Aug, 4, 2025

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

Regression Analysis on EV Charger Coverage

Jesus Torres-Carbajal
Western Governors University
Masters in Data Analytics Capstone

About Me

Bachelors in Physics – University of Washington

Masters student in Data Analytics – WGU

4+ years environmental laboratory analysis experience

Passion for applying data-driven insights to sustainability and clean energy initiatives

Electric vehicle (EV) adoption is rapidly increasing across the U.S., especially on the West Coast.

Public charging infrastructure has not grown at the same pace, leading to potential accessibility gaps.

Previous studies show disparities in charger distribution, often favoring higher-income or less diverse communities.

Understanding where chargers are most needed can help public policy and businesses invest resources more effectively.

"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.)

1. **Problem Statement & Hypothesis**
2. Data Analysis Process
3. Key Findings
4. Limitations
5. Proposed Actions
6. Expected Benefits

This study examines the gap in public EV charger coverage across West Coast ZIP codes, using demographic and geographic data to model where chargers are most needed.

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.

1. Problem Statement & Hypothesis
2. **Data Analysis Process**
3. Key Findings
4. Limitations
5. Proposed Actions
6. Expected Benefits

Data Analysis Process

Data Collection: Gathered from state EV registration datasets, U.S. Census Bureau, and DOE Alternative Fuel Data Center.

Data Cleaning: Used Python to merge datasets, handle missing values, and remove inconsistencies.

Feature Engineering: Created new prediction variables, (e.g Total EV Ownership).

Modeling: Applied Random Forest regression to predict optimal chargers per ZIP code.

Statistical Testing: Compared predicted vs. actual counts using one-sample Wilcoxon signed-rank test.

Data Sources

State EV Registration Data: Counts of registered EVs by ZIP code (WA, OR, CA).

U.S. Census Bureau (ACS): Demographic and socioeconomic data, including population, median income, and homeownership rates.

USDA Rural-Urban Classification: Identification for rural vs. urban ZIP codes.

DOE Alternative Fuel Data Center: Public charger location and count data.

Data Preparation & Feature Engineering

Data Integration: Merged state EV registration data, U.S. Census Bureau demographics, USDA rural–urban codes, and DOE charger location data by ZIP code.

Cleaning: Handled missing values, removed duplicates, standardized ZIP code formats.

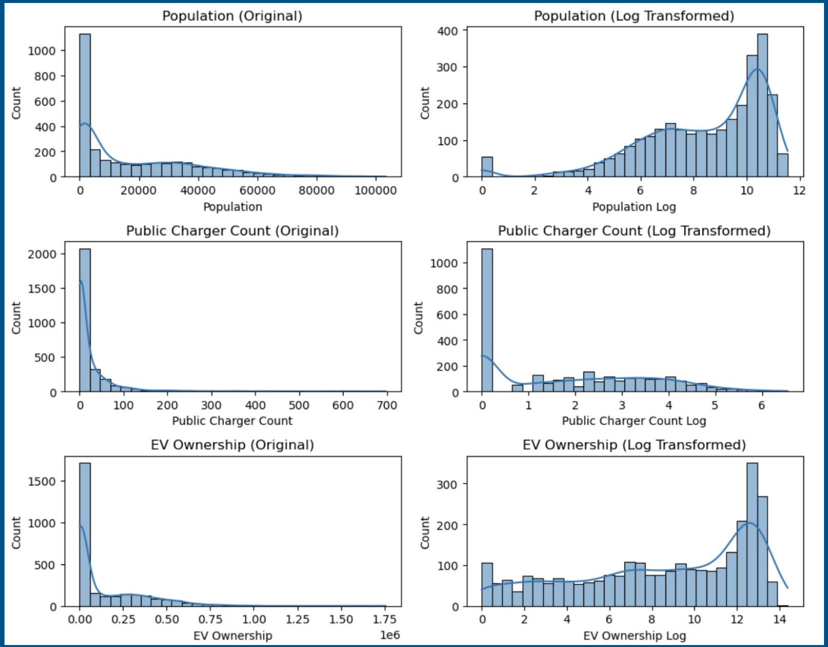
Feature Engineering:

- Created Charger-to-EV ratio.
- Encoded rural/urban classification.
- Calculated population density and homeownership rate from raw census data.

