EV Infrastructure Research

Daniel Kong
dskong@ucsd.edu

Phi Nguyen
pnguyen@sdge.com

Abstract

This project examines electric vehicle (EV) infrastructure within the context of San Diego Gas & Electric (SDG&E), focusing on exploratory data analysis (EDA) of key datasets to support infrastructure planning and development. Leveraging the Alternative Fuels Data Center (AFDC) fuel station dataset, the OpenStreetMap NetworkX (OSMnx) package, the US Census API python package (Cenpy), and the California vehicle registration dataset from the Department of Motor Vehicles, this study provides foundational insights into the distribution of EV stations and vehicle registration trends across the SDG&E service area. By working with these datasets, this project aims to build familiarity with critical data sources, software packages, and methodologies essential to EV infrastructure analysis, enhancing domain knowledge necessary for effective EV ecosystem development.

Code: https://github.com/dskong07/DSC180-ev-infra

1 Introduction

1.1 AFDC Dataset

The AFDC application programming interface (API) and dataset (Alternative Fuels Data Center 2024) provide a comprehensive overview of alternative fueling stations across the United States and Canada. This dataset, maintained by the U.S. Department of Energy, includes critical information such as station locations, opening date, connector types, fuel types, accessibility, and operational status, making it a valuable resource for investigating both EV and fueling station infrastructure at large.

The EDA of the AFDC dataset primarily looks at the spatial and temporal distribution of EV and alternative fuel stations within both the state of California as well as the SDG&E service area. The analysis examines station positioning, geographic coverage, and opening dates over time, providing insights into the state of primarily EV charging infrastructure. This exploration serves to help the ideation of key patterns, correlated statistics, and potential methodological visual representation of metrics such as service availability, laying a foundation for future infrastructure investigation or planning.

1.2 Vehicle Registration Dataset

The Vehicle Fuel Type Count by Zip Code dataset (California Department of Motor Vehicles 2019-2024) provides a breakdown of registered vehicles by fuel type across California zip codes for the years 2024, 2023, 2022, 2021, 2020, and 2019. This dataset includes counts for gasoline, electric, hybrid, and other alternative fuel vehicle registration numbers in each zip code, providing insight into the distributions of fuel preferences and changes over time at a localized level.

The EDA of the Vehicle Registration dataset investigates the distribution of different fuel-type cars across zip codes in California, as well as the changes to those distributions across the years stated above.

1.3 Census Data and Geospatial Feature Exploration

In this project, the usage and exploration of Python packages OSMnx and Cenpy play a pivotal role in extending any understanding of spatial accessibility and population demographics, respectively, within the context of EV station infrastructure.

OSMnx facilitates the acquisition and analysis of spatial street network data, allowing for an in-depth exploration of EV station accessibility, routing, and geographic distribution. This library allows us to explore the data in graph format, utilizing node and edge notation to calculate distance, paths, and more complex integrations with additional data to create visualizations such as geographic heat maps.

Cenpy serves as a valuable resource for integrating U.S. Census data, providing an examination of demographic statistics and patterns that intersect with vehicle trends and fuel station infrastructure.

2 Exploratory Data Analysis Methods

2.1 AFDC EDA

The AFDC Datasets were gathering using the public AFDC API with a personal key. Additional parameters were set to create distinct datasets for the United States entirely and the state of California alone. The analysis comprises of 3 primary components: raw data inspection, geographical analysis, and time series analysis.

Within the raw data inspection, the formatting and structure of the data is explored, and missing values, duplicate entries, and erroneous inputs are investigated. The missing values in the datasets were primarily due to the differing types of fuel stations, therefore leaving columns such as types of ev charging connectors empty in a hydrogen fuel cell data entry. Duplicate entries exist, but are likely due to multiple stations clustered in one location. Impossible data points, such as a latitude longitude in China within the California dataset, are then either purged or corrected using google maps as a reference. A preliminary investigation for the distributions of alternative fuel stations across states is also conducted.

Geographical analysis primarily focuses on California as a whole. The stations are plotted onto a map of California, and overlaid with additional layers, such as demographic data including per capita income and population density, as well as a map of power transmission lines (California Energy Commission 2024) in the relevant regions. To properly see localized patterns, the python libraries geopandas and folium are used to create interactive cloropleth maps for demographic data with regards to station locations.

