**Influence of a county’s socio-economic variables - electric vehicle (EV) population, per capita income, population estimates, and unemployment rates – on the distribution and availability of charging stations in various counties**

**Pradhi Kohli**

**November 14, 2024**

**Abstract**

Electric vehicle (EV) charging stations are essential for supporting the widespread adoption of electric vehicles (EVs). Convenient and easily accessible charging options help alleviate range anxiety for drivers. However, EV charging infrastructure is unevenly distributed. This study aims to evaluate how key socio-economic variables — EV population, per capita income, population estimates, and unemployment rates — influence the distribution and availability of EV charging stations across various counties. The county-level data for Washington State and New Jersey from 2020 to 2022 is cleaned and normalized in Excel. Using Python programming language, Exploratory data analysis and Descriptive Statistics analyses are conducted. Data Visualization demonstrates the relationship between the independent variable (IV) - EV population, per capita income, population estimates, unemployment rates - and the dependent variable (DV) - EV charging stations. These factors are crucial to understand the pattern where there is a higher demand for charging infrastructure and a greater economic capacity to support infrastructure development. The ordinary least square (OLS) regression test indicates the effect of each predictor on the dependent variable. It reveals that population Estimate and EV registered show strong relationships with charging stations, while unemployment rate and per capita personal income appear to be less significant. Therefore, as per the model, counties with higher EV populations, and more densely populated areas tend to have more robust charging station networks. However, the unemployment rate and per capita personal income variables appear to be less significant and might be removed or further investigated. Understanding these socio-economic dynamics is crucial for policymakers and stakeholders to ensure equal access to EV charging infrastructure as the transition to electric mobility accelerates. By addressing these disparities, a more inclusive and sustainable adoption of EVs could be achieved, ultimately contributing to global efforts for reducing carbon emissions and limiting climate change.

**Introduction**

The transition to EVs is a critical component of global efforts to reduce carbon emissions and minimize climate change (Haidar et al., 2022). As governments and industries collaborate to promote EV adoption, the expansion of charging infrastructure has become a key challenge (Debnath et al., 2021). The availability and distribution of charging stations is essential for ensuring that EV users have reliable access to charging, which in turn affects the broader adoption of electric mobility (Debnath et al., 2021).

However, the deployment of EV charging infrastructure is not uniform across counties and is influenced by a variety of socio-economic factors that influence local demand, investment capacity, and planning priorities (Chen et al., 2020). Key socio-economic variables such as income distribution, technology development trends, demographics, and lifestyle factors play a significant role in EV adoption and charging stations distribution. (Debnath et al., 2021)

Areas with a higher concentration of EVs are more likely to see increased demand for charging stations, driving local governments and businesses to invest in infrastructure (Ai et al., 2018). This higher demand can be attributed to the need for convenience and accessibility, which in turn, encourages further EV adoption. Similarly, counties with higher per capita incomes tend to have the financial resources to support more extensive infrastructure development. Wealthier regions are often able to allocate more funds to public and private charging networks, offer incentives for EV purchases, and support broader public awareness campaigns (Chen et al., 2020).

Conversely, economically disadvantaged areas, as indicated by higher unemployment rates, may face challenges in funding or prioritizing the expansion of EV charging networks, potentially leading to disparities in access to charging stations (Esmaili et al., 2024). In such regions, a lack of financial resources can hinder the development of necessary infrastructure, creating a barrier to EV adoption and contributing to the divide between affluent and less affluent areas.

Population density also plays a crucial role in the implementation of charging stations. Urban areas with higher population densities typically experience more significant demand for charging infrastructure due to the larger number of potential EV users (Esmaili et al., 2024). These areas often benefit from more substantial investments in public infrastructure, making it easier to justify the cost of installing and maintaining charging stations. In contrast, rural and low-density areas may struggle with the economics of infrastructure investment due to the lower number of EV users and longer distances between charging points (Haidar et al., 2022).

