Machine Learning Approach to EV Charging Infrastructure Planning

X. Liu, E. Luo, Q. Sun, Q. Zhang

SYSEN 6880 Final Project

1 Introduction

The electric vehicle market has already seen a boom in recent years. In the graph below, you can see that U.S. sales of EV almost doubled from 2020 to 2021. Some experts say that the current high gas price will send that demand soaring even more in the near future.

With the growth of EVs, their charging availabilities became critical as well. Taking one of the hottest EVs in the market as an example, Tesla 3 will need between half to a full day to fully charge itself in a residential environment. If time is a constraint, public electric charging could reduce the time needed up to 10 times, given the much higher voltage it provides. If the EV owners are simply away from home, then public charging becomes an inevitable question. Let alone, some EV owners do not even have EV charging capability at home. A strategic solution to the location and scale of public charging planning is required when considering the current electric vehicle revolution.

The construction of an electric vehicle charging station is the premise of the popularity of electric vehicles. Unlike gas stations, EV stations will require more land to service the same amount of vehicles as the minimum charging time to refuel an average electric vehicle is still more than 40 mins. Therefore, it is more important to plan out the locations and scale of EV charging stations in an area to ensure that the charging and energy needs are met in a cost-effective manner.

A local EV community shall be considered as a closed system, whereas the location of each EV is based on its owner's residential location. Given the varieties of EV with longer to shorter ranges. It is necessary to include all scenarios so that all EV owners can travel within the area worry-free. Most importantly, the location of all EVs is required to plan out the ideal locations for the electric charging stations. Furthermore, the number of electric charging ports per charging station could be estimated and predicted in addition to the public EV charging demand growth in the future. EV owners, who live in an optimized state with well-planned EV charging stations, shall be able to travel within the state even when there are no home charging capabilities.

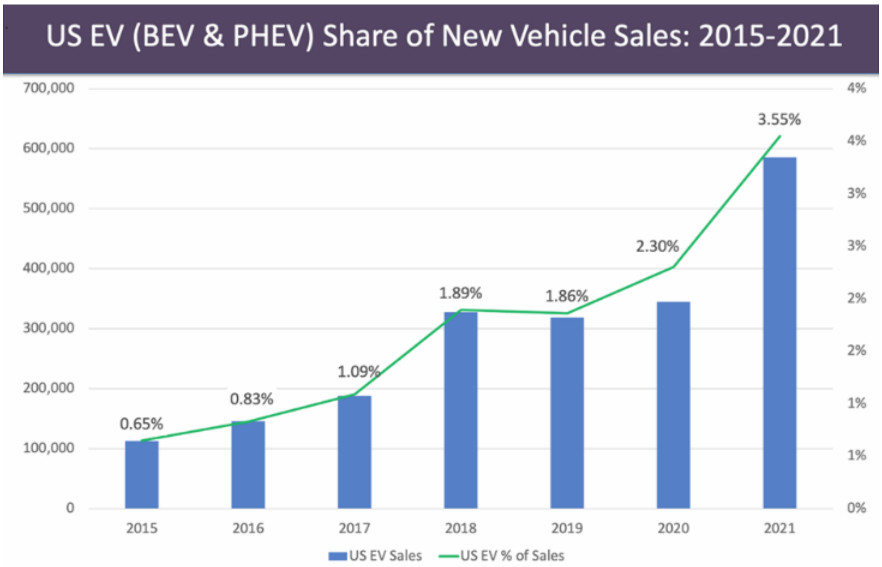


Figure 1: US EV Sale 2015 to 2021 [11]

2 Background

The use of electric vehicles (EVs) is growing in popularity each year, and a considerable increase in demand is expected in the distribution network. One of the main challenges preventing more customers to incorporate EVs into their homes is the difficulty in access to reliable charging stations, especially along highways and interstate roads. This project aims to supply an improved solution to the EV charging experience on the road, especially during longer trips within a state.

T. Donna Chen et al. [1] discussed solutions for EV charging station (CS) locations by supplying behavioral models to predict when and where vehicles are likely to be parked. The article used mixed-integer optimization program to build a regression model that supply key inputs for determining efficient charging station locations. S. Deb et al.[2] supplied a comprehensive review of the machine learning applications used for solving different aspects of charging infrastructure planning.

