Regression Analysis on EV Charger Coverage

Predictive Modeling for Public EV Charger Sufficiency on the U.S. West Coast

Executive Summary

Jesus Torres-Carbajal

Western Governors University

August 2025

Executive Summary

This study explores the need for public electric vehicle (EV) charger coverage across ZIP codes on the U.S West Coast. Regression and statistical modeling were used to predict the number of public chargers needed per ZIP code based on demographic and geographic variables such as population, median household income, EV ownership, urban/rural classification, and housing characteristics. The research question was, "Does the distribution of publicly available EV chargers significantly deviate from the modeled estimates of public need across ZIP codes on the West Coast?"

To address the research question, a dataset was compiled from multiple sources: state EV registration data, the U.S Census Bureau, and the Department of Energy's Alternative Fuel Data Center to address the research question. The datasets were cleaned and processed using Python, with necessary pre-processing steps such as merging, handling missing values, and feature engineering to create variables such as Charger-to-EV ratio.

A Random Forest regression model estimated the "optimal" number of public chargers needed per ZIP code based on demographics. The predicted values were then statistically compared to charger counts using a one-sample Wilcoxon signed-rank test. The model then identified several ZIP codes as potentially underserved in charging infrastructure.

The finding can inform EV infrastructure planning, mainly in ensuring equitable access across various urban and rural regions on the West Coast. One limitation of this analysis is that the model depends on ZIP-level EV registration data as a proxy for charger demand, which does not include outside influential factors like commuters or tourism. Future work could incorporate charger usage data to more accurately estimate demand and identify high-utilization EV chargers. GPS-based traffic data could also identify high-traffic zones where additional chargers within each ZIP code can be placed to provide the most benefit.

Section A: Introduction

Research Question

Does the distribution of publicly available EV chargers across ZIP codes on the West Coast significantly deviate from the modeled estimates of public need?

Context and Justification

The contribution of this study to the field of Data analytics is creating a predictive model that estimates the optimal number of public electric vehicle charging stations (EVCSs) for ZIP codes across the West Coast. The number of EVs in the United States is projected to rise to 27 million by 2030, up from just 3 million in 2022 (PwC, n.d.), making assessing where current infrastructure meets growing demand critical. The analysis uses a Random Forest regression model to identify key predictors, such as EV ownership per ZIP code, population, median household income, homeownership and renter occupancy rates, urban/rural classification, and current charger counts, to estimate how many public chargers a ZIP code should reasonably support.

To define "optimal" coverage, the charger-to-EV ratio is calculated for each ZIP code, with the top 25% of ZIP codes serving as a benchmark. The model is trained on this group and used to predict expected charger counts for all other ZIP codes. A one-sample Wilcoxon signed-rank test assesses whether actual charger counts significantly deviate from the model’s estimates.

From a business perspective, this model can help EV infrastructure companies and investors identify underserved areas that represent opportunities for profitable charger deployment. These underserved ZIP codes are places where there may not be enough chargers to meet demand, making them ideal locations to expand infrastructure and to offer a good return on investment. Prior research has found that communities with fewer advantages have 64% fewer chargers per capita than other communities, despite having higher public charging needs due to limited home charging access (Yu et al., 2025). Yu et al. (2025) confirmed this difference using the Mann–Whitney U test, revealing a statistically significant charger availability gap across these community types. This study builds on those findings by offering a predictive tool to guide infrastructure planning and highlight the differences in charger availability across ZIP codes, particularly in areas with higher public charging needs and lower access.

Hypothesis

- **Null Hypothesis:** There is no statistically significant difference between the actual and predicted number of public EV chargers per ZIP code on the West Coast, where the mean difference is zero.
- **Alternative Hypothesis:** A statistically significant difference exists between the actual and predicted number of public EV chargers per ZIP code on the West Coast, where the mean difference is not zero.

The hypothesis tests whether the model's predictions of charger need align with the current distribution of chargers across these ZIP codes. Suppose the analysis fails to reject the null hypothesis. Failing this may suggest that charger placement does not significantly differ from what the model predicts as optimal based on demographic factors. However, this result does not confirm that the current infrastructure is sufficient. Further investigation would be needed to validate this. Further investigation would be needed to validate this. If the analysis rejects the null hypothesis, it would indicate a statistically significant gap between actual and predicted optimal charger counts, revealing areas where infrastructure needs are unmet. These insights help identify ZIP codes that could benefit from further expansion in EV charger placements.

Section B: Data Collection

Data Collection

The analysis uses multiple publicly available datasets grouped by ZIP code to analyze whether the availability of public electric vehicle charging stations (EVCSs) is sufficient to meet local demand. The data includes the number of public EV chargers, EV registrations, population, median household income, homeownership rate, renter occupancy rate, and urban/rural classification for ZIP codes across California, Oregon, and Washington.

Demographic, Housing, and Socioeconomic Data

Demographic data for population, income, and housing tenure were collected from the U.S. Census Bureau’s 2019–2023 American Community Survey (ACS) 5-Year Estimates (U.S. Census Bureau, 2024). The dataset provides reliable estimates for over 33,000 ZIP Codes and is commonly used for public planning. While the ACS offers a wide range of data, it is based on survey samples and includes a margin of error. Another limitation is that these estimates may not reflect the most recent changes in 2024 or 2025.

Urban and Rural Classification

ZIP codes for urban/rural areas were classified using the U.S. Department of Agriculture’s Rural-Urban Commuting Area (RUCA) approximation file, which classifies ZIP codes on a 1-10 scale based on population density and commuting patterns (U.S. Department of Agriculture Economic Research Service, 2020). The dataset includes over 41,000 ZIP codes nationwide. In this study, RUCA codes 1-3 were classified as urban, while codes 4-10 were classified as rural. These classifications are based on data from 2006–2010 and may not fully reflect recent demographic changes.

EV Registration Data

EV registration counts were gathered from each state's most recent published datasets. Washington’s EV registration data was obtained from the Washington State Department of

Licensing dataset available through Data.gov (Washington State Department of Licensing, 2025). Oregon's data was obtained from the Oregon Department of Energy's EV Dashboard (Oregon Department of Energy, 2025). California's dataset came from the California Energy Commission's ZEV and Infrastructure Statistics website (California Energy Commission, 2025). These datasets provided counts of battery electric vehicles and plug-in hybrid electric vehicles by ZIP code. Each state releases these datasets according to different timelines, so that slight inconsistencies may exist across these three states.

Public Charger Locations

The public charger dataset was collected from the U.S. Department of Energy's Alternative Fueling Station Locator (Alternative Fuels Data Center, 2025). The dataset includes the number of public charging ports by ZIP code and is updated daily for networked stations. While the dataset is actively maintained, some chargers, such as non-networked or newly installed stations, may not appear due to delays. Another limitation is that the dataset was selected to exclude private, workplace-only, and residential-only chargers, which may result in undercounting available charging options.

An advantage of using these datasets is that they come from official government sources, which makes the data more trustworthy and suitable for public infrastructure analysis. These sources are regularly updated and freely available for use, which also helps ensure reproducibility in future studies. A disadvantage is that each dataset had to be merged using ZIP Codes as the shared key, which assumes that all datasets report data based on ZIP codes similarly. This method would add some uncertainty, especially for ZIP codes with inconsistent data across these sources.