Time series analysis investigates the growth and adoption of alternative fuel stations across the included time periods, additionally stratifying the dataset by state. To investigate country-wide patterns on the state level, a logarithmic scaling is used to visualize state fuel station adoption.

2.2 DMV EDA

A similar approach is taken initially for the Vehicle Registration dataset. This dataset is gathered via the ca.data.gov public API. Overall, there are 6 individual datasets each for the years of 2024, 2023, 2022, 2021, 2020, 2019. The data is then split by zip code to analyze vehicle registration distributions, and then combined to create plots of registration trends over the years included. The registration data is then refined to contain EV registrations only, and grouped by zip code and year. Any missing values are imputed via the implied distribution of non-EV registration trends in a given zip code. Zip codes missing too much data to be properly or reasonably imputed are given a registration count of 0, as missing values are stated by the organization maintaining the dataset to have a true value of 10 or lower. After imputation, the data is used to calculate the change in EV registrations year over year for each individual zip code , and fitted to a poisson distribution, modeling the growth curves of EV adoption over time.

3 Geospatial and Cross-sectional Dataset Exploration

The DMV vehicle registration dataset and government popluation data gathered via Cenpy were joined on their zip codes (California Department of Education 2024) to create a dataset containing registration trends as well as demographic and census data in each zip code. Utilizing these combinations, metrics such as EV registrations per capita in a given zip code were calculated, allowing for a more thorough and wholistic perspective of analyzing the relationship between demographic and geospatial patterns. This joined dataset, along with the newly generated features, was then used to plot chloropleth maps overlaid by EV charging stations in order to identify any potential trends between vehicle registration rates, population, and the locality and density of charging stations. Futher exploration on the capabilities and potential use cases of geospatial visualizations were enhanced via the use of the OSMnx package. Utilization of a graph-based network model allows for representations of distance by way of driving or foot traffic, and further enables a more granular level investigation into demographic trends across geographic distributions experienced in real life, such as optimal driving paths, or overlaying demographic data to visualize realistic and practical access to facilities.