By examining the impact of socio-economic variables on the distribution and availability of EV charging stations at the county level from 2020 to 2022, policymakers can develop future strategies to address variations and support the widespread adoption of EVs (Chen et al., 2020). This approach may help ensure that the benefits of EV adoption, such as reduced emissions and improved public health, are shared equally across all regions, contributing to a more sustainable future (Debnath et al., 2018).

**Problem Statement:**

How do a county’s selected socio-economic variables - EV population, per capita income, population estimates, and unemployment rates - from 2020 to 2022 influence the distribution and availability of charging stations in various counties?

**Objectives:**

Higher EV populations, increased per capita income, larger population estimates, and lower unemployment rate are positively associated with a greater distribution and availability of charging stations across various counties from 2020 to 2022.

**Significance**

The significance of understanding the relationship between socio-economic factors and the distribution of EV charging stations study resides in the potential to inform policies and strategies that support a fair and balanced transition to EVs (Debnath, et al., 2018). As the adoption of EVs grows, this research strives to ensure that charging infrastructure is distributed in a way that reflects the diverse needs and economic capabilities of different regions.

Access to charging infrastructure must be similar across both urban and rural areas, as well as affluent and economically disadvantaged regions. Without a comprehensive understanding of how socio-economic factors influence charging station placement, underserved communities could face barriers to EV adoption due to limited access to charging infrastructure. (Chen et al., 2020)

By targeting infrastructure investment in areas with higher demand potential or where socio-economic conditions suggest greater need, stakeholders must optimize the increase the impact of their infrastructure development efforts. Accelerating EV Adoption has a direct relation to the availability of charging stations. With a well-distributed charging infrastructure and sufficient economic resources, consumers are more likely to make the switch to EVs.

Supporting environmental goals by understanding how socio-economic variables influence charging infrastructure, policymakers must develop strategies to accelerate the adoption of EVs in all regions, contributing to national and global climate goals.

Data-driven policy formulation based on socio-economic factors will help shape future policy decisions regarding the expansion of EV charging networks. For example, policies that target areas with high unemployment or low per capita income could ensure that infrastructure development is not only driven by demand but also by a commitment to easy access.

In summary, understanding the interplay between socio-economic factors and charging station distribution is essential for developing effective policies and the growth of electric mobility.

**Methodology**

*Data Description*

The data from the counties of two distinct states - WA and NJ - are selected. These 2 states are located several miles apart (Map 1). Washington (WA) state is about 9 times bigger than New Jersey (NJ). NJ is 8.9% more expensive than WA. There are 39 counties in WA and 21 counties in NJ. Both WA and NJ offer a broad spectrum of potential areas for analysis such as social, economic, transportation, etc. Both these states have been striving to do well in EV adoption and infrastructure. Differences between these states provide a rich context for understanding role of regional variations in EV adoption, infrastructure development, policy impact, and other socio-economic variables. Additionally, the comparison could reveal valuable insights into how different states approach sustainability, technological adoption, and public policy.

The following data and websites are used to obtain datasets:

Charging Station data: Alternative Fuels Data Center (AFDC) provides the number of EVs per state and charging stations per county. [https://afdc.energy.gov/data\_download?download[data][api]=alt\_fuel\_stations&download[data][timeframe]=historical](https://afdc.energy.gov/data_download?download%5bdata%5d%5bapi%5d=alt_fuel_stations&download%5bdata%5d%5btimeframe%5d=historical)

Electric vehicles population data for counties is available on the website: <https://www.atlasevhub.com/materials/state-ev-registration-data/#data> <https://catalog.data.gov/dataset/mva-electric-and-hybrid-vehicle-registrations-by-county-as-of-october-2020>

Source for Per Capita Income is accessible by the Bureau of Economic Analysis weblink: <https://www.bea.gov/data/income-saving/personal-income-county-metro-and-other-areas>

Population Estimates data link is available on the Economic Research Service website: <https://www.ers.usda.gov/data-products/county-level-data-sets/county-level-data-sets-download-data/>

Source for Unemployment data is downloaded from website: <https://www.ers.usda.gov/data-products/county-level-data-sets/county-level-data-sets-download-data/>