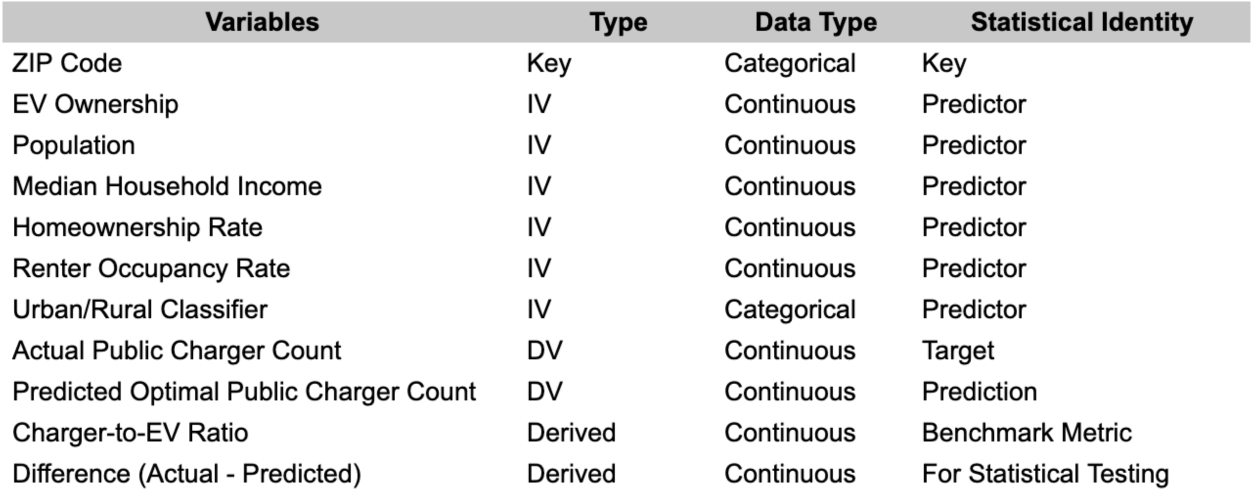
One challenge during the data collection process was that no single source provided all the data needed for this analysis. Each dataset only covered a part of the whole picture behind determining EV charger need, such as demographics, EV registrations, and available public chargers. Obtaining the necessary data required searching across multiple state and federal websites. Once trusted sources were found, each dataset was confirmed to be publicly available and up-to-date, then manually compiled into a single dataset, grouped by each Zip code.

Datasets Used:

- Washington EV Registration Data: Updated July 2025.
- Oregon EV Dashboard: Updated March 2025.
- California Vehicle Population: Updated April 2025.
- RUCA ZIP Code Dataset: Updated July 2025
- AFDC Station Locator Data: Updated daily via API.
- ACS 5-Year-Estimates: From 2019-2023.

Figure 1

Table of Variables Used



Variables	Type	Data Type	Statistical Identity
ZIP Code	Key	Categorical	Key
EV Ownership	IV	Continuous	Predictor
Population	IV	Continuous	Predictor
Median Household Income	IV	Continuous	Predictor
Homeownership Rate	IV	Continuous	Predictor
Renter Occupancy Rate	IV	Continuous	Predictor
Urban/Rural Classifier	IV	Categorical	Predictor
Actual Public Charger Count	DV	Continuous	Target
Predicted Optimal Public Charger Count	DV	Continuous	Prediction
Charger-to-EV Ratio	Derived	Continuous	Benchmark Metric
Difference (Actual - Predicted)	Derived	Continuous	For Statistical Testing

Section C: Data Extraction and Preparation

Data Extraction and Preparation

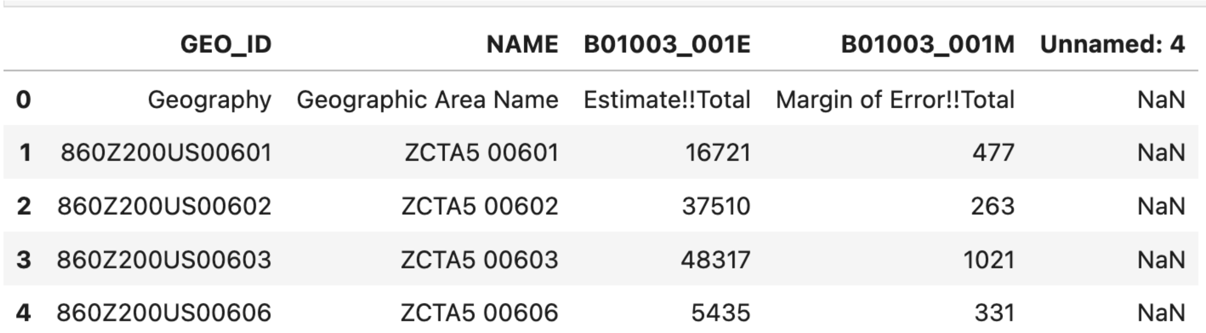
Python was used for data processing and preparation, and Pandas and Numpy libraries were used for cleaning and transformation (McKinney, 2010; Harris et al., 2020). Python is justified for this analysis due to its flexibility and the strength of its data science libraries. While R is known for its powerful statistical packages, recent studies comparing programming languages have found that Python produces results as accurate as R in everyday analytical tasks (Hill et al., 2024). Additionally, tools like Scikit-learn make Python especially equipped for building and testing machine learning models within a single environment.

ACS Datasets

Each ACS dataset was already aggregated at the ZIP Code level. The ZIP Code columns were standardized as 5-digit strings to ensure compatibility, relevant columns were selected, and datasets were to be later merged using ZIP Code as the shared key.

Figure 2

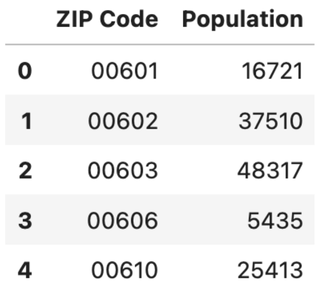
Example Raw ACS Dataset



	GEO_ID	NAME	B01003_001E	B01003_001M	Unnamed: 4
0	Geography	Geographic Area Name	Estimate!!Total	Margin of Error!!Total	NaN
1	860Z200US00601	ZCTA5 00601	16721	477	NaN
2	860Z200US00602	ZCTA5 00602	37510	263	NaN
3	860Z200US00603	ZCTA5 00603	48317	1021	NaN
4	860Z200US00606	ZCTA5 00606	5435	331	NaN

Figure 3

Example Raw ACS Dataset - Formatted



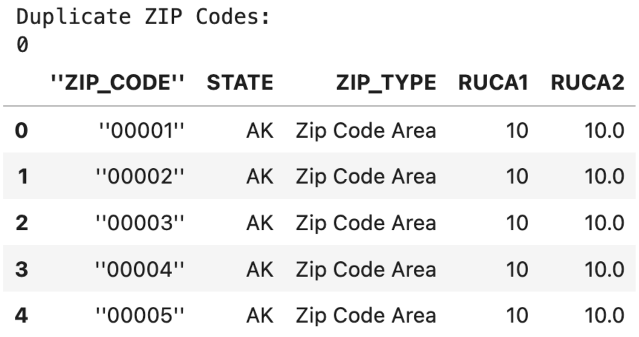
	ZIP Code	Population
0	00601	16721
1	00602	37510
2	00603	48317
3	00606	5435
4	00610	25413

RUCA Dataset

RUCA ZIP Code files were assigned a binary urban/rural classification for each ZIP code. Only ZIP codes from WA, OR, and CA were included, and the classification was defined using RUCA codes 1-3 as urban and 4-10 as rural.