4 Results, Findings, and Discussion

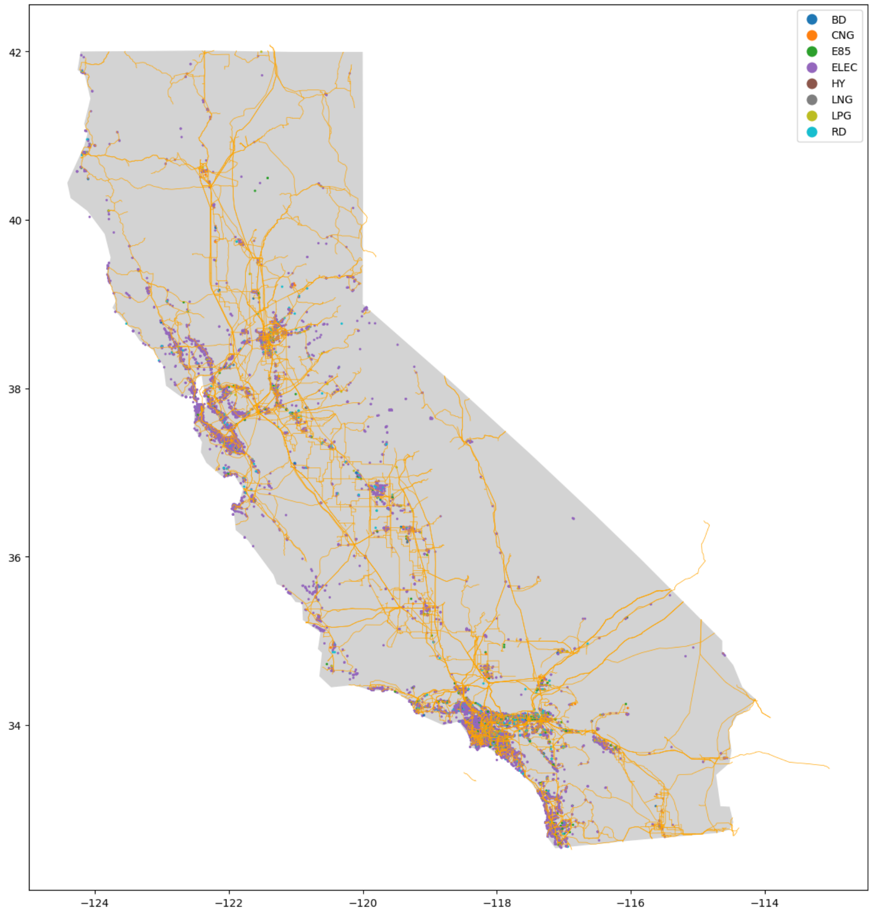

Figure 1: Alternative fuel stations overlaid with electrical transmission lines in CA.

Figure 1 plots alternative fuel station locations across California, overlaid with transmission line data gathered from the California State Geoportal, maintained by the government of California. Notably, electric car charging stations, represented by purple dots on 1, present a significant clustering behavior around the major transmission lines across the state. However, the correlation is somewhat shared by other alternative fuel stations, and may arise for many reasons, the most prominent of which being that population dense regions are more likely to have higher needs for energy and resources, and also contain the highest number of electric vehicles, ergo needing more charging stations. However, this pattern still remains an interesting potential topic for future investigations.

Figures 2 and 3 reveal notable trends regarding the placement of fuel stations. Specifically, while fuel stations are often situated near high-income areas, they are less frequently located within the wealthiest locales. Observations from fig.3 indicate that these stations tend to cluster in areas with high population density and are generally aligned along major highways. This aligns with expectations, as high-density areas are typically urban, and a greater population concentration generally implies a higher number of vehicle owners.

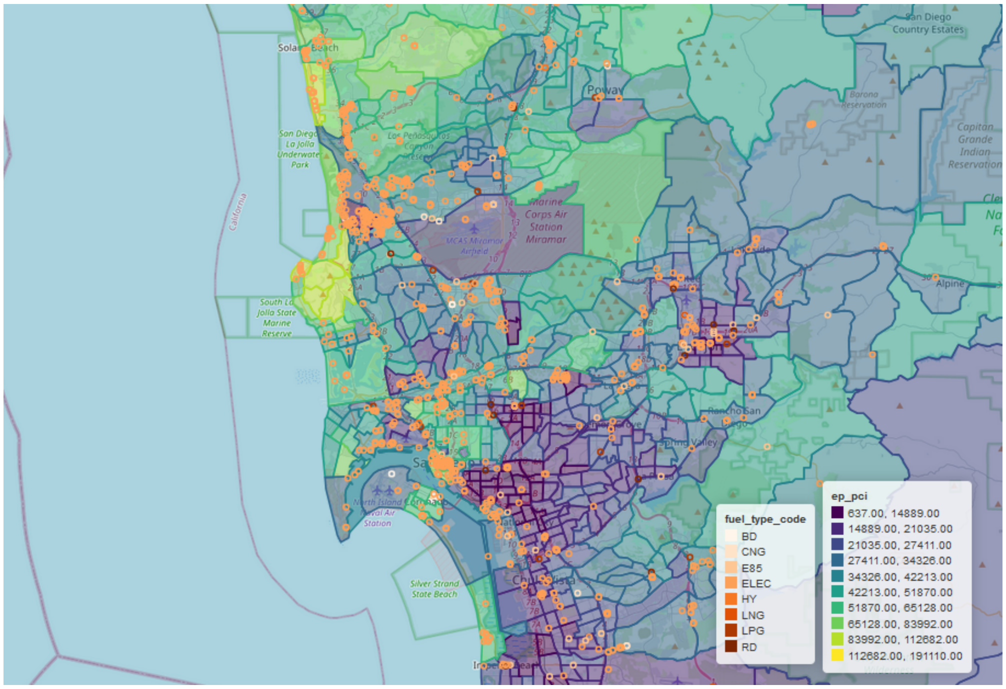

Figure 2: Fuel stations plotted on a cloropleth of per capita income, where lighter means higher income and darker means lower income.

