**[Analyzing Electric Vehicle Trends and Charging Infrastructure in Washington State](https://public.tableau.com/views/WashingtonElectricVehicleAnalysis-final/Story?:language=en-US&publish=yes&:display_count=n&:origin=viz_share_link)**

Washington State has enacted laws aimed at lowering its total greenhouse gas emissions. Nearly 40% of the state's yearly greenhouse gas emissions come from transportation and increasing electric vehicle (EV) adoption is key in achieving this goal.

The state has adopted new standards for low-emission vehicles, set to take effect in 2024, and has banned the sale of new gas-powered vehicles by 2035. In 2022, the US federal government allocated $5 billion to create a nationwide network of EV charging stations (EVCS), prioritizing Level 3 fast charging stations located along highways.

Approximately 80% of EV charging occurs at homes. Public EVCS are important for travel, fast charging, serving residents of multi-unit dwellings, and more.

EVCS equipped with Level 3 fast charging ports (a.k.a. DCFC) allow for rapid EV charging, taking about 30 minutes compared to a couple of hours required by a Level 2 charging port. Level 1 chargers, typically used at home, can take 8 to 12 hours to fully charge a vehicle. Fast charging is important for drivers during long journeys or users needing to recharge their vehicles quickly.

**Exploratory Analysis**

Among all the relationships investigated, the one of most interest was between the number of EV and the number of EVCS. The correlation between the total EV per county and the total EVCS per county was **0.94**. The correlation between the total EV per county and the estimated population was **0.86**. The correlation between the estimated population and the total EVCS per county was **0.69**.

These strong correlations led to the formation of my hypothesis: **As the number of electric vehicles per capita increases, there is a corresponding increase in the number of EVCS.**

**Seattle**, Washington is one of the **most populous** cities in the Pacific Northwest and is located in King County. **King County** contains **86%** of the state population, **94%** of the registered **EV** within the state, and **95%** of the **EVCS** in the state.

When evaluating the data on a county level, the inclusion of populous King County has the potential to disproportionately influence the relationships observed, potentially introducing skewed or erroneous conclusions when aggregated with smaller counties.

To assess the relationship across counties of varying sizes, a cluster analysis was conducted to group the counties based on similar characteristics, allowing for an independent evaluation of the relationships between variables in each cluster.

**Cluster Analysis**

I analyzed the counties using the k-means algorithm and created clusters based on their EV, EVCS, population statistics, and Level 3 charging ports.

The '**High**' cluster is composed of **only** **King County**. This cluster is an **urban center**.

The '**Medium**' cluster consists of **six** counties that **represent mixed suburban** **areas**.

The '**Low**' cluster includes **thirty-two** counties that primarily encompass **rural areas**.

**Linear Regression Analysis**

I performed a linear regression analysis for each cluster with **EV per capita** as the independent value (X) and with the **cumulative number of EVCS** as the dependent variable (y) to test the null hypothesis that there is no significant linear relationship between the increase in the total number of EVs per capita and the cumulative number of EVCS. The alternative hypothesis was that there is a significant relationship between the two variables. The confidence interval used was 95%.

For all clusters, the regression analysis produced a p-value that was **<0.001**, indicating a **strong relationship** between the two variables and providing evidence to reject the null hypothesis and accept the alternative hypothesis for each cluster.

The analysis for each cluster produced a notably different R-squared value. R-squared is a value that ranges from 0 to 1 and indicates how much of the variation in the dependent variable can be explained by the independent variable. A low value indicates that only a small amount of variability can be explained by the independent variable.

**Machine Learning Modelling - Linear Regression**

In order to limit the scope of linear regression modeling, I focused on the High and Medium clusters, where EV per capita is a significant factor influencing the number of EVCS.

For each cluster, I split the respective dataset into a training and test set, fitted a Linear Regression algorithm to the training data, created a model, and used the model to predict the values of the test data to compare the results to the actual data. I then calculated the model summary statistics to evaluate its’ accuracy.

For the **High** cluster, R-squared values of **0.957** (training) and **0.935** (test) indicating a model with a **strong fit**. This model should effectively predict numbers of EVCS in King County.

In the **Medium** cluster, R-squared values of **0.352** (training) and **0.324** (test) indicate a model with only **moderate predictive accuracy**. When **plotted individually**, each county in the Medium cluster has an R-squared value **above 0.94**, suggesting that **a portion of the cluster variance came from grouping different counties together**.

**CONCLUSIONS**

The predictive model for the **High** cluster could be used by planners to **budget for infrastructure needs, plan for additional electrical demand**, and **could help inform decisions about charging infrastructure**.

While the **Medium** cluster’s predictive model could be used to **budget for infrastructure needs,** and **inform decisions about charging infrastructure** in its counties, the model's **moderate R-squared value** indicates that the accuracy of its predictions will vary depending and **should be used with caution when making decisions**.

Performing a cluster analysis allowed the High cluster (King County) to be evaluated independently of the other clusters, however the differences between the counties within the Medium and Low clusters likely prevented accurate predictive models being calculated for them.

**PROJECT LIMITATIONS**

The training dataset for the High cluster and the Medium cluster (77 and 462 observations, respectively) are **very small datasets** and are **not large enough to make robust predictive models**.

The linear regression analysis indicates a relationship between the number of EVs per capita and the number of EVCS. However, the causality remains uncertain—whether one variable leads to the increase of the other, or if **both are influenced by external factors** such as **governmental policies promoting EV adoption** and **charging infrastructure development**.

The model makes predictions based on historical trends. As new data becomes available, it would require regular updates and retraining to maintain accuracy and relevance over time.

This project focused on investigating the relationship between EVs per capita and EVCS within a county. External factors such as government policies, commuting patterns, and proximity to major highways could affect the installation of EVCS. These factors could result in EVCS being installed beyond the needs of local EV users and affect the strength of the relationship between these variables.

**FURTHER ANALYSIS**

Dividing counties into smaller, more detailed sections using census tract information could create an opportunity for better clustering and make for more precise predictive models.

Expanding the analysis to incorporate areas outside of Washingtion State would offer a larger dataset and potentially strengthen the predictive power of the model.

Other factors, such as traffic patterns and strategic positioning of charging stations along major roads, are other areas to investigate when determining effective EVCS distribution for Washington State's goals of increasing EV adoption and achieving substantial reductions in greenhouse gas emissions.

In addition to EV and EVCS initiatives, Washington State includes other zero-emission vehicles in its strategy for reducing greenhouse gases. Further analysis could incorporate hydrogen fuel cell vehicles and compressed hydrogen charging stations into the analysis.

Further analysis to improve the adoption of EVs could be done with usage data on EVCS :

- Are charging stations busier during specific times?

- Do users prefer multiple partial charges or driving until a full charge is needed?

- Are usage patterns different between Level 2 and Level 3 ports?

- Does charging price impact user behavior?

- Are specific charging connectors more common or utilized more frequently than others?