Figure 4

Raw RUCA Dataset

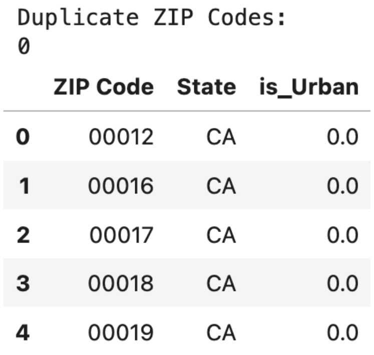


```
Duplicate ZIP Codes:
0
```

	''ZIP_CODE''	STATE	ZIP_TYPE	RUCA1	RUCA2
0	''00001''	AK	Zip Code Area	10	10.0
1	''00002''	AK	Zip Code Area	10	10.0
2	''00003''	AK	Zip Code Area	10	10.0
3	''00004''	AK	Zip Code Area	10	10.0
4	''00005''	AK	Zip Code Area	10	10.0

Figure 5

RUCA Dataset - Formatted



```
Duplicate ZIP Codes:
0
```

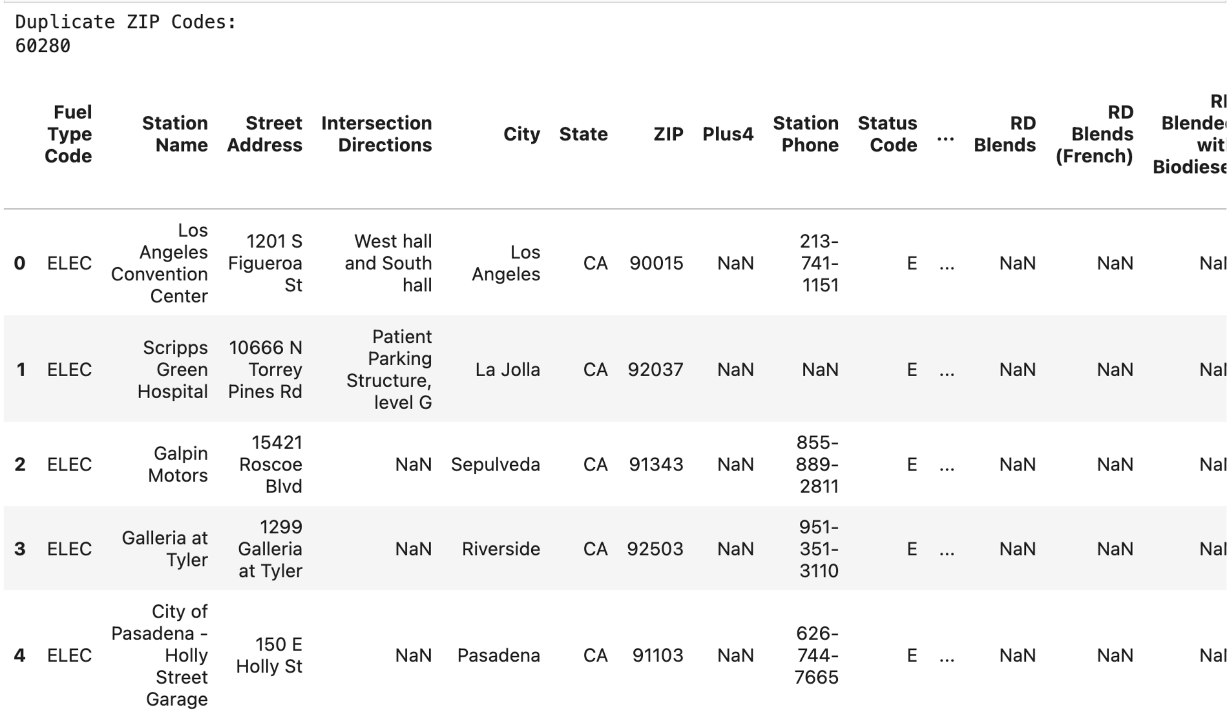
	ZIP Code	State	is_Urban
0	00012	CA	0.0
1	00016	CA	0.0
2	00017	CA	0.0
3	00018	CA	0.0
4	00019	CA	0.0

Public Charger Locations Dataset

Public charger data were filtered to include only available, public, non-workplace, Level 2, and DC Fast chargers. EV registration data were cleaned by dropping missing ZIP codes and aggregating vehicle counts by ZIP code.

Figure 6

Raw Public Charger Dataset

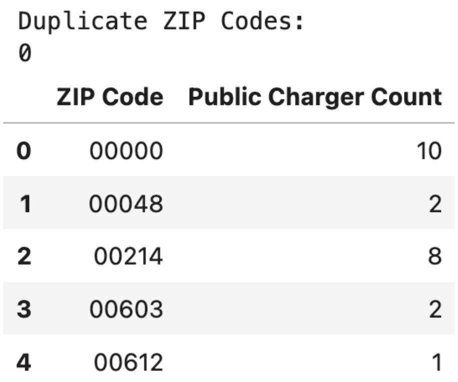


```
Duplicate ZIP Codes:
60280
```

	Fuel Type Code	Station Name	Street Address	Intersection Directions	City	State	ZIP	Plus4	Station Phone	Status Code	...	RD Blends	RD Blends (French)	RI Blende wit Biodiese
0	ELEC	Los Angeles Convention Center	1201 S Figueroa St	West hall and South hall	Los Angeles	CA	90015	NaN	213-741-1151	E	...	NaN	NaN	Nal
1	ELEC	Scripps Green Hospital	10666 N Torrey Pines Rd	Patient Parking Structure, level G	La Jolla	CA	92037	NaN	NaN	E	...	NaN	NaN	Nal
2	ELEC	Galpin Motors	15421 Roscoe Blvd	NaN	Sepulveda	CA	91343	NaN	855-889-2811	E	...	NaN	NaN	Nal
3	ELEC	Galleria at Tyler	1299 Galleria at Tyler	NaN	Riverside	CA	92503	NaN	951-351-3110	E	...	NaN	NaN	Nal
4	ELEC	City of Pasadena - Holly Street Garage	150 E Holly St	NaN	Pasadena	CA	91103	NaN	626-744-7665	E	...	NaN	NaN	Nal

