Distribution of Electrical Vehicle Charging Stations in Bay Area

By Su Yu

Introduction

Since the last few decades, electrical vehicles (EV) have gained an increasing attention and demand as a replacement of traditional fossil fuel-powered vehicles. As of 2018, 45 series production highway-capable all-electric cars are available in various countries, with 300,000 Nissan Leaf and 200,000 Tesla Model S sold across the globe.

While the EVs are rapidly expanding its business, a major drawback of these vehicles is that their utility relis heavily on the availability of electrical vehicle charging stations (EVCS). The sparse distribution of EVCS in some cities is a primary concern that limits the popularity of EVs.

This project studies the distribution of EVCS in the Bay area. This project aims to develop statistical and machine learning methods in analyzing the current distribution of EVCS, and making suggestions to addition of future stations. This project uses geographical data from Foursquare to help develop this analysis.

The rest of this report is organized as follows. Section 2 describes the data with variable definitions and initial data analysis. Section 3 lists the methods our team used to model this dataset. Section 4 presents theresults of model fitting from each method. Section 5 discusses the analytical findings from the applied methodsand summarizes the outcomes of this class project.

Data Collection

In this project, the distribution of EVCS in the Bay area is collected using Foursquare. Union City, CA is selected as the geographical center of search. A total of 39 EVCS are found within 10 kilometers from the center of search, as displayed in Figure 1. A sub-table containing the location of all 39 EVCS is created from Foursquare data, of which its head is shown in Table 1.

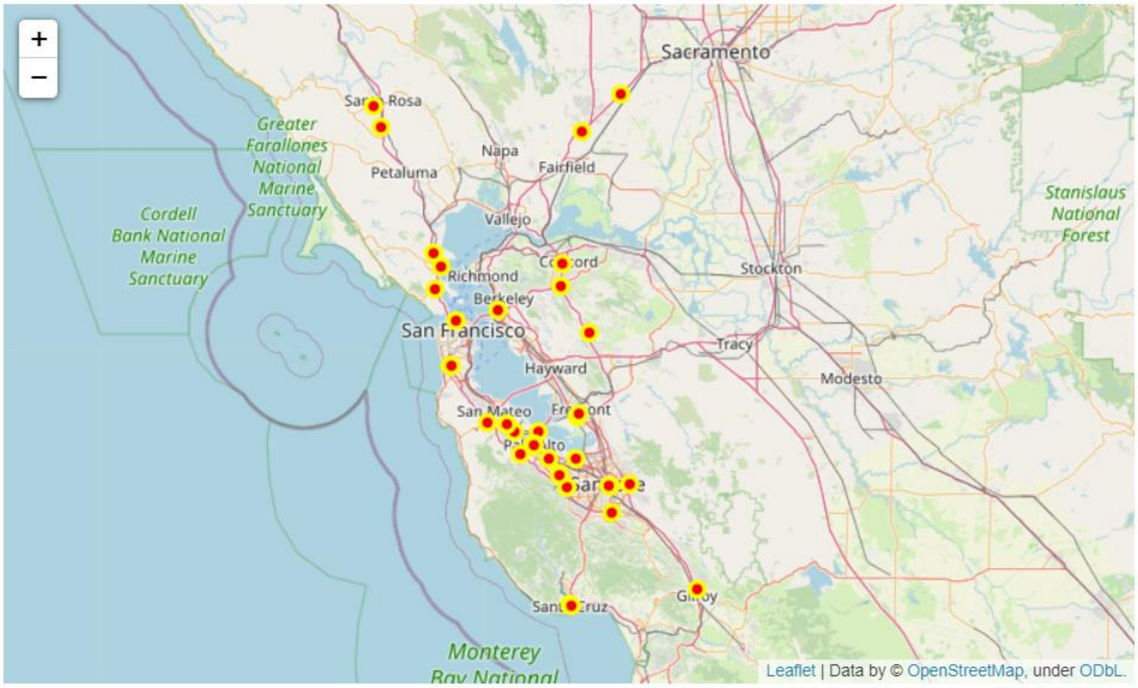

Figure 1. Distribution of EVCS in the Bay Area.

Table 1. Head of dataframe containing the locations of EVCS

	name	categories	lat	lng	City	County
0	Phone Charging Station	Shopping Mall	37.526129	-122.000650	Newark	Alameda
1	Lucky EVgo Charging Station	Automotive Shop	37.534214	-121.996956	Fremont	Alameda
2	Chargepoint Charging Station	EV Charging Station	37.444565	-122.165583	Palo Alto	Santa Clara
3	Charging Station	EV Charging Station	37.485030	-122.146750	Menlo Park	San Mateo
4	Chargepoint Charging Station	EV Charging Station	37.446567	-122.162415	Palo Alto	Santa Clara

The data on population and land areas of cities and counties of bay area is collected from Wikipedia. Python module "BeautifulSoup" is used to scrape the table from Wikipedia page. A sub-table containing the population in 2010 and land areas of each city in bay area is created from the Wikipedia data, of which its head is shown in Table 2.

