County despite Seattle being the most likely assumption.

One likely cause for this unforeseen outcome may be their overall wealth. The cost of living in Redmond is 13.2% more than Seattle <sup>[11]</sup>. Nonetheless, the difference is still relatively low given how many more EVs reside in Redmond. Although, this suggests that Redmond may have a higher level of disposable income.

Another cause may be their crime rate. Seattle has a 76.9% property crime rate compared to Redmond's 44.2% property crime rate <sup>[12]</sup>. Correlating this with our data may suggest that higher levels of property crime may deter EV purchases. Nevertheless, these assumptions cannot be made with 100% certainty without socioeconomic analysis, which is not explored in this paper.

## EV Brand Distribution Geospatial Analysis

With growing demand for EVs, more automakers are jumping on the bandwagon to sell their own EVs. Figure 10 portrays the distribution of EV brands across WA using vehicle\_data. Tesla is the top brand for all counties and holds the majority of WA's EV market with 73411 EVs, or 45% of the total. This is followed by Nissan with only 13786 EVs.





Figure 10: Distribution of EV Brands in WA

Tesla being the highest seller of EVs is not particularly surprising. Their appeal can be traced to its main selling point—autopilot. Tesla is well known for its integration of AI and reliability for being in the market the longest. This makes them a highly trusted and enticing company to purchase EVs from. The drastic disparity between first and second indicates that Tesla may remain the dominant EV brand in WA for at least the next few years. However, given that other manufacturers have only recently begun stocking up on EVs [13], it is hard to say how the distribution may change in the far future.

## PREDICTION

One of the main purposes of this study is to analyze the EV trend in WA and predict the population of EVs in the coming years. A model was already selected when analyzing the annual count of active EVs in WA.

### Accuracy Evaluation: MAPE

Our model responsible for *forecasting* future EV purchases will be analyzed using its MAPE against its own training set. The metric allows us to compute the error percentage between our sample and prediction values [4].

$$MAPE = \frac{1}{n} \sum_{t=1}^n \frac{|A_t - F_t|}{A_t}$$

Like the Mean Percentage Error, the calculation for MAPE is dependent on the sample size and average deviation from its actual value. MAPE was chosen over the Mean Squared Error due to the wavering

trend we analyzed earlier. MAPE is less sensitive to these relative outlier behaviors and will provide a more accurate evaluation of our time-series predictor, especially with our large dataset [3].

### Exponential Regression

Upon initial inspection of the earlier visualized trend (Figure 5), the data appears to be exponential. An exponential regression was fitted using scikit-learn's `curve_fit` function.

$$y = a \cdot e^{b \cdot x} + c$$

This model generated coefficients a: -9.685e-14, b: 1, c: -17287208152.249, with a MAPE score of 1.2589e+18%. This is extraordinarily high especially considering this score is compared against itself. However, as it was mentioned earlier, the trend for this dataset occasionally wavers, which may confuse and underfit our basic exponential model.

### Linear Regression

Looking at the trend again, it could be argued that it is linear. Testing this, our model was adjusted to a linear regression with NumPy's `polyfit` and `polyval` functions with order 1.

$$y = a \cdot x + b$$

This model generated coefficients a: 1519.821 and b: 6475.202, with a MAPE score of 14.7834%. This is already a major improvement from the exponential regression; however, the accuracy is still relatively low.

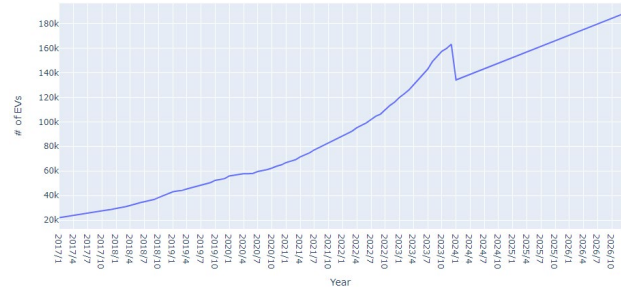


Figure 11: EVs Registered with Predictions up to 2026 (LR)

Figure 11 plots a prediction of how many EVs we are likely to see in the next three years to 2026 using our linear regression. The cause for 14.7834%

error rate immediately becomes clear as our predicted number of EVs drastically drops after 2023 to follow the linearity of the model. Thus, our model still requires some fine tuning.

### Piecewise Exponential-Linear Regression

Figure 11 demonstrates that the exponential trend we observed originally has an observable impact on the latter years. While it begins linearly, the trend eventually becomes exponential in June 2020. A piecewise regression will be more suitable here where a linear regression is initially used until June 2020 (41 months since January 2017), where it switches to an exponential regression <sup>[14]</sup>.

$$y = \alpha_0 + \beta_0 x + \alpha_1 \cdot e^{\beta_1 \cdot (x-41)} + \varepsilon$$

This model generates  $\alpha_0$ : 19320.685,  $\beta_0$ : 965.464,  $\alpha_1$ : 29574.878,  $\beta_1$ : 0.036,  $\varepsilon$ : 28783.733, with a MAPE score of 1.9618%. Our piecewise exponential-linear regression model improves our prediction accuracy to 98%. This indicates that the model is good at predicting the expected total EVs registered with the DOL in the next three years. Meanwhile, there is some error to argue against overfitting.