As discussed above, most of the previous studies on EV charging station planning focus on EV demand, parking demand and electric price. This project will first use the registered EV population and distribution to predict future EV growth and charging station planning. We used 2021 EV population data from Washington State. The data includes each registered EV's location, model, range, price, etc. An example of the dataset we used in this project is shown below.[3] This dataset allows us to analyze the investment, model preference, and frequency of use of electric vehicles in each region of Washington State. This peace information can further be used to predict the distribution and density of EV charging station in each area.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	VIN (1-10)	County	City	State	ZIP Code	Model Yea	Make	Model	Electric Vehi	Clean Altern	Electric Rang	Base MSRP	Legislative D	DOL Vehicle	Vehicle Location
2	KNDCE3LG3L	King	RENTON	WA	98059	2020	KIA	NIRO	Battery Elect	Clean Altern	239	0	11	128569878	POINT (-122.132064 47.494834)
3	KNDCE3LG9K	King	SEATTLE	WA	98108	2019	KIA	NIRO ELECTF	Battery Elect	Clean Altern	239	38500	37	105987073	POINT (-122.31336800000001 47.54411)
4	KNDCC3LG5L	King	RENTON	WA	98057	2020	KIA	NIRO	Battery Elect	Clean Altern	239	0	37	114300301	POINT (-122.215501 47.476576)
5	1G1FY6S00K	Thurston	YELM	WA	98597	2019	CHEVROLET	BOLT	Battery Elect	Clean Altern	238	36620	2	113112303	POINT (-122.558621 46.888349)
6	1N4AZ0CP6F	Chelan	LEAVENWOF	WA	98826	2015	NISSAN	LEAF	Battery Elect	Clean Altern	84	29010	12	330944169	POINT (-120.73040299999998 47.74993200000001)

Figure 2: WA EV Population Data Overview[3]

This project will use the electric vehicle population to determine the location and the number of charging stations corresponding to each county or postcode in Washington State. Two primary data sets will be used in this project.

The first data set is the electric vehicle population S_P,

$$S_P = \{'County', 'City', 'State', 'ZIP\ Code', ..., 'Electric\ Range', 'Vehicle\ Location'\}$$

In the data set S_P, 'County', 'City', 'State', 'ZIP Code', and 'Legislative District' provides the registration location of the electric vehicle, which helps determine the population (amount) of electric vehicles in each county or postcode. To reduce our training time, we plan to do away with these five columns of data set, and instead use the Vehicle Location data set to estimate the top 3 to 5 areas where EV populations are high in the state of Washington, using a 2D-distribution plot. The 'Vehicle Location' data set contains the satellite longitudinal and lateral points of each vehicle, which can be plotted over an x-y grid. Therefore, allowing us to visualize the population density of EVs in the x-y plane that is the state of Washington.

In S_P, 'VIN', 'Model Year', 'Make', and 'Model' represent the electric vehicle's vehicle identification number and brand. 'Electric Vehicle Type' has two types in this dataset, one is Battery Electric Vehicle (BEV), and the other is Plug-in Hybrid Electric Vehicle (PHEV). BEV is Clean Alternative Fuel Vehicle Eligible, and PHEV is Not eligible due to low battery range. 'Electric Range' represents the range of the electric vehicle after fully charging. 'Base MSRP' stands for manufacturer's suggested retail price.

This project will use the vehicle location and range of the electric vehicle to estimate the mean distance from population EV location to nearby highway/interstate locations to install EV charging stations. The population of the electric vehicle and the running range will affect the number of the charging station. In addition, MSRP may correlate with the income in the county. Households in this location may prefer a shorter charging time, which means the supercharging station can be in this county.

$$S_C = \{T_0, T_f, T, E, GHG_s, G_s, P\}$$

3 Methods

3.1 Data Cleaning

We use Washington state electric vehicle population data for this project. We focused on using the vehicle registration location data and electric range data to address where charging stations should be placed. First, we keep "Model Year", "Make", "Model" and "Electric Vehicle Type" to represent the electric vehicle's brand. Inside the "Electric Vehicle Type" column there are two types, one is battery electric vehicle (BEV), and the other is Plug-in Hybrid Electric Vehicle (PHEV). According to the clarification in the dataset, the Plug-in Hybrid Electric Vehicle (PHEV) is not eligible due to the low battery range, and We are only focusing on electric vehicles, not including hybrid vehicles. So we removed Plug-in Hybrid Electric Vehicle (PHEV) from the dataset. Since we are not considering the price of the electric vehicle, we remove the "Base MSRP" in the dataset. Next, we drop the none value in our dataset to avoid outliers. Furthermore, we add the longitudinal and lateral columns based on the registration location to analyze the vehicle location better.