Figure 7

Public Charger Dataset - Formatted



```
Duplicate ZIP Codes:
0
```

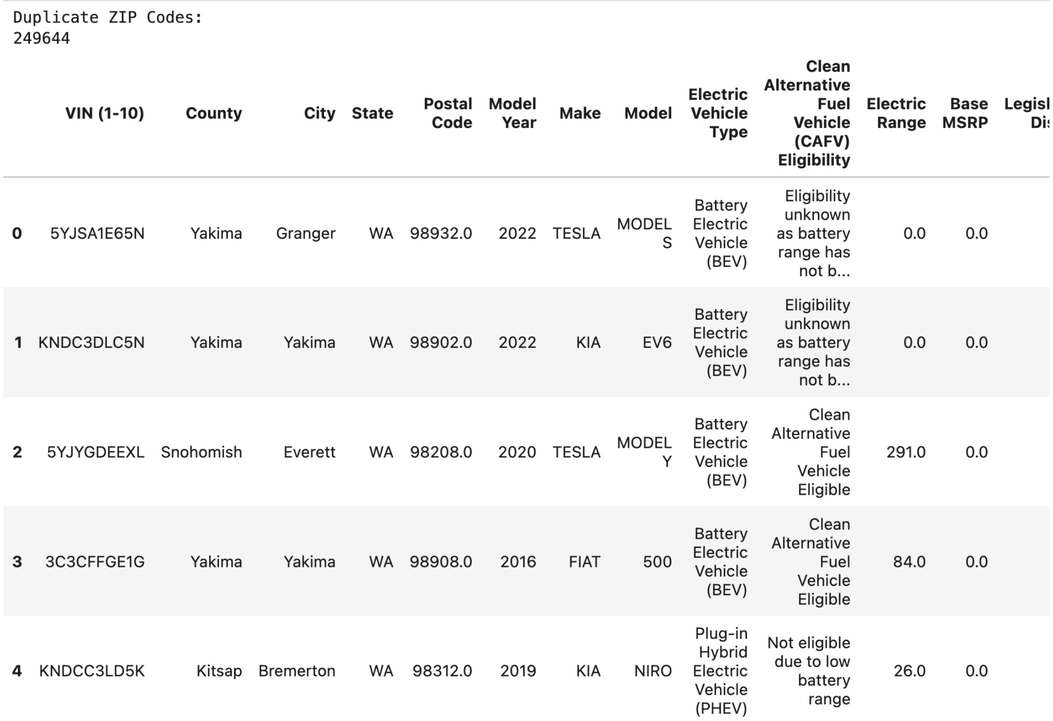
	ZIP Code	Public Charger Count
0	00000	10
1	00048	2
2	00214	8
3	00603	2
4	00612	1

EV Registrations Dataset

EV registration data, such as for Washington State, required additional preprocessing. Missing ZIP codes were dropped, ZIP codes were padded to 5 digits as a string, and vehicles were grouped by ZIP code by counting unique VINs.

Figure 8

Example Raw EV Registration Dataset

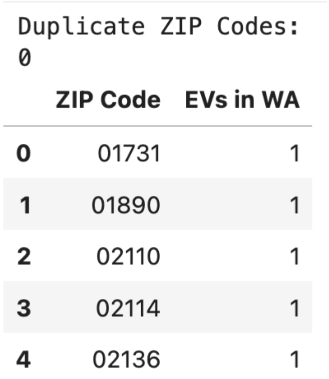


```
Duplicate ZIP Codes:
249644
```

	VIN (1-10)	County	City	State	Postal Code	Model Year	Make	Model	Electric Vehicle Type	Clean Alternative Fuel Vehicle (CAFV) Eligibility	Electric Range	Base MSRP	Legisl Di
0	5YJSA1E65N	Yakima	Granger	WA	98932.0	2022	TESLA	MODEL S	Battery Electric Vehicle (BEV)	Eligibility unknown as battery range has not b...	0.0	0.0	
1	KNDC3DLC5N	Yakima	Yakima	WA	98902.0	2022	KIA	EV6	Battery Electric Vehicle (BEV)	Eligibility unknown as battery range has not b...	0.0	0.0	
2	5YJYGDEEXL	Snohomish	Everett	WA	98208.0	2020	TESLA	MODEL Y	Battery Electric Vehicle (BEV)	Clean Alternative Fuel Vehicle Eligible	291.0	0.0	
3	3C3CFFGE1G	Yakima	Yakima	WA	98908.0	2016	FIAT	500	Battery Electric Vehicle (BEV)	Clean Alternative Fuel Vehicle Eligible	84.0	0.0	
4	KNDCC3LD5K	Kitsap	Bremerton	WA	98312.0	2019	KIA	NIRO	Plug-in Hybrid Electric Vehicle (PHEV)	Not eligible due to low battery range	26.0	0.0	

Figure 9

Example EV Registration Dataset- Formatted



```
Duplicate ZIP Codes:
0
```

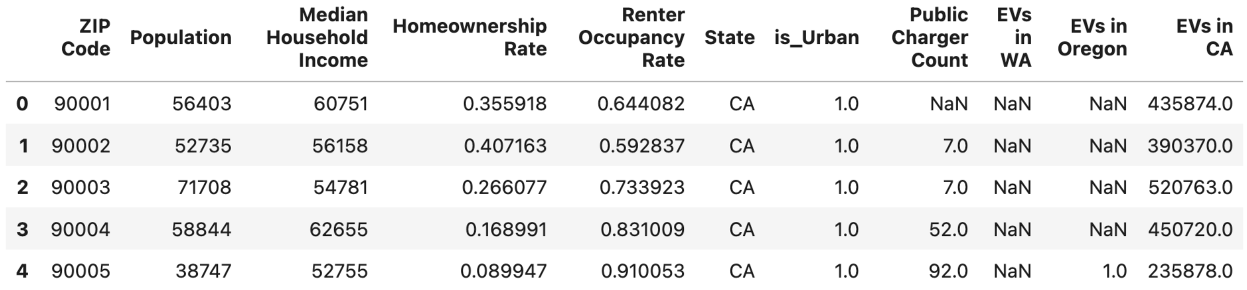
	ZIP Code	EVs in WA
0	01731	1
1	01890	1
2	02110	1
3	02114	1
4	02136	1

Merging

Once the datasets were formatted and contained standardized ZIP Codes across each dataset, they were merged based on each ZIP Code as the shared key. Zeros were filled in for missing values for EV registration counts for other states, then summed to create a new column for EV Ownership.

Figure 10

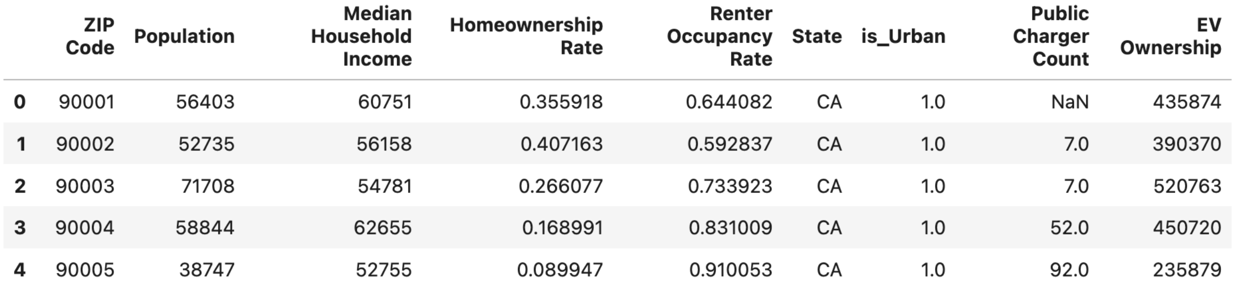
Merged Dataset



	ZIP Code	Population	Median Household Income	Homeownership Rate	Renter Occupancy Rate	State	is_Urban	Public Charger Count	EVs in WA	EVs in Oregon	EVs in CA
0	90001	56403	60751	0.355918	0.644082	CA	1.0	NaN	NaN	NaN	435874.0
1	90002	52735	56158	0.407163	0.592837	CA	1.0	7.0	NaN	NaN	390370.0
2	90003	71708	54781	0.266077	0.733923	CA	1.0	7.0	NaN	NaN	520763.0
3	90004	58844	62655	0.168991	0.831009	CA	1.0	52.0	NaN	NaN	450720.0
4	90005	38747	52755	0.089947	0.910053	CA	1.0	92.0	NaN	1.0	235878.0