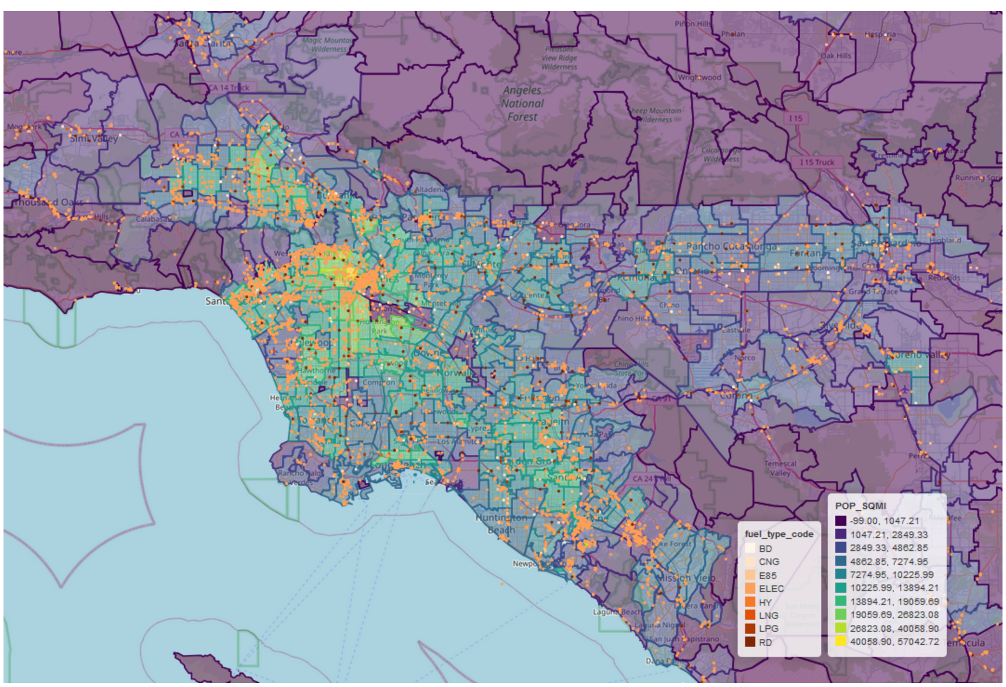

Figure 3: Fuel stations plotted on a cloropleth of population density, where lighter means higher density and darker means lower density.

To further investigate these findings, additional research utilizing data on EV sales by zip code, or combining vehicle registration data in the context of the regions investigated, would be valuable to assess whether EV ownership patterns align with the observed fuel station densities. Regarding EV charger placement relative to power lines, plotting the locations of traditional gasoline pumps could provide insights into whether the power line distribution correlates with areas of high traffic flow, rather than solely infrastructure design. This comparison would help clarify whether the clustering of fuel stations is influenced more by traffic volume than by the underlying energy infrastructure.

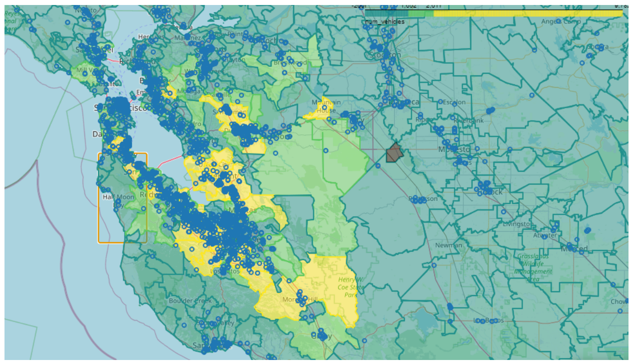

Figure 4: Fuel stations plotted on a cloropleth of ev registrations per zip code in the San Francisco bay area, where lighter means more registrations and darker means fewer.