Table 2. Head of dataframe containing the population (2010) and land areas of each city in CA

	City	County	Population	LandArea[Sqmi]
0	Alameda	Alameda	73812	10.61
1	Albany	Alameda	18539	1.79
2	American Canyon	Napa	19454	4.84
3	Antioch	Contra Costa	102372	28.35
4	Atherton	San Mateo	6914	5.02

The geospatial data on California counties is collected from an open-source github repository. A geojason file containing the boundaries of each county in California is downloaded, and used for map plotting.

In this project, the two sub-tables are merged for clustering and data analysis. The geospatial data is used for choropleth plotting in folium maps.

Methodology

This project uses "folium" map module in Python to visualize the geospatial data in a real map. To address the different demands in EV across counties, the number of EVCS is divided by the population or land area per county. The computed ratios are normalized as Equation 1,

$$r = \frac{n_{EVCS}}{PLA} / \max\left(\frac{n_{EVCS}}{PLA}\right) \qquad \textit{Equation 1}$$

where r is the normalized ratio, n_{EVCS} is the count of EVCS in the county, and PLA is the population or the land area of the county. These computed ratios are also reported in Table 3. Choropleth is then created based on above-computed ratios for better visualization of density distribution of EVCS.

This project uses K-means for clustering of the above collected geographical data of EVCS distribution. The results of clustering are plotted on the folium map.

Results

The collected results and computed ratios are summarized in Table 3. The results here are group by county, because there are too many cities in the Bay Area.

Table 3. EVCS distributions results summarized per county in the Bay Area

	County	Population	LandArea[Sqmi]	EVCS Count	Normalized EVCS-Pop Ratio	Normalized EVCS-Land Ratio
0	Marin	184982	68.02	3	1.000000	0.792947
1	San Mateo	657229	143.83	8	0.750553	1.000000
2	Santa Clara	1691682	352.86	15	0.546740	0.764273
3	Sonoma	348463	83.73	2	0.353901	0.429446
4	Solano	394510	127.16	2	0.312594	0.282774
5	Contra Costa	889240	270.90	3	0.208023	0.199100
6	Alameda	1369005	315.24	3	0.135121	0.171096
7	San Francisco	805235	46.87	1	0.076575	0.383588
8	Napa	110271	31.80	0	0.000000	0.000000

Shaded maps are generated using folium choropleth based on normalized EVCS-population ratio and normalized EVCS-land area ratio in Figure 2 and 3. The shade is generated in a yellow-red scale, with light yellow being scarce in EVCS and dark red being rich in EVCS. In-map clustering of EVCS is shown as blue labels (1 EVCS) or green-yellow circles (multiple EVCS, with counts in center of circle).