Purpose: Ensure consistent, analysis-ready dataset for accurate modeling.

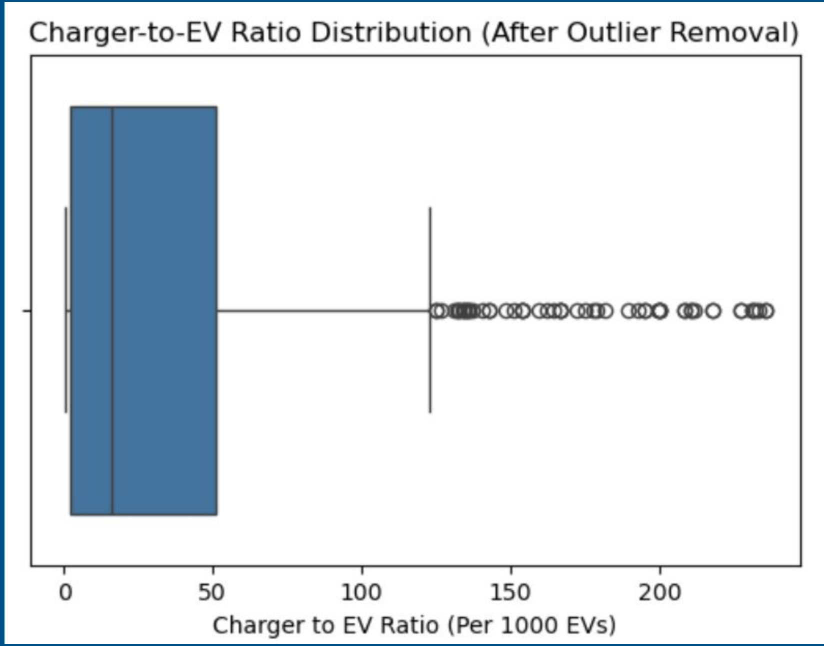
Distribution of Key Variables

Most variables were highly right-skewed, so log transformations were applied to improve model performance and meet regression assumptions.



Charger-to-EV Ratio Distribution for Threshold

- Outliers above the top 10% removed for clarity
- Boxplot shows remaining ratio spread across ZIP codes
- Used to inform Top 25% coverage threshold



Random Forest Model Performance & Fit

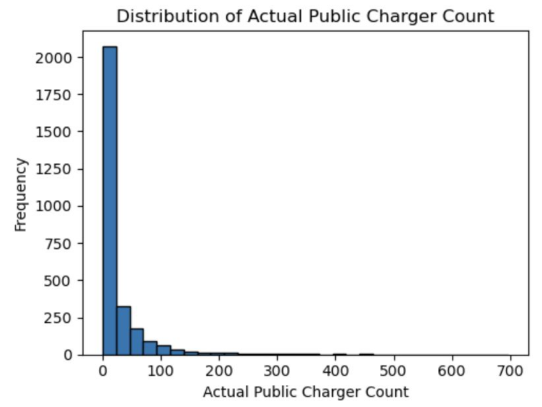
The Random Forest model predicts optimal charger counts by ZIP code.

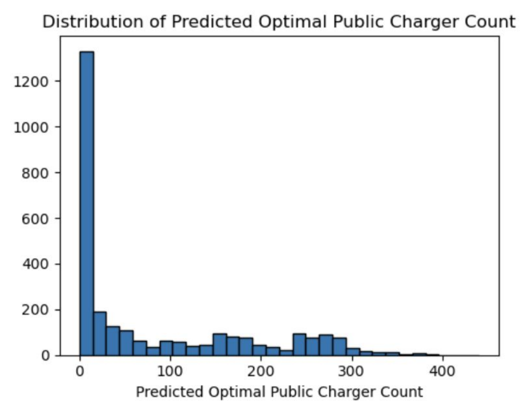
Predicted distribution better reflects infrastructure needs, though extreme values remain harder to estimate.