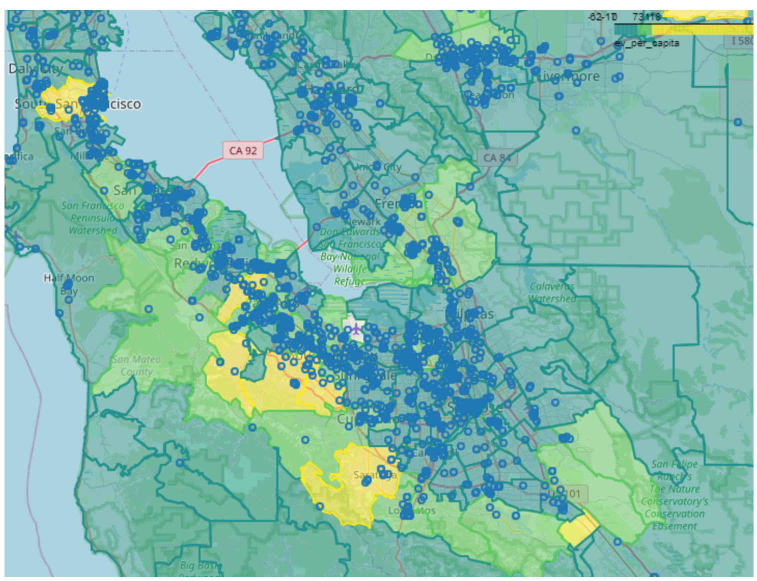

Figure 5: Fuel stations plotted on a cloropleth of ev registrations *per capita* in the San Francisco bay area, where lighter means more registrations and darker means fewer.

By looking at total EV registration numbers from figure 4 and per capita EV registration rates in figure 5 overlaid with charging station data, we can see that regions with higher numbers of registered EVs are indeed within closer proximity to the dense clusters of charging stations. This supports the suggestion made prior that station positioning is likely strongly influenced by the factor of overall EV registrations and traffic in a local area.

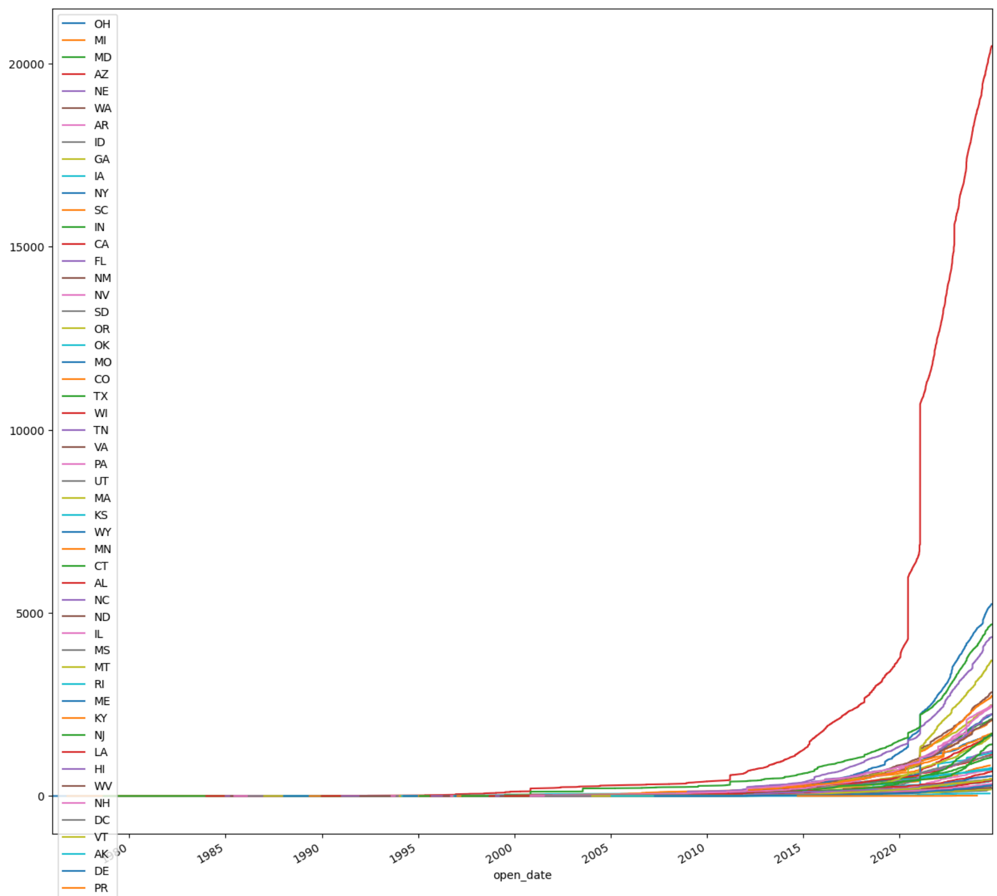

Figure 6: Time series growth of individual state alternative fuel station openings, on a linear scale.