3.2 K-mean Clustering Model

We use K-mean clustering to divide WA based on EV population, and K cluster centroids correspond to N charging stations.

K-means [5] clustering is a complex unsupervised learning algorithm for solving clustering problems. It classifies a given dataset into given K clusters such that the sum of squares within the clusters is minimized. The data set is divided into K mutually exclusive clusters such that the data points within each cluster are as close to each other as possible but as far as possible from the data points in the other clusters.

3.2.1 Elbow Method

In cluster analysis, the Elbow Method is a heuristic method used to determine the number of clusters in a data set. The method consists of plotting explained variance as a function of the number of clusters and selecting the elbow of the curve as the number of clusters used. [6] We use the Elbow method to find the best K for our first clustering at the first step. K=6 is chosen.

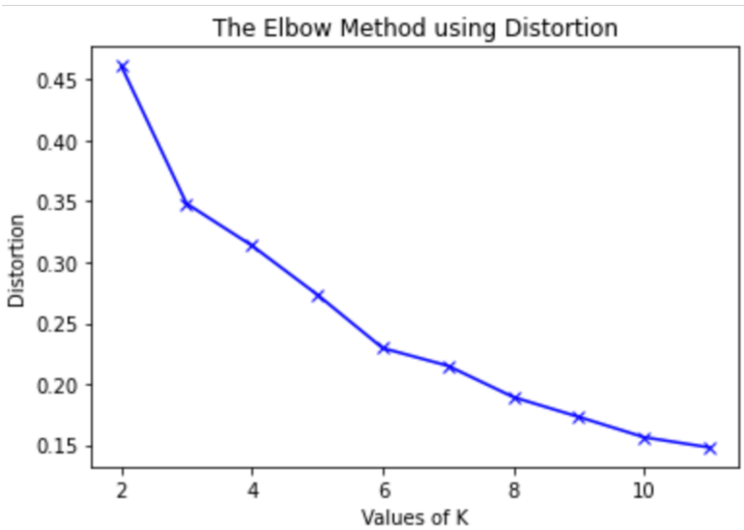


Figure 3: Elbow Method Results for 1st Cluster.

3.2.2 Clustering

After first clustering, we keep implementing the elbow method and K-mean to each small cluster, as shown in the figure on the right, until each Charging station area has no more than 5000 EV and the maximum EV to the nearest Charging Station distance is less than the minimum EV range.

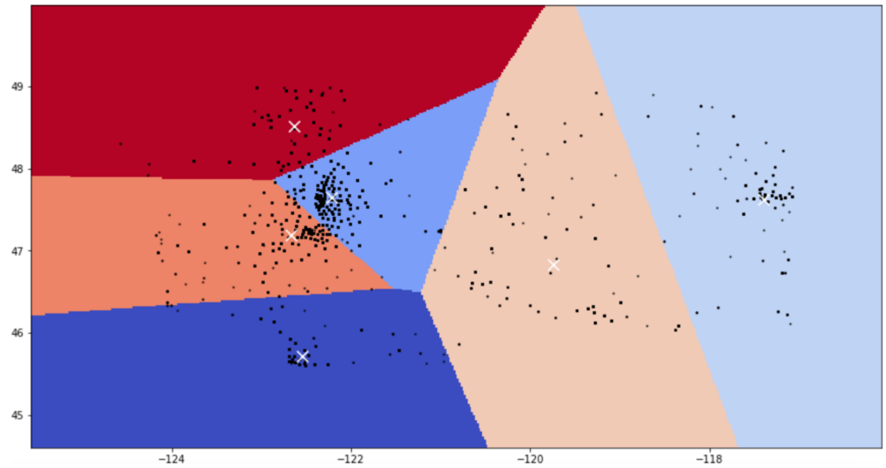


Figure 4: 1st Cluster.