**Map 1**: Map of United States. WA state and NJ are highlighted in green.

A map of the united states

Description automatically generated

*Preprocessing*

Excel software is utilized extensively to organize, clean, and combine various data files. This data is sourced from five distinct origins, ensuring a diverse set of information. When working with time-series data covering the period from 2020 to 2022, consistency and comparability across different years is maintained, and Excel’s PivotTables play a significant role in achieving this. Sum, average, and ratio functions are used to link multiple variables, such as different types of EV charging stations and the number of EVs relative to the total number of vehicles. To refine the dataset, the following methods were used:

Data Integration: Merging data from different sources and ensuring that the combined dataset is coherent and consistent. This involved using joins, concatenations, and aligning data fields.

Removing Duplicate Data: To avoid skewed analysis, duplicate entries were identified and eliminated.

Handling Missing Values: Rows/columns with a significant amount of missing data were removed.

Standardizing Formats: Consistent format is ensured. For example, dates are all made to follow the same format, and text data is standardized about casing and spelling.

Correcting Errors: Consistent attention is paid while identifying and fixing errors or inconsistencies within the data. This also included correcting typos.

Data Transformation: Data is converted into a suitable format for analysis. This includes aggregation, pivoting, and encoding categorical variables into numerical values. Furthermore, filtering and sorting methods are applied, allowing for a focused analysis of specific subsets of the data. VLOOKUP () functions are used to merge data from various tables based on common keys, streamlining the integration process. Additionally, new columns are created, and the data is carefully formatted to ensure compatibility with Python software, preventing errors and facilitating seamless data processing.

*Analytical Methods*

Data visualization using the Python libraries is successfully implemented to illustrate the relationship between the IV and the DV. The process began with the selection and preparation of the dataset, ensuring it is clean and ready for visual analysis. The following analytical methods are used:

Exploratory data analysis (EDA) on the dataset is performed using the Python programming language, utilizing libraries such as Pandas, NumPy, and Matplotlib. To understand the relationship between the IV and the DV, measures of central tendency including mean, median, mode, minimum, and maximum values are calculated for both variables (Table 1). This approach assists in exploring the distribution, spread, and potential outliers within the dataset, along with a comprehensive understanding of the data characteristics. Visualizations achieved via histograms and scatter plots illustrate these statistical summaries, aiding in the interpretation of the patterns and relationships between the variables (Figure1 and 2).

A comprehensive Descriptive Statistics analysis is conducted on the counties of Washington State and New Jersey for the years 2020, 2021, and 2022. There are 39 counties in WA state and 21 counties in NJ. In addition to calculating central tendency, this analysis involved calculating measures of dispersion including standard deviation, variance, and range values (table 1). The purpose is to identify trends and variations in key demographic and economic indicators across the specified period of 3 years. Frequency distributions are generated to understand the data distribution within each county. Comparative analyses between the two states are performed to highlight any significant differences or similarities in their statistics. Trend line data visualizations (Graph 1, 2, 3, 4) are created to view a clear representation of the relationship between the IV and the DV. These visual tools facilitated the identification of patterns, outliers, and potential areas of interest within the data. These statistical measures provided insights into the demographic shifts, economic changes, and overall trends affecting the counties in WA and NJ over the three years.

**Figure 1:** New EV Registered: NJ vs WA (2020-2022)

A graph with green and blue squares

Description automatically generated

**Figure 2:** Change in EV charging stations: NJ vs WA (2020-2022)

A graph of a bar chart

Description automatically generated with medium confidence

**Table1:** Descriptive Statistics



*Validation:*

For the selected dataset, after the initial data is preprocessed and cleaned, the multiple regression algorithm OLS is performed. Since the data is continuous, OLS regression is selected.

The supervised machine learning algorithm, OLS, provides the linear relationship between the socio-economic factors -- EV population, per capita income, unemployment rates, the availability of EV, and the charging stations. The coefficients obtained from the OLS regression provide a clear measure of the impact that each IV has on the DV charging stations. The scatter plots with regression graphically depict the relationship between the DV and each IV, highlighting the fit of the model and key findings such as trends, patterns, and outliers. The regression line demonstrates how the changes in socio-economic factors are expected to impact the availability of charging stations. This comprehensive analysis presents the overall strength and direction of the relationships between variables.