Figure 11

Merged Dataset - Add EV Ownership



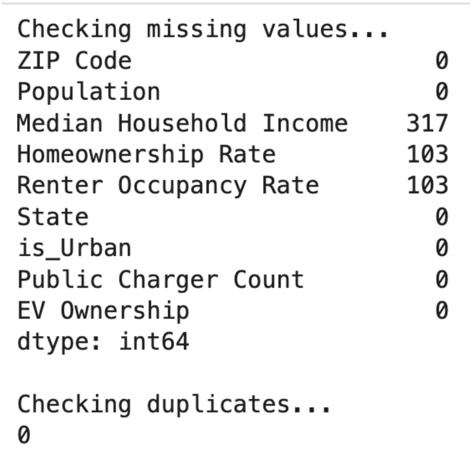
	ZIP Code	Population	Median Household Income	Homeownership Rate	Renter Occupancy Rate	State	is_Urban	Public Charger Count	EV Ownership
0	90001	56403	60751	0.355918	0.644082	CA	1.0	NaN	435874
1	90002	52735	56158	0.407163	0.592837	CA	1.0	7.0	390370
2	90003	71708	54781	0.266077	0.733923	CA	1.0	7.0	520763
3	90004	58844	62655	0.168991	0.831009	CA	1.0	52.0	450720
4	90005	38747	52755	0.089947	0.910053	CA	1.0	92.0	235879

Handling Missing Values

Missing values in the Public Charger Count column were imputed with zero to reflect ZIP codes lacking public chargers. ZIP codes without a rural or urban label were removed to keep the data consistent when comparing different types of areas. After that, any missing values in numeric columns were handled later within the model pipeline using median imputation after the train-test split. This technique helped avoid data leakage from the test set from influencing the model. Using the median to fill in missing numbers was justified to avoid outliers from extreme values.

Figure 12

Remaining Missing Values



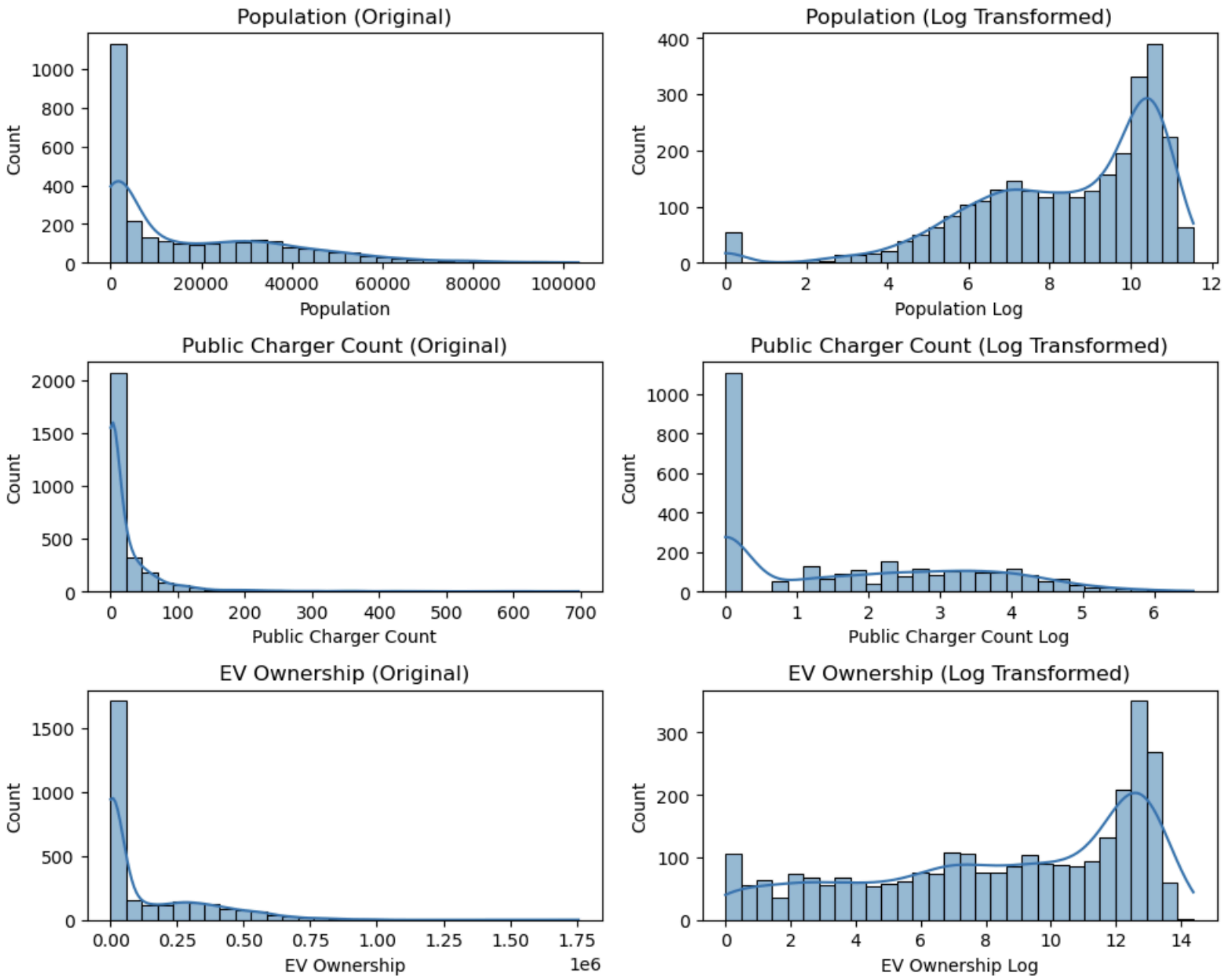
```
Checking missing values...
ZIP Code                    0
Population                  0
Median Household Income   317
Homeownership Rate        103
Renter Occupancy Rate     103
State                       0
is_Urban                    0
Public Charger Count        0
EV Ownership                0
dtype: int64

Checking duplicates...
0
```

Addressing Skewness

Log transformations were used on the Population, Public Charger Count, and EV Ownership columns to reduce skewness and improve model performance. As shown in the figure below, the original distributions were highly right-skewed, but the log-transformed versions were able to spread more evenly. This technique was justified to help make the patterns in the data easier for the Random Forest model to learn and to reduce the effect of extreme values.

Figure 13

Transformations



Applying the Threshold

To define what “optimal” charger coverage looks like, a Charger to EV Ratio was calculated by dividing the number of public chargers by the number of registered EVs per ZIP Code and scaling it per 1,000 EVs. ZIP codes in the top 25% of this ratio were selected as the benchmark group for modeling. Then, the most extreme 10% of this top group was removed to avoid letting unusually high outliers skew the model. This technique was justified as it ensured the model learned from areas with good charging access by keeping only the top 25% of ZIP codes, also removing extreme values so the model would not focus on the unusual cases.