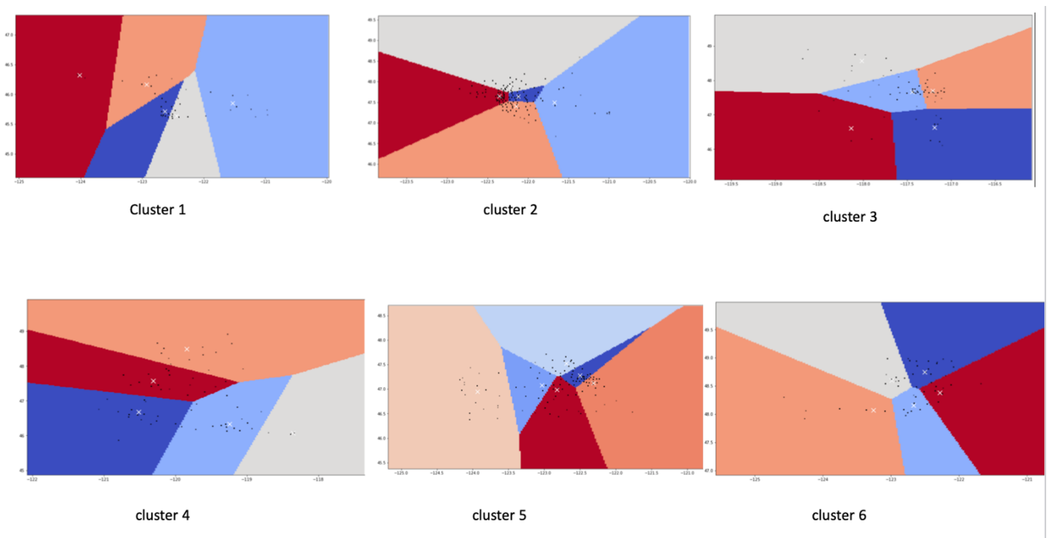


Figure 5: 2nd Cluster.

3.2.3 Charging Station Need Estimation

The number of EV charging ports in the planning area will depend on the number of EVs, the model and range of EVs, and the daily electricity demand of the EVs. The EV number, model and range are given in our dataset. The total electricity demand of an EV per day is denoted by W.[7] W is calculated as below:

$$W = P \times r \times cycles \times N_{EV}$$

Where P is the power consumption of the EV in kWh/mile, r is the battery range of EV in miles, cycles is the average charge cycles per day.N_{EV} is the number of EV in a planning area.

The number of charging ports N_{CP} is calculated below:

$$N_{CP} = \beta \times \frac{W}{P_{CS} \times T_{EV}}$$

Where β is an adjustment constant factor for the theoretical estimation of charging ports need, P_{CS} is the capacity of a Charging Station area, and T_{EV} is the average time needed to fully charge an EV. [8][9]

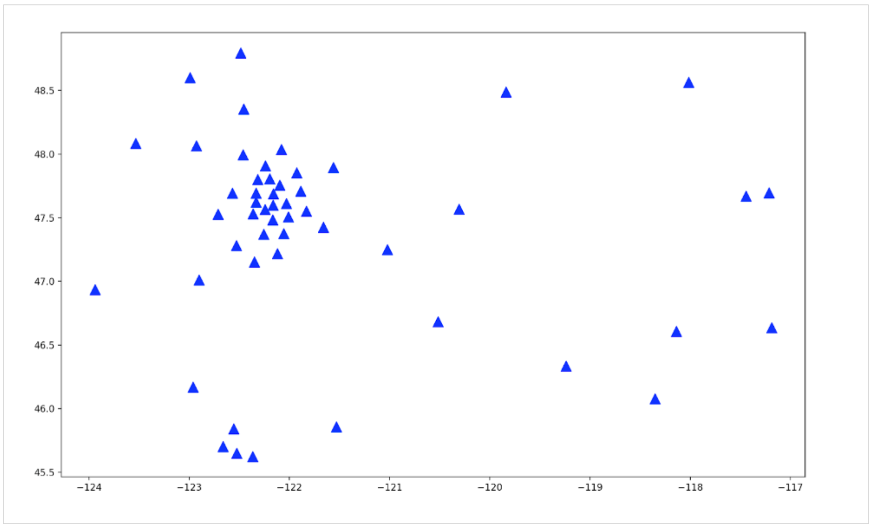


Figure 6: 50 Charging Station Gathering Area in WA

3.2.4 K-mean results

We ended up with 50 charging station gathering areas using k-mean. The location we found for the charging station suggests that it's the most efficient location to install the charging station. It could be multiple charging ports around this area. Figure 6(a) shows the maximum EV to charging station distance is less than the minimum EV range for all clusters, which means every EV is able to drive to the nearest charging station. Figure 6(b) shows every cluster's EV number and charging port number.