For model forecasting, the dataset is divided into two subsets: a training set and a testing set, using an 80/20 split. The training set comprises 80% of the data and is used to train the OLS regression model, enabling it to learn the underlying patterns and relationships. The remaining 20% of the data is reserved as the testing set, which is used to evaluate the model's performance and validate its predictive accuracy.

A model's performance is evaluated on the test set using metrics such as accuracy, precision, MSE, RMSE, and MAE for regression (Table 2).

KFold metrics represent the performance of the model during the KFold cross-validation process and is a more reliable performance estimate because it averages over multiple training and test splits. MSE (Mean Squared Error) and RMSE (Root Mean Squared Error) values are listed in Table 2. A low test KFold MSE (as compared to train KFold MSE) indicates that the model generalizes well to unseen data. KFold RMSE computed on the test set has lower value as compared to the train set, providing an indication of low error the model has on new, unseen data. A lower RMSE means the model's predictions are more accurate on test data.

OLS provides performance metrics on a single split of the data (Table 2). A higher train MSE means that the model is not fitting the data very well. OLS RMSE model on the training data represents the average error in predicting the target variable on the training data. A high test MSE means the model does not generalize well to new data. The OLS RMSE model on the test set suggests the number of errors (88.831523) the model has when predicting new data. OLS MSE/RMSE provides performance metrics on a single split of the data; therefore, it may not be a reliable an estimate of generalization performance as compared to KFold performance estimates.

**Table 2:** Model's Performance:

KFold MSE train: 97704.880349

KFold RMSE train : 312.577799

KFold MSE test: 854.920559

KFold RMSE test: 29.239025

OLS MSE train: 2216372.524527

OLS RMSE train: 1488.748644

OLS MSE test: 7891.039539

OLS RMSE test: 88.831523

*Visualization:*

**Graph1:** Density Plot for EV charging stations vs unemployment rate

**A graph showing the amount of unemployment rate and ev charging stations

Description automatically generated**

**Graph2:** Density Plot EV charging stations vs population estimate

**A graph with a red line and blue dots

Description automatically generated**

**Graph3:** Density Plot EV charging stations vs per capita personal income

**A graph with a red line and blue dotted line

Description automatically generated**

**Graph4:** Density Plot EV charging stations vs EV registered

**A graph with a red line

Description automatically generated**

**Results and Discussion**

The correlation matrix in Table 2 represents the pairwise correlations between five variables - EV population, per capita income, population estimates, unemployment rates, and the EV charging stations. Correlation indicates measures of the strength and the direction of the linear relationship between two variables.

There is a strong positive correlation between EV charging stations and Population estimate (0.830) indicating that with more people, there are more charging stations. Also, positive correlation between EV charging stations and EV registered (0.641) indicates more EVs tend to correlate with more charging stations.

The Moderate Positive Correlations between Population estimate, and EV registered (0.561) indicate larger populations tend to have more EV registrations. Population estimates and Per capita income (0.573) correlations indicate areas with larger populations tend to have higher per capita income. EV charging stations and Per capita income (0.481) correlation indicates higher income areas tend to have more charging stations.

There is a Weak or No Correlations between Unemployment rate and other variables, meaning that the relationships with the unemployment rate are weak and mostly negative, with no significant correlations to population size, per capita income, or EV registrations. Additionally, the weak Negative Correlations between Unemployment rate and EV charging stations (-0.113), means that there is a slight tendency for areas with higher unemployment rates to have fewer EV charging stations. Also, the correlation between Unemployment rate and Per capita personal income (-0.246) indicates that higher unemployment leads to weak association with lower personal income.

In summary, Population size is a key factor driving the number of EV charging stations and EV registrations, as evidenced by the strong positive correlations with both. Per capita personal income is moderately positively correlated with both EV charging stations and EV registrations, but the relationship is weaker than with population size. Unemployment rate has minimal impacts on the other variables in this dataset, as most correlations are weak or negligible.