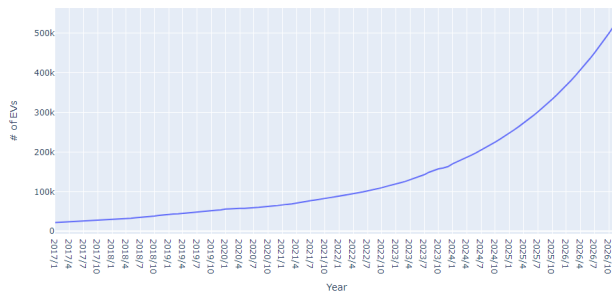


Figure 12: EVs Registered with Predictions up to 2026 (PWR)

Figure 12 illustrates our prediction model estimating the likely cumulative amount of actively registered EVs in WA. Unlike our linear regression model, it is not underfitted and provides a good continuation of the observed trend up to 2023.

### CONCLUSION

In the last few years, demand for EVs has been on a consistently exponential rise. Data pertaining to the trend of EV ownership was relevantly complicated to work with. This study was repeated three times

with different assumptions of the descriptions before a complete understanding of the data values was gained. Nonetheless, each variation produced the same result of a relatively exponential trend with some linearity at the beginning.

It is important to note that scientific breakthroughs and federal and economic have played a significant role dictating EV purchases. Predicting how many EVs we can expect cannot be constrained to the predictions provided by a singular model. As we had explored, combined regression formulas were key in defining an accurate model with limited annual data. Still, we cannot expect this model will remain accurate. With more data, additional models may be incorporated in the piecewise regression. Or the trend can be fitted to a single, most likely logistic, more representative regression model.

Nevertheless, using our predictions and analysis with the current data, we can conclude that WA can eventually see a dominance of EVs in the automobile industry in the next few years. The bulk of these EVs are likely to be Tesla's found in denser, safer, and wealthier cities. This observation can also be applied to other states or the entire U.S. in areas under similar conditions.

Overall, a boom in EV sales and ownership in the U.S. can be expected for the next few years. However, a key reminder that this trend will not stay consistent; external factors from the government, economy, and demographics can all play a part in accelerating or hindering the growth.

### REFERENCES

- [1] "Electric Vehicle Population Size History By County", 2023, <https://catalog.data.gov/dataset/electric-vehicle-population-size-history-by-county>
- [2] "Electric Vehicle Population Data", 2023, <https://catalog.data.gov/dataset/electric-vehicle-population-data>
- [3] "RdR score metric for evaluating time series forecasting models", 2022, <https://medium.com/@dave.cote.msc/rdr-score-metric-for-evaluating-time-series-forecasting-models-1c23f92f80e7>

- [4] "What Is MAPE? A Guide to Mean Absolute Percentage Error", 2023, <https://www.indeed.com/career-advice/career-development/what-is-mape#:~:text=Calculate%20the%20MAPE&text=Add%20all%20the%20absolute%20percent,final%20result%20is%20the%20MAPE>
- [5] "Timeline: History of the Electric Car", <https://www.energy.gov/timeline-history-electric-car>
- [6] "Electric Vehicle (EV) and Fuel Cell Electric Vehicle (FCEV) Tax Credit", 2023, <https://afdc.energy.gov/laws/409>
- [7] "The end of range anxiety: how has the range of electric cars changed over time?", 2023, [https://www.sustainabilitybynumbers.com/p/electric-car-range?utm\\_campaign=post&utm\\_medium=web](https://www.sustainabilitybynumbers.com/p/electric-car-range?utm_campaign=post&utm_medium=web)
- [8] "GeoJSON and KML Data for the United States", 2022, <https://eric.clst.org/tech/usgeojson/>
- [9] "How many electric cars are in the US?", 2023, <https://www.hertz.com/us/en/blog/electric-vehicles/how-many-electric-cars-are-in-the-us>
- [10] "Avoid Mistakes in Machine Learning Models with Skewed Count Data", 2021, <https://towardsdatascience.com/avoid-mistakes-in-machine-learning-models-with-skewed-count-data-e3512b94d745>
- [11] "Cost of Living Comparison Calculator Redmond, Washington vs Seattle, Washington", <https://www.bestplaces.net/cost-of-living/redmond-wa/seattle-wa/349971>
- [12] "2023 Compare Crime Rates: Seattle, WA vs Redmond, WA", <https://www.bestplaces.net/crime/?city1=55363000&city2=55357535>
- [13] "Automakers Are Adding Electric Vehicles to Their Lineups. Here's What's Coming", 2023, <https://www.consumerreports.org/cars/hybrids-evs/why-electric-cars-may-soon-flood-the-us-market-a9006292675/>
- [14] "STAT 501: Regression Methods, Piecewise Linear Regression Models". Dr. Iain Pardoe. Pennsylvania State University, <https://online.stat.psu.edu/stat501/lesson/8/8.8>
- [15] "Create and visualize Choropleth map with Folium", 2020, <https://medium.com/analytics-vidhya/create-and-visualize-choropleth-map-with-folium-269d3fd12fa0>
- [16] "Modeling Functions", 2020, <https://towardsdatascience.com/modeling-functions-78704936477a#7b2b>
- [17] "225Project: Analysis and Prediction of Electric Vehicle Ownership in WA Source Code Repository", <https://github.com/SDRepos-StephenDong/225Project.git>