Performance metrics:

R-squared = 0.67

RMSE = 54 chargers

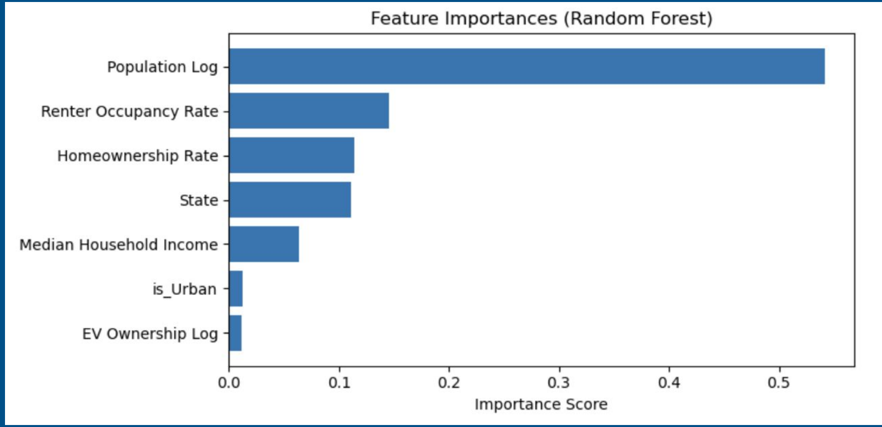




Random Forest Feature Importance

Key Variables Identified:

- Population size is the biggest factor, more people means more chargers needed
- Housing type matters, more renters or homeowners changes charging needs
- Income and location still matter, but not as much

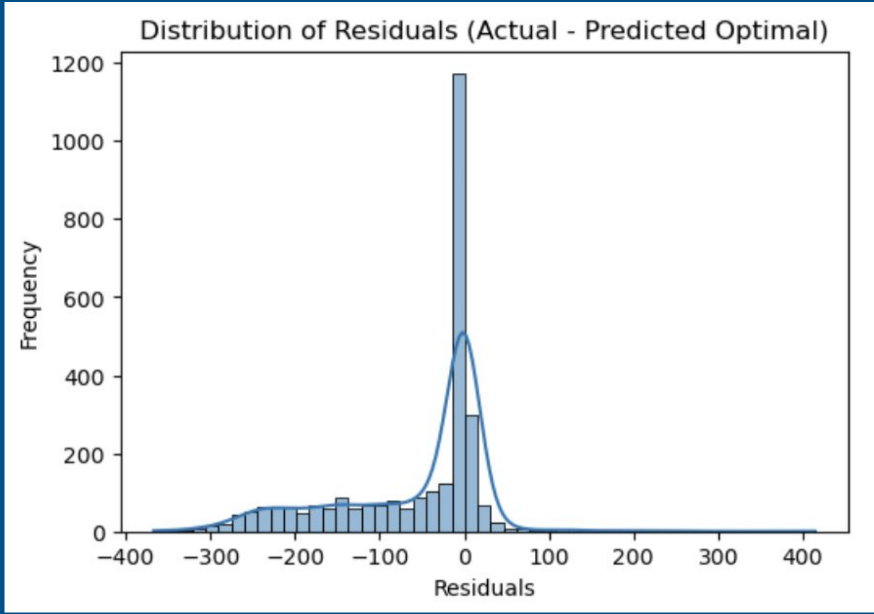


Residual Analysis & Statistical Test

Residuals mostly near 0, showing decent fit

Some extreme under predictions remain

Wilcoxon test: Stat = 278,645.5, $p < 0.005$, indicating a significant difference between predicted and actual counts.