**Table2:** Correlation Matrix

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **EV charging stations** | **Unemployment Rate** | **Population Estimate** | **Per capita personal income** | **EV registered** |
| **EV charging stations** | 1 | -0.112590986 | 0.830488353 | 0.480999431 | 0.64093022 |
| **Unemployment Rate** | -0.112590986 | 1 | -0.065340423 | -0.24642863 | 0.039650602 |
| **Population Estimate** | 0.830488353 | -0.065340423 | 1 | 0.572603998 | 0.560952294 |
| **Per capita personal income** | 0.480999431 | -0.24642863 | 0.572603998 | 1 | 0.265570467 |
| **EV registered** | 0.64093022 | 0.039650602 | 0.560952294 | 0.265570467 | 1 |

**Table 2:** OLS Regression Results, where DV is charging stations and IVs are unemployment rate, population estimate, per capita personal income, EV registered

**A screenshot of a computer screen

Description automatically generated**

The R-squared of 0.759 suggests that the model explains about 76% of the variance in the DV, which is a good fit. The proportion of the variance in the charging stations (DV) is explained by the IVs in the model.

The adjusted R-squared of 0.752 accounts for the number of predictors in the model, indicating a strong fit. In other words, after accounting for the number of predictors, about 75.2% of the variance in DV is explained by the model.

The large F-statistic value (109.3) and its p-value (6.37e-42) show that the model is statistically significant and explains a substantial portion of the variability in the dependent variable. With a low p-value (almost zero), the null hypothesis is rejected, and the model provides meaningful insights.

The AIC (Akaike Information Criterion) and BIC (Bayesian Information Criterion) values are useful for comparing this model with other models. Lower value indicates a better-fitting model.

With 144 observations and 4 predictors, the model seems reasonably detailed and provides useful insights into the relationships between the IVs and the DV.

**Model Equation**: From the regression output (table 2), the following equation is obtained:

Y = 96.6776 – 23.1522 ⋅ x1 + 184.6341 ⋅ x2 + 1.2000 ⋅ x3 + 65.7721 ⋅ x4

Where, x1= Unemployment Rate, x2 = population Estimate, x3 =Per capita personal income, x4 = EV registered. 96.6776 is a constant value.

From table 2, the coefficients indicate the effect of each predictor on the DV. The significant predictors population Estimate x2 (184.6341) and EV registered x4 (65.7721) shows strong relationship with charging stations (y), while unemployment rate x1 (-23.1522) and per capita personal income x3 (1.2) appear to be less significant.

Model Fit: In table 2, the high t-statistics and low p-values for population Estimate (x2, t=10.991, p=0) and EV registered (x4, t=4.984, p=0) suggest they are important predictors, while Unemployment Rate (x1, t=-1.787, p=0.076) and Per capita personal income (x3, t=  0.073, p=0.942) might be removed or further investigated.

Residual Diagnostics: The Omnibus test, Jarque-Bera test, and the skewness and kurtosis values suggest potential issues with normality of the residuals, indicating that some model assumptions are violated. An omnibus value of 138.711 with a p-value of 0.000 suggests that the residuals do not follow a normal distribution, which may indicate potential issues such as outliers or non-normality. The Durbin-Watson value 1.793 indicates some positive autocorrelation but since it is relatively close to 2, therefore there is no autocorrelation and is not a major concern. JB value 3034.635 with p-value = 0.00 suggests that the residuals are not normally distributed. The skew of 0 means it is perfectly symmetric, while a positive value indicates that the distribution is skewed to the right. The skew of 3.299 suggests that the residuals are right skewed, indicating a non-normal distribution. Kurtosis value of 24.500 indicates that the residuals have tails, implying the presence of outliers. The condition number is a measure of multicollinearity in the model. A large condition number indicates that there may be high correlation among the predictors, which can lead to numerical instability in the regression estimates. Cond. No. of 2.74 suggests that multicollinearity is not a major concern in this model.