Figure 14

Charger to EV Ratio Before Applying Threshold

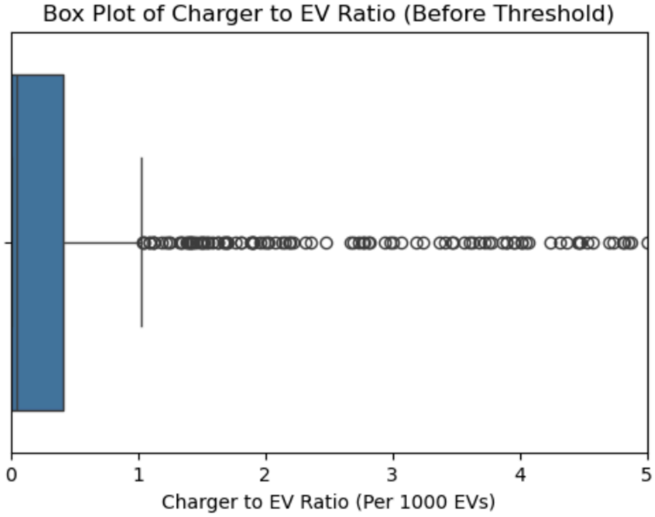
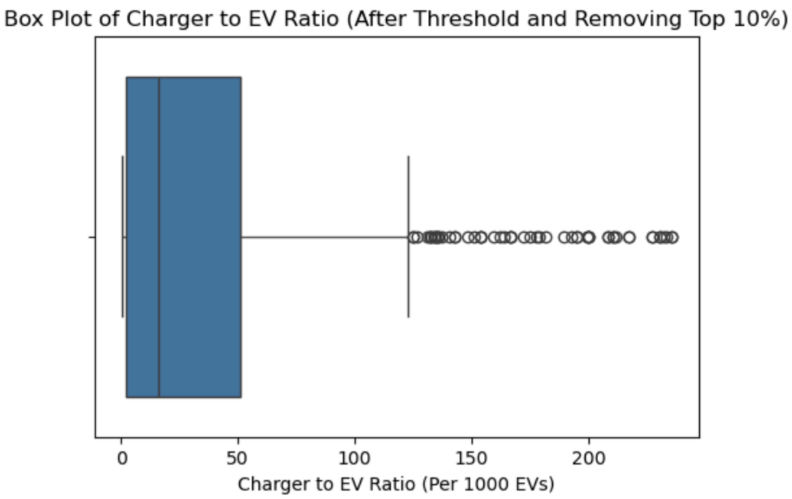


Figure 15

Charger to EV Ratio Before Applying Threshold

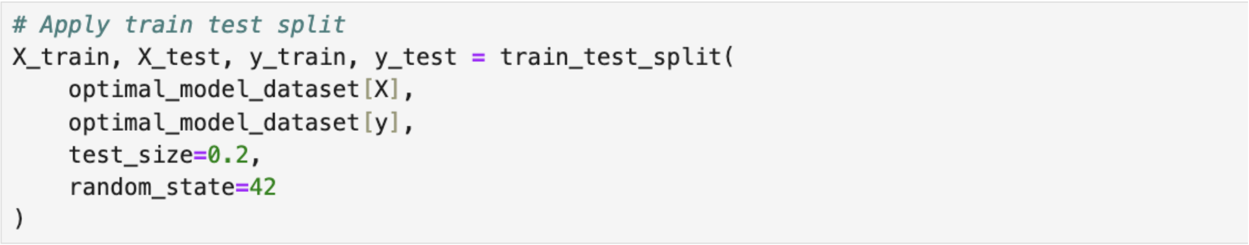


Applying Train Test Split

The dataset was split into training and testing sets using Scikit-learn to help ensure the model generalizes to unseen data (Pedregosa et al., 2011). A training and testing split was applied to the filtered dataset, using 80% of the data to train the model and 20% to test its performance. This technique was justified to help check whether the model can make accurate predictions on unseen data.

Figure 16

Apply Train Test Split



```
# Apply train test split
X_train, X_test, y_train, y_test = train_test_split(
    optimal_model_dataset[X],
    optimal_model_dataset[y],
    test_size=0.2,
    random_state=42
)
```

One advantage of this project's data preparation steps was performing a few essential cleaning steps before merging datasets. These steps included standardizing ZIP Code formats, filtering targeted chargers, and applying aggregation for EV registrations by ZIP code. Doing this early ensured a smoother process for establishing joins and consistent keys across all data sources. The technique reduced the likelihood of errors to help prevent missing values in the final dataset.

A disadvantage of the chosen preparation steps was using a threshold to define the top 25% of ZIP codes using the charger-to-EV ratio, then excluding those top 10% of outliers from this subset to train the model. While this helped focus the analysis on areas with strong charger coverage, it may have excluded some ZIP codes with valid, but unusual, charger counts. While these values may represent areas with high charger counts, such as towns with high tourism, they were extreme enough to distort model performance and were removed to keep the model reliable.

Section D: Analysis

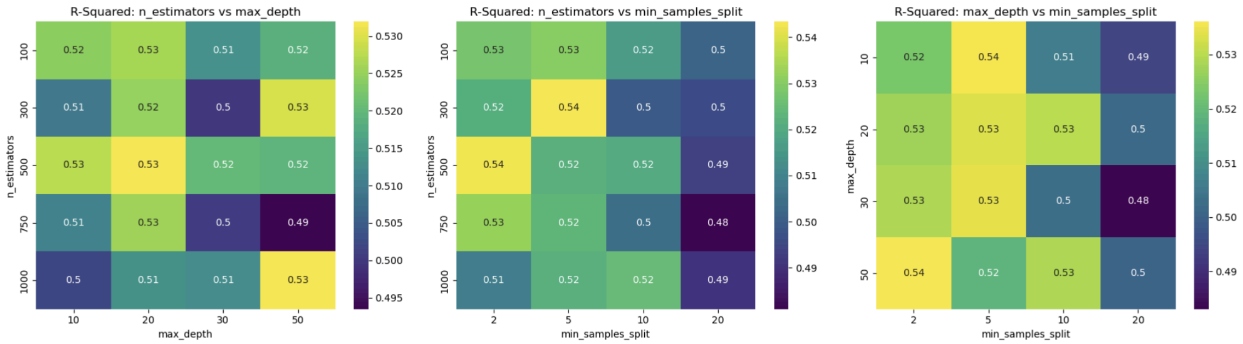
Analysis

A Random Forest regression model was used to estimate the optimal number of EVSCs per ZIP Code. A Random Forest model was used in this analysis as the model can capture complex, nonlinear relationships between predictors, and does not assume normality or linearity constraints. An advantage to using the Random Forest regression model was that it offered strong predictive power and handled complex nonlinear relationships between features and the target variable. However, a disadvantage is that Random Forest may be less interpretable than simpler models. Still, its ability to determine the importance of features provides valuable insights for stakeholders, offering a safer alternative to fully black-box models, which can be risky in high-cost decision-making (Rudin, 2019).

A preprocessing pipeline imputes the median for the remaining missing numeric values described in the previous section. This step was conducted in this order, after the train-test-split, to avoid data leakage, where the testing set includes the training process. The median was selected as the means for imputing due to the skewness of the data. Hyperparameter tuning was performed using Randomized Search, which used 500 iterations and a 5-fold cross-validation. The best model was selected based on the R-squared of the validation folds. Randomized Search was used to efficiently search through a wide range of hyperparameter combinations without conducting an exhaustive search, saving computation time compared to Grid Search. An advantage of this method is that it can identify strong model combinations quickly, without wasting time on weaker hyperparameters. One disadvantage is that it might miss the best combination in the search space as the technique relies on random sampling rather than a complete search.