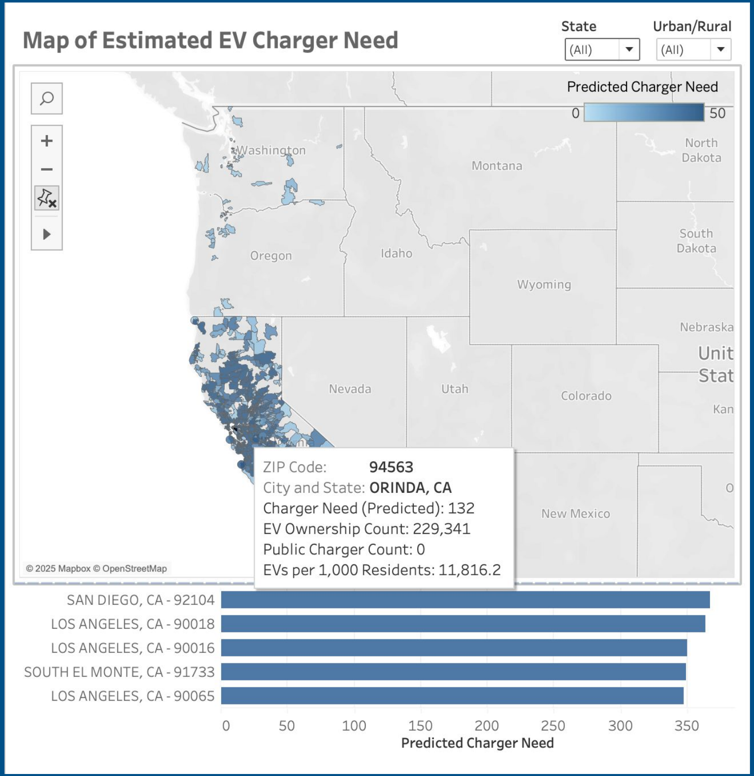


Geographic Distribution of Predicted EV Charger Needs

Map displays ZIP codes across WA, OR, and CA with predicted charger gaps.

Darker blue indicates higher estimated charger need.

Many of the ZIP codes with highest need are in California's urban centers, especially Los Angeles and San Diego.



Limitations

Proxy for Demand: EV registrations may not capture charger use by commuters, tourists, or fleets.

Weak Predictors: EV ownership showed limited value in predicting charger needs.

Data Timelines: Demographic data may be outdated.

Outlier Removal: May have excluded real high-demand areas (e.g., tourist hubs).

Geographic Scope: Results apply only to CA, OR, and WA.

Proposed Actions

Use this model to guide public & private EV infrastructure planning.

Prioritize ZIP codes with highest identified charger needs.

Target underserved and historically marginalized communities.

Incorporate real charger usage data to identify high-utilization vs. underused stations.

Analyze GPS traffic and commuting patterns to spot high-traffic areas lacking chargers.

> **“Public EV charger access has the tendency to favor higher-income, non-Hispanic White communities.”**
> **- Jiao et al., 2024**

1. Problem Statement & Hypothesis
2. Data Analysis Process
3. Key Findings
4. Limitations
5. Proposed Actions
6. **Expected Benefits**

Expected Benefits

Directs chargers to underserved and historically marginalized communities, to help prevent unequal access.

Place chargers in high-demand areas, improving utilization rates.

Support state and local EV adoption goals with data-driven infrastructure planning.

Reduce wasted resources by avoiding overbuilt or underused areas.

References

Alternative Fuels Data Center. (2025). Alternative fueling station data download. U.S. Department of Energy. 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

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

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

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

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

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

Interactive Dashboard:
https://public.tableau.com/views/Capstone_Project_MSDA_Tableau/EstimatedEVChargerNeedMap?:language=en-US&publish=yes&:sid=&:display_count=n&:origin=viz_share_link