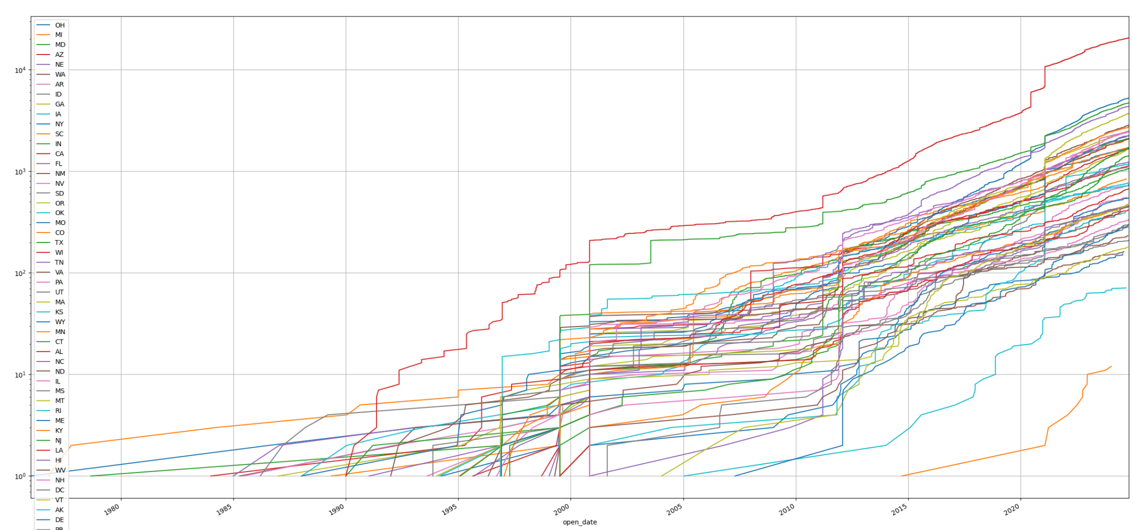

Figure 7: Time series growth of individual state alternative fuel station openings, on a logarithmic scale.

Fig.6 of alternative fuel station counts reveals a general trend of exponential growth across all states, with California leading prominently. However, when the data is plotted on a logarithmic scale as in fig.7, distinct periods of accelerated growth emerge, marked by sharp increases in station numbers across many states. While these periods vary slightly by state, there are several consistent surges, notably in the periods of late 1999–2001, 2005–2006, 2012–2013, and 2021–2022. The following hypotheses are proposed to explain these trends:

- National government policies, such as subsidies and incentives for clean energy infrastructure, may have influenced these growth periods.
- Broader adoption of low-emission vehicles could have increased demand for alternative fuel stations.
- Widespread upgrades to national energy infrastructure might have enabled or facilitated these expansions.

To investigate these hypotheses, further research could involve examining government policy enactments within these time frames, analyzing clean energy vehicle sales trends, or reviewing energy infrastructure upgrade reports.

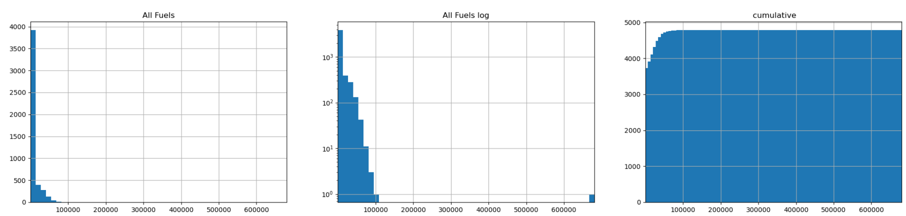

Figure 8: Distribution of all vehicle registrations per zip code on a linear, log and cumulative scale (left to right).