[a]

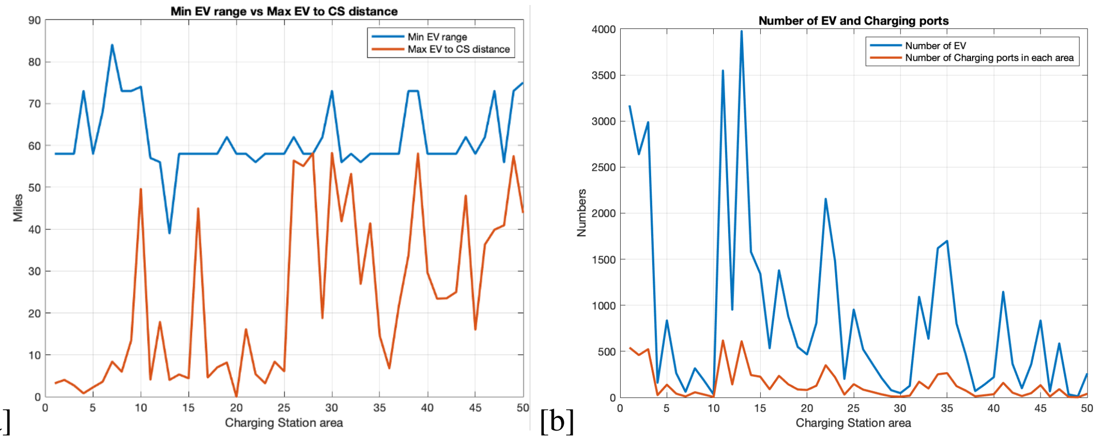


Figure 7: (a) min EV range vs. max EV to CS distance (b) Number of EV vs. Number of Charging Ports

3.3 Gaussian Mixture Model

3.3.1 Silhouette Score

The silhouette score is applied to check whether the number of clusters is the same as applying the elbow method. The silhouette score is to measure the distance between each point to its nearest neighbor points, which provides information about the overall clustering performance. From Figure 7, it shows that the first highest silhouette score at the n=3, and the second highest silhouette score at n=6. It is not precise to apply two clusters only in this model, so dividing the dataset into six clusters is much more persuasive here.

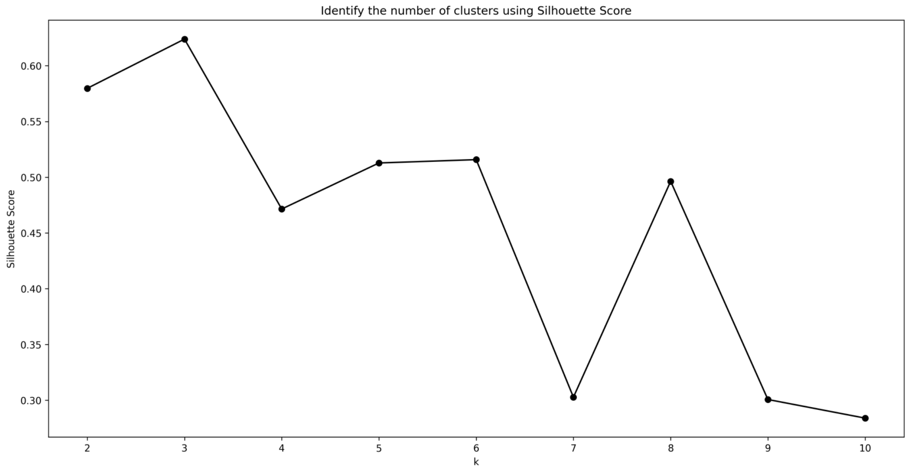


Figure 8: Silhouette Score

3.3.2 Clustering

K-means method is applied in the previous part to determine the location of the charging stations. However, K-means may exist some limitations to cluster the dataset. For example, when a data point is close to more than one cluster centroid, K-Means is hard to handle the uncertainty. In addition, K-means may not handle the non-linear decision boundaries of clustering. Hence, Gaussian Mixture Model is applied to check whether the result is different. The Figure 8 shows the clusters distribution under the Gaussian Mixture Model.

The clustering from the Gaussian Mixture Model is significantly different from the output from K-means. To better compare the result, the mean location of 6 clusters is shown in Figure 9(a). When comparing the center location, the result is very close to the K-means.

3.3.3 Gaussian Mixture Model result

To better determine the EV charging location, the Gaussian Mixture model is applied to cluster 5 EV charging locations in each cluster shown in Figure 8. The total number of 35 EV charging locations are shown in Figure 9(b).