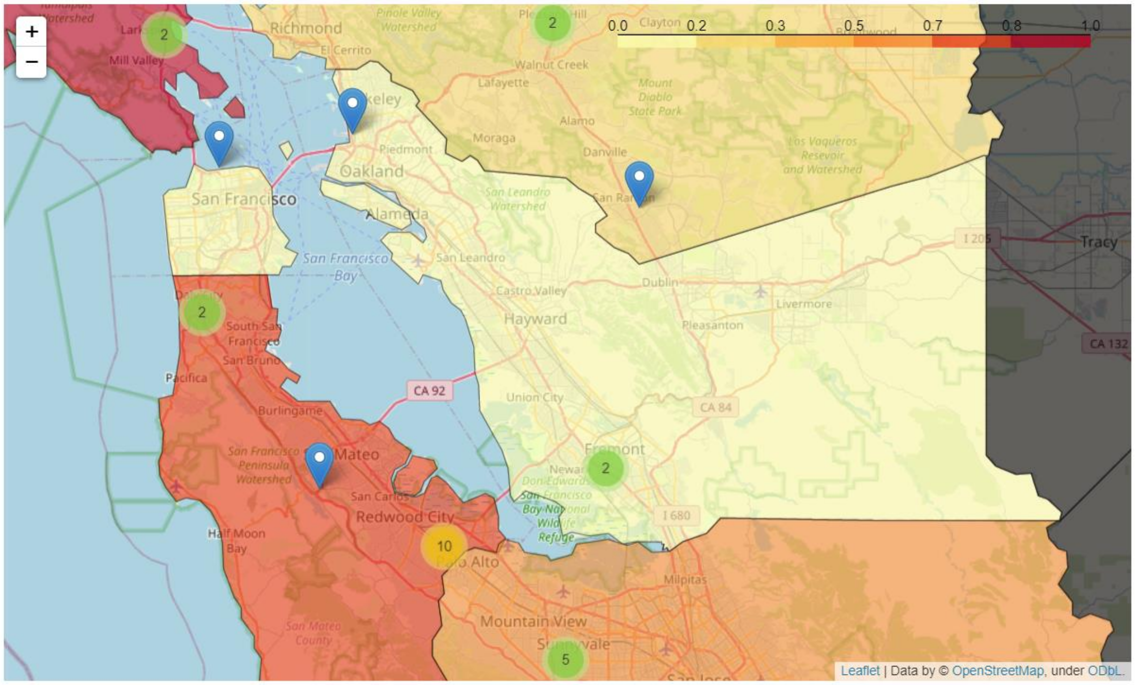

Figure 2. EVCS distribution clustering and normalized density per county based on population in the Bay Area

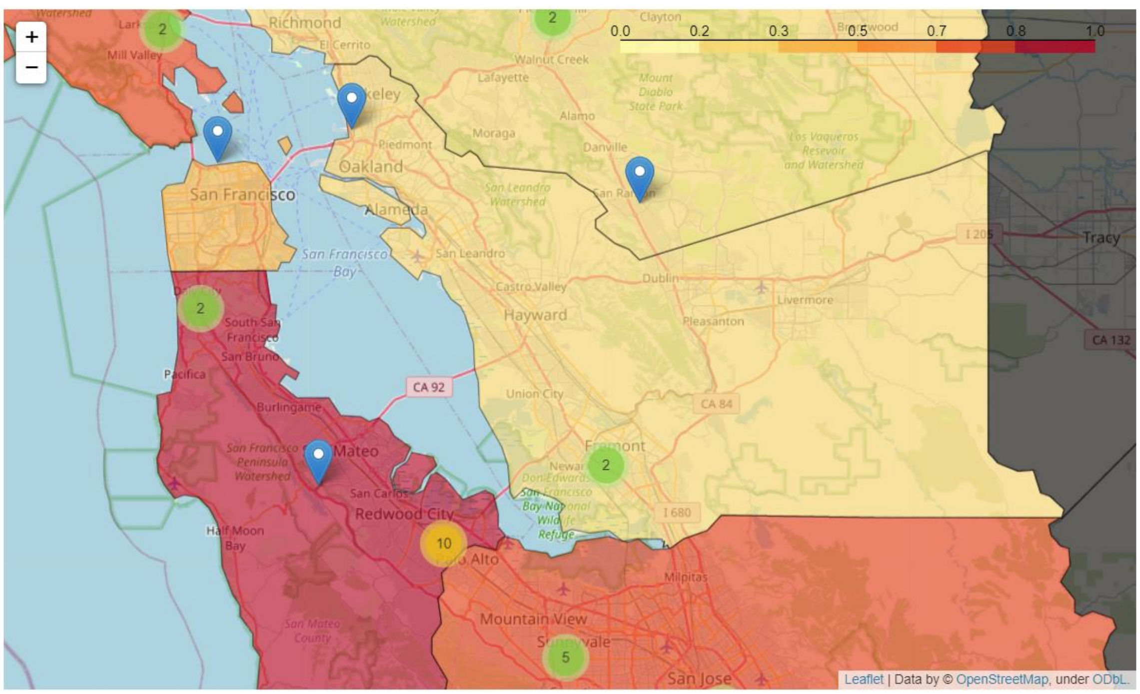

Figure 3. EVCS distribution clustering and normalized density per county based on land area in the Bay Area

Discussion

Generally speaking, the analysis based on population (Figure 2) and land area (Figure 3) are overall consistent, with some minor differences. In the scope of this project, both map plots suggest a need of more EVCS in county of Alameda, San Francisco, and Contra Costa based on the choropleth shading. Considering that people in San Francisco likely relies less on personal vehicles due to public transportations in the city, county of Alameda and Contra Costa would have a greater need for future EVCS.

More specifically, this projects suggest to build future EVCS in region of Hayward, Livermore, and Brentwood/Antioch. Based on clustering and qualitative inspection in Figure 2 and 3, these regions are located in county of Alameda and Contra Cost with no existing EVCS nearby. From data in Table 2, these regions are densely populated cities/towns that would amplify the benefit of a new EVCS. In addition, the Hayward Executive Airport and Livermoore Municipal Airport would also boost the need of a new EVCS in the above described regions. Therefore, Hayward, Livermore, and Brentwood/Antioch should be prioritized in adding future EVCS.

Conclusion

In conclusion, this project analyzes the density of existing EVCS per county based on population, land area, and geospatial distribution. From this analysis, the researcher suggests to add future EVCS in Hayward, Livermore, and Brentwood/Antioch regions.