Figure 17

Hyperparameter Tuning

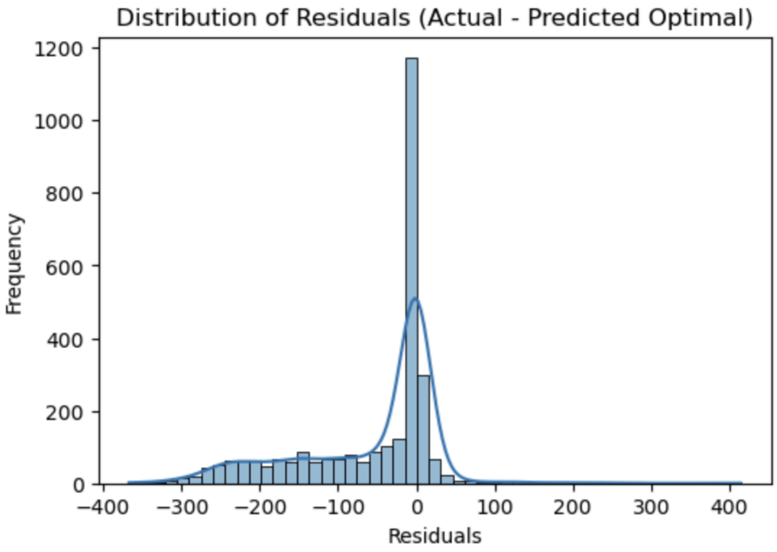


Model performance was evaluated using the R-squared and Root Mean Squared Error (RMSE). R-Squared measured how much variation in charger counts was explained by the input features, while RMSE quantified the average prediction error in the same units as the target variable. R-squared was used for this analysis as the objective was to understand how well ZIP code features explain differences in EV charger counts. R-squared became useful as it shows how much variation the model can explain. However, one downside is that it can overestimate accuracy for a complex or overfit model, so RMSE and cross-validation were also used to verify the results. Visualizations such as scatterplots and bar charts were created using Matplotlib and Seaborn to help interpret model results (Hunter, 2007; Waskom, 2021). The model was trained using the top 25% of ZIP codes with the highest charger-to-EV ratios, which served as the benchmark for "optimal" coverage. After training, the model predicted charger counts across all ZIP codes in the dataset.

A statistical test was conducted on the difference between the two values to evaluate whether the actual number of public chargers per ZIP code significantly differed from the predicted optimal amount. The test checked whether the average of this difference was statistically different from zero, suggesting that a need for additional EV chargers existed. The Shapiro-Wilk test was used to check if these residuals were normally distributed (Virtanen et al., 2020). Since the normality assumption was violated, the Wilcoxon signed-rank test, a non-parametric alternative, was performed. An advantage of the Wilcoxon test is that it does not assume a normal distribution, allowing the test to be used for skewed data. A disadvantage is that the Wilcoxon test is less likely to detect slight but fundamental differences than parametric tests when the data is normally distributed.

Figure 18

Statistical Testing Calculations



Model Performance

The optimized Random Forest model achieved an R-squared of 0.5906 on the validation set and 0.6722 on the test set, demonstrating that the model could explain approximately 67.2% of the variance in public charger counts. The RMSE on the test set was 54.50, representing the average prediction error in charger counts. These results suggest that the model provided reasonably strong predictive performance, especially given the complexity and skewness of the dataset.

Section E: Data Summary and Implications

Data Summary

The analysis developed a predictive model to estimate the optimal number of public EV chargers per ZIP code using regional and demographic variables. Figure 19 displays the distribution of actual charger counts against the predicted optimal counts from the model. The actual charger counts are highly right-skewed, with many ZIP Codes having little to no EV chargers. While in the predictions, the model presents a broader range of optimal charger counts, nearing halving ZIP codes with little to no chargers, and redistributes the amount of chargers, often higher than the current counts.

A Wilcoxon signed-rank test comparing actual and predicted charger counts yielded a test statistic of 278645.5 and a p-value of 9.53×10^{-215}, indicating an extremely significant difference between actual and predicted values. This analysis supports the idea that many ZIP codes across the West Coast may be underserved by the current EV charging infrastructure.

Figure 19

Actual vs. Predicted Charging Counts

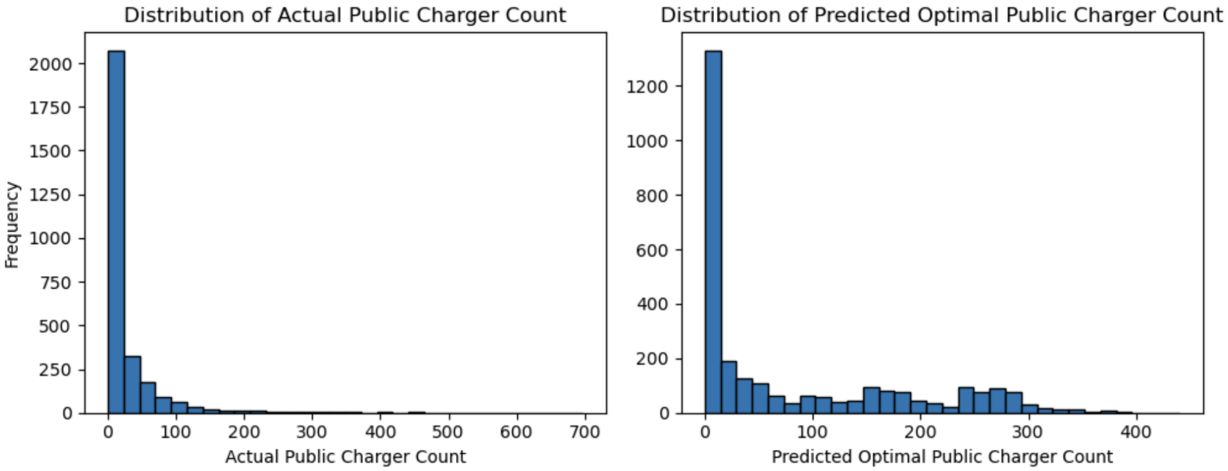
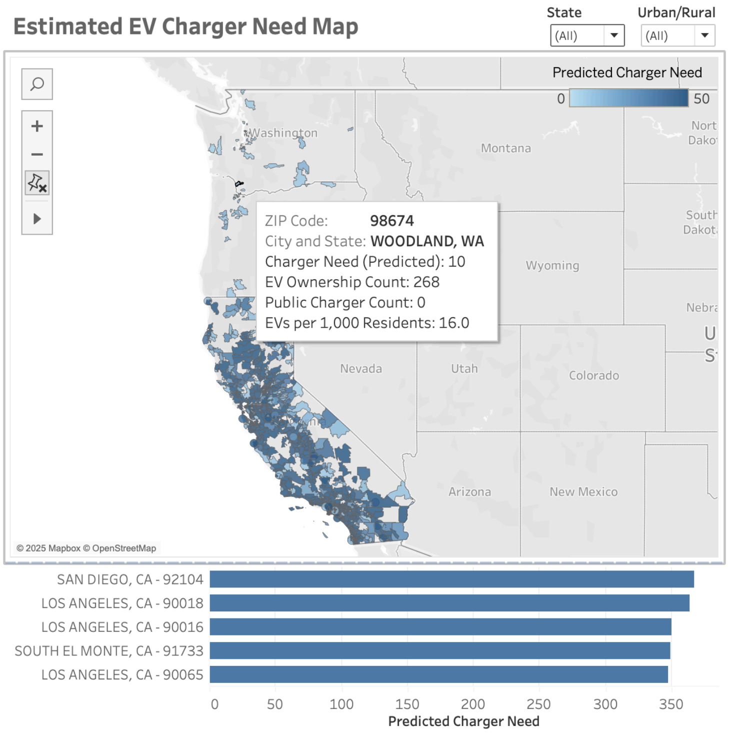