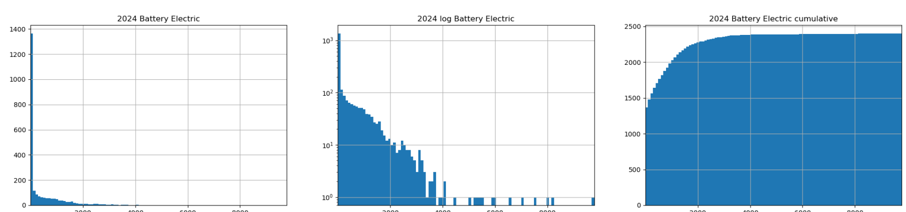

Figure 9: Distribution of EV vehicle registrations per zip code on a linear, log and cumulative scale.

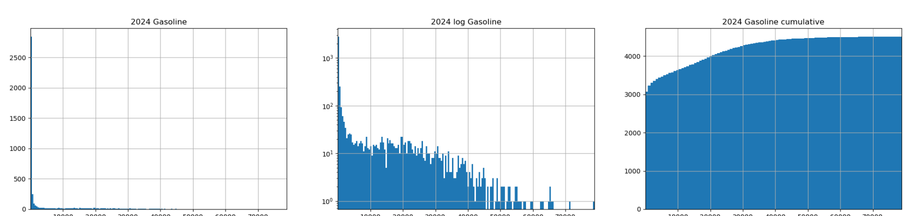

Figure 10: Distribution of ICE vehicle registrations per zip code on a linear, log and cumulative scale.

From fig. 8, 9, and 10, it is evident that overall, vehicles per zip code largely cluster towards lower registrations per zip code, and taper off as registrations increase. However, different trends are exposed when looking at EV and gasoline vehicle registrations. EV registration rates show to have a more extreme clustering towards lower values, as the logarithmic representation in fig. 9 shows a sudden decline in the bins. There also seem to be various zip codes that have unusually high amounts of EVs registered. In fig. 10, we can see via the logarithmic scale that the decrease in high amounts of registration tapers off quickly but not suddenly, and continues to remain at stably decreasing levels compared to EV registrations. This is likely due to the historical dominance of gas-powered motor vehicles, and the continued dominance they have in regards to vehicle representation despite the growing market for EVs.

Figure 11: The path from SDG&E HQ to an EV station at 7544 Girard Ave.

Utilizing the OSMnx package for analyzing and visualizing road networks, is essential for understanding accessibility and path distances within urban environments, and was instrumental in calculating the shortest path distance between SDG&E HQ and an chosen EV charging station at 7544 Girard Ave, shown in fig. 11 This use case can easily be extrapolated to various locations, such as residential areas and commercial zones. By utilizing the capabilities to retrieve, map, and measure route distances, OSMnx has attractive use cases in defining metrics on spatial coverage of EV infrastructure relative to high-demand areas, such as an abstract scalar field representing accessibility to chargers for any given location. Additionally, combining these use cases with datasets from sources such as Cenpy or the DMV vehicle registration dataset allows for the creation of spatial data visualizations that integrate both network and demographic data. By overlaying OSMnx-generated road networks with Cenpy-derived census data—such as population data in fig. 12 or vehicle registration rates such as in fig. 4 and 5 as mentioned above, a user can easily highlight potential socio-economic, adoption rates, or population density related factors influencing EV station placement and accessibility. For instance, mapping income levels alongside path distances to the nearest EV stations could allow for an assessment of whether certain demographic groups have adequate access to EV charging infrastructure, potentially visualized in a format showcased in fig. 13 using a centrality heatmap biased by various parameters. Utilizing these tools in conjunction with cross-sectional data integration and utilization provides a cohesive framework for spatial analysis and visualization, enabling data-driven insights into infrastructure planning via consideration of physical and demographic landscapes. This approach supports targeted infrastructure development and allows for more informed decision-making regarding the equitable distribution of resources.