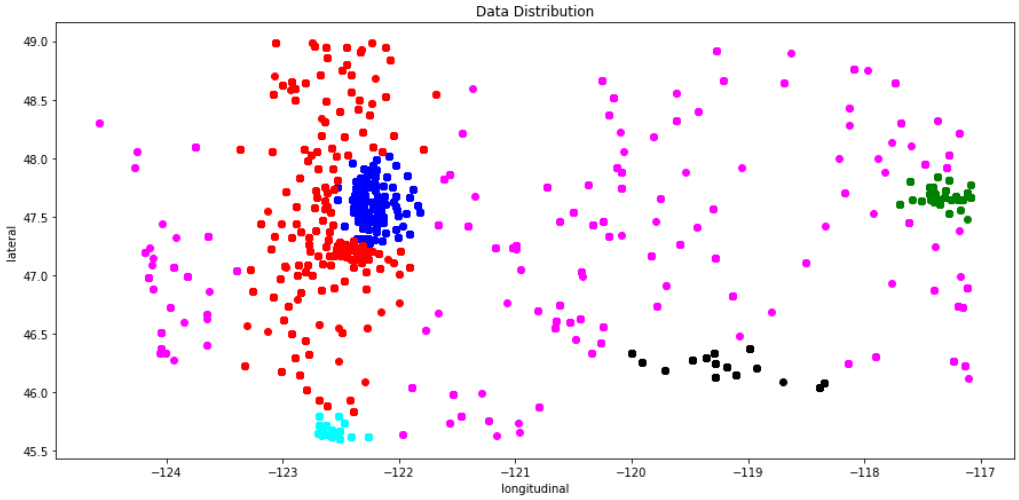
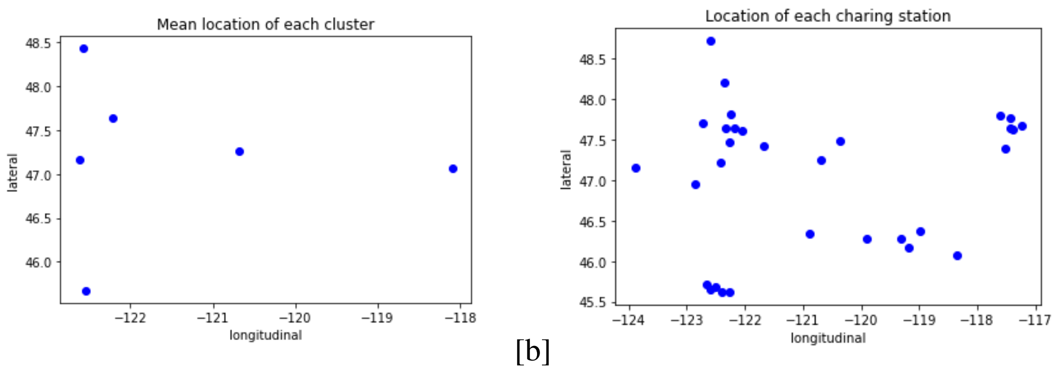


Figure 9: Gaussian Mixture Model cluster distribution of the dataset.



[a] [b]

Figure 10: (a) Mean location of each cluster (b) Location of EV charging station

4 Conclusion

Our results in this project can be used to guide in the making of better EV charging station infrastructure planning decisions when many charging stations (in the order of hundreds or more) need to be placed for charging at different locations. We used the clustering Machine Learning method, which mainly focused on K-mean clustering to locate charging stations. This method allows us to locate the charging station of a planning area in an efficient manner, and prepare for the future by using estimated charging station needs with the increasing EV population. On the contrary, potential EV owners could use this EV charging station optimization method to their benefit when deciding the types of EV to purchase.

5 Reference

[1] Chen, T. D., Kockelman, K. M., amp; Khan, M. (2013). Locating Electric Vehicle Charging stations. Transportation Research Record: Journal of the Transportation Research Board, 2385(1), 28–36. https://doi.org/10.3141/2385-04
[2] Deb, S. (2021). Machine learning for solving charging infrastructure planning problems: A comprehensive review. Energies, 14(23), 7833. https://doi.org/10.3390/en14237833
[3] Dusi, V. K. (2020, November 12). Evpopulation. Kaggle. Retrieved April 10, 2022, from https://www.kaggle.com/datasets/vijayakishoredusi/evpopulation?resource=download
[4]Sheldon-Zhang. (n.d.). Sheldon-Zhang/Gatech