Figure 20

Map of Charging Need Across the West Coast


Estimated EV Charger Need Map

State (All)

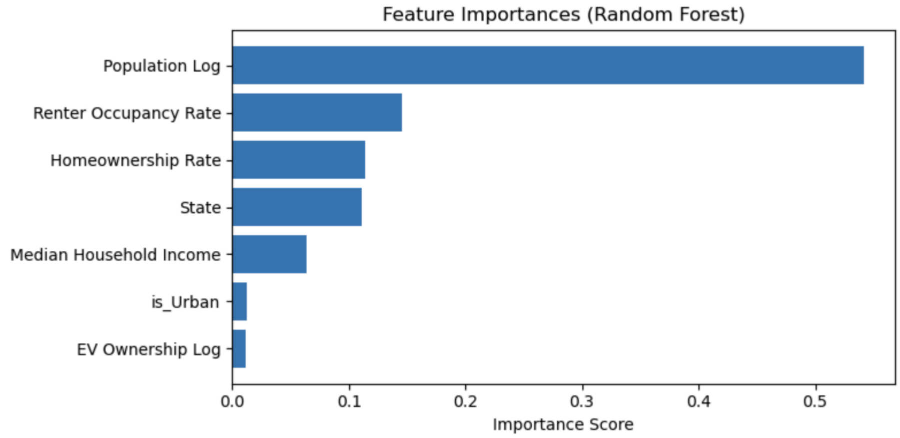
Urban/Rural (All)

ZIP Code: 98674
City and State: WOODLAND, WA
Charger Need (Predicted): 10
EV Ownership Count: 268
Public Charger Count: 0
EVs per 1,000 Residents: 16.0

One limitation of this study is that it uses EV registration data as a proxy for actual demand. However, the model could not find much value in the number of EVs registered to each ZIP code, suggesting this was not a strong predictor of modeling public charging infrastructure. This shortfall could mean that individuals who live outside the area, like commuters or tourists, are using public chargers more than those who live there, which would not be found in local EV registration data.

Figure 20

Important Features in the Random Forest Model



Implications

The intention for this analysis was to build off of similar previous research, such as that by Jiao et al. (2024), who found that public EV charger access in Austin, Texas, disproportionately favored higher-income, non-Hispanic White communities with more registered EVs. Their regression analysis found income and race to be statistically significant predictors of charger access. Similarly, this analysis applies predictive modeling to gaps in charger availability across ZIP codes on the U.S. West Coast that may exist based on socioeconomic and regional characteristics.

Public and private stakeholders should use this model to help guide the deployment of future EV infrastructure and prioritize ZIP codes where charger need is found. This tool can help close equity gaps by identifying underserved communities, especially historically marginalized ones, and support efforts to meet future EV adoption goals. A future direction to build off this analysis is to incorporate charger usage data generated from EV chargers connected to the internet, to see where high utilization areas can be. This data could distinguish EV chargers that may be underused or frequently occupied, to help determine where demand could be. Peak hours and seasonal trends could help guide decisions on where new chargers should be placed and the number of stations built. Another direction is to include traffic and commuting patterns based on GPS data to see where areas of high traffic could be, such as in commercial districts where current predictors may be less relevant. Utilizing heatmaps or clustering techniques could reveal chokepoints that lack sufficient charging support.

References

Alternative Fuels Data Center. (2025). *Alternative fueling station data download.* U.S. Department of Energy. Retrieved July 30, 2025, from https://afdc.energy.gov/data_download

California Energy Commission. (2025). *Zero-Emission Vehicle and Infrastructure Statistics.* https://www.energy.ca.gov/files/zev-and-infrastructure-stats-data

Harris, C.R., Millman, K.J., van der Walt, S.J. et al. *Array programming with NumPy*. Nature 585, 357–362 (2020). https://doi.org/10.1038/s41586-020-2649-2

Hill, C. A., Du, L., Johnson, M., & McCullough, B. D. (2024). *Comparing programming languages for data analytics: Accuracy of estimation in Python and R*. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 14(4), e1531. https://doi.org/10.1002/widm.1531

Hunter, J. D. (2007). *Matplotlib: A 2D graphics environment*. Computing in Science & Engineering, 9(3), 90–95. https://doi.org/10.1109/MCSE.2007.55

Jiao, J., Choi, S. J., & Nguyen, C. (2024). *Toward an equitable transportation electrification plan: Measuring public electric vehicle charging station access disparities in Austin, Texas.* PloS one, 19(9), e0309302. https://doi.org/10.1371/journal.pone.0309302

McKinney, W. (2010). *Data structures for statistical computing in Python*. In S. van der Walt & J. Millman (Eds.), Proceedings of the 9th Python in Science Conference (pp. 56–61). https://doi.org/10.25080/Majora-92bf1922-00a

Oregon Department of Energy. (2025). *Oregon Electric Vehicle Dashboard.* https://www.oregon.gov/energy/Data-and-Reports/Pages/Oregon-Electric-Vehicle-Dashboard.aspx

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., … & Duchesnay, E. (2011). *Scikit-learn: Machine learning in Python.* Journal of Machine Learning Research, 12, 2825–2830.

PwC. (n.d.). *US electric vehicle charging market growth*. https://www.pwc.com/us/en/industries/industrial-products/library/electric-vehicle-charging-market-growth.html

Rudin, C. (2019). *Stop explaining black box machine learning models for high-stakes decisions and use interpretable models instead.* Nature Machine Intelligence, 1, 206–215. https://doi.org/10.1038/s42256-019-0048-x

U.S. Census Bureau. (2024). *American Community Survey 5-year estimates, 2019–2023*. United States Department of Commerce. https://data.census.gov

U.S. Department of Agriculture Economic Research Service. (2020). *Rural-Urban Commuting Area Codes - ZIP Code Approximation File.* https://www.ers.usda.gov/data-products/rural-urban-commuting-area-codes/

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., ... & SciPy 1.0 Contributors. (2020). *SciPy 1.0: Fundamental algorithms for scientific computing in Python.* Nature Methods, 17(3), 261–272. https://doi.org/10.1038/s41592-019-0686-2

Washington State Department of Licensing. (2025). *Electric Vehicle Population Data.* https://catalog.data.gov/dataset/electric-vehicle-population-data

Waskom, M. L. (2021). *seaborn: statistical data visualization.* Journal of Open Source Software, 6(60), 3021. https://doi.org/10.21105/joss.03021

Yu, Q., Que, T., Cushing, L. J., & others. (2025). *Equity and reliability of public electric vehicle charging stations in the United States.* Nature Communications, 16, Article 5291. https://doi.org/10.1038/s41467-025-60091-y