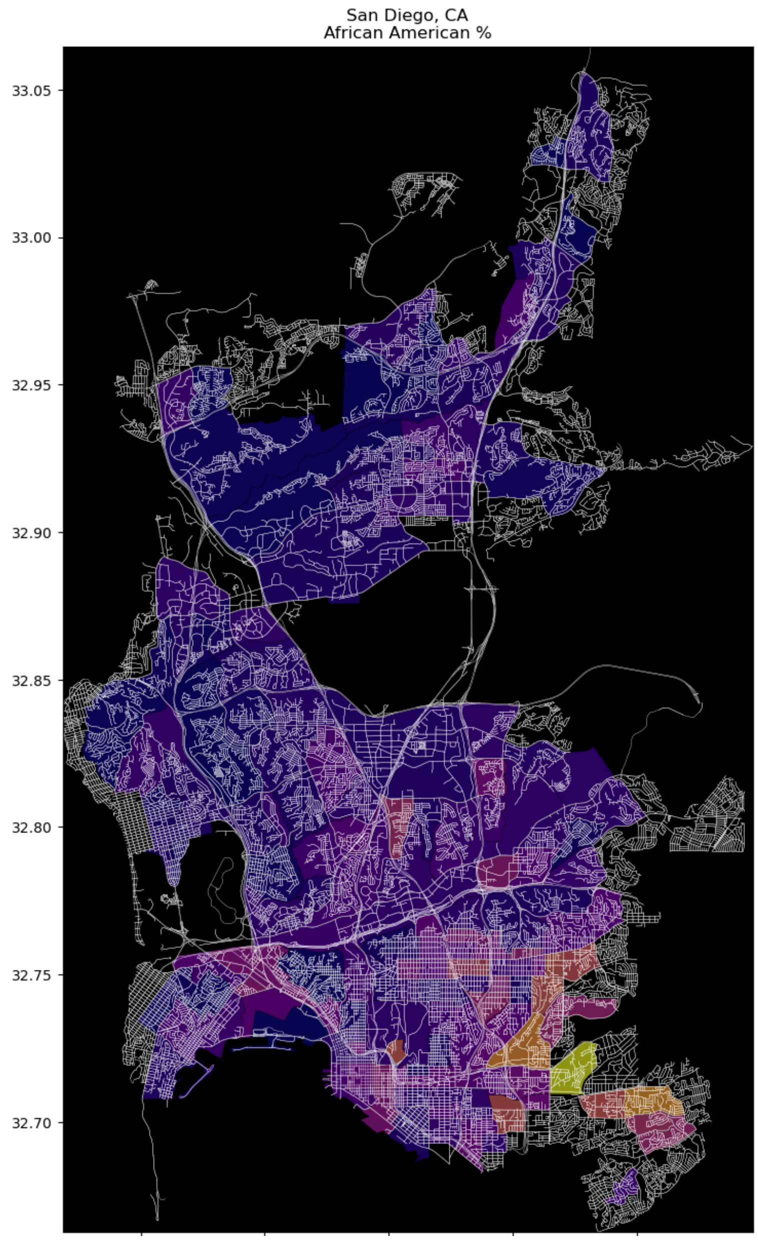

Figure 12: Proportion of a population demographic in San Diego, CA.

Figure 13: Edge centrality heatmap.

5 Conclusion

The exploration of primary datasets and the familiarization of core python packages is paramount to the understanding of EV domain knowledge. As proposed in the discussion section, future insights and relationships between EV and demographic data can be easily integrated to drive future infrastructure related decision making and understanding. Additionally, the usage the topics covered in this project serve as the framework for creating any myriad of future applications or studies - such as creating a charging station app, or in the usage of developing reports for shareholders. Overall, the explorations helped to greatly deepen the domain knowledge and expertise held, and will continue to serve as vectors to approach further investigations.

References

Alternative Fuels Data Center, The National Renewable Energy Laboratory. 2024. "Alternative Fuel Stations." [Link]

California Department of Education. 2024. "California Zip Codes." [Link]

California Department of Motor Vehicles. 2019-2024. "Vehicle Fuel Type Count by Zip Code." [Link]

California Energy Commission. 2024. "California Electric Transmission Lines." [Link]