**Conclusion**

This study extracts results from the regression analysis to reveal how each of the four socio-economic factors vary in their effectiveness in determining the distribution of EV charging stations. EV population and county population density have strong positive relationships with the presence of charging stations, while unemployment rates and per capita income are less influential. These findings suggest that infrastructure development should be prioritized in regions with higher EV adoption and larger populations. While unemployment rates and per capita income show weak associations, further investigation into these variables could reveal additional insights. Policymakers and stakeholders should focus on addressing these disparities to ensure equitable access to EV charging stations, fostering the widespread adoption of EVs. By targeting underserved regions, governments can support a more inclusive transition to electric mobility, which is essential for advancing climate goals and reducing carbon emissions. Understanding these socio-economic dynamics is vital for creating sustainable and effective policies that support the growth of EV infrastructure.

**References**

Ai, S., Chakravorty, A., & Rong, C. (2018). Household EV charging demand prediction using machine and Ensemble Learning. *2018 IEEE International Conference on Energy Internet (ICEI)*, 163–168. https://doi.org/10.1109/icei.2018.00037

Chen, C., Zarazua de Rubens, G., Noel, L., Kester, J., & Sovacool, B. K. (2020). Assessing the socio-demographic, technical, economic and behavioral factors of Nordic electric vehicle adoption and the influence of vehicle-to-grid preferences. *Renewable and Sustainable Energy Reviews*, *121*, 109692. https://doi.org/10.1016/j.rser.2019.109692

*Cost of living in Washington (US) compared to New Jersey (US)*. MyLifeElsewhere. (n.d.). https://www.mylifeelsewhere.com/cost-of-living/washington-usa/new-jersey-usa

*Data downloads*. Alternative Fuels Data Center: Data Downloads. (n.d.). https://afdc.energy.gov/data\_download?download%5Bdata%5D%5Bapi%5D=alt\_fuel\_stations&download%5Bdata%5D%5Btimeframe%5D=historical

Debnath, R., Bardhan, R., Reiner, D. M., & Miller, J. R. (2021). Political, economic, social, technological, legal and environmental dimensions of electric vehicle adoption in the United States: A social-media interaction analysis. *Renewable and Sustainable Energy Reviews*, *152*, 111707. https://doi.org/10.1016/j.rser.2021.111707

*Download Data*. USDA ERS - County-level Data Sets: Download Data. (n.d.). https://www.ers.usda.gov/data-products/county-level-data-sets/county-level-data-sets-download-data/

Esmaili, A., Oshanreh, M. M., Naderian, S., MacKenzie, D., & Chen, C. (2024). Assessing the spatial distributions of public electric vehicle charging stations with emphasis on equity considerations in King County, Washington. *Sustainable Cities and Society*, *107*, 105409. https://doi.org/10.1016/j.scs.2024.105409

Ruder, A. (2024, November 4). *State EV registration data*. Atlas EV Hub. https://www.atlasevhub.com/materials/state-ev-registration-data/#data  https://catalog.data.gov/dataset/mva-electric-and-hybrid-vehicle-registrations-by-county-as-of-october-2020

Haidar, B., & Aguilar Rojas, M. T. (2022). The relationship between public charging infrastructure deployment and other socio-economic factors and electric vehicle adoption in France. *Research in Transportation Economics*, *95*, 101208. https://doi.org/10.1016/j.retrec.2022.101208

*Interactive map of visited states*. Get A Clickable Interactive US Map. (n.d.). https://www.fla-shop.com/visited-states/?st=NJ%2CWA&vc=1ca032&uc=90cfea&hc=40bfa6&bc=ffffff&sl=on

*Personal income by county, Metro, and other areas*. Personal Income by County, Metro, and Other Areas | U.S. Bureau of Economic Analysis (BEA). (n.d.). https://www.bea.gov/data/income-saving/personal-income-county-metro-and-other-areas

Ruoso, A. C., & Ribeiro, J. L. (2022). The influence of countries’ socioeconomic characteristics on the adoption of electric vehicle. *Energy for Sustainable Development*, *71*, 251–262. https://doi.org/10.1016/j.esd.2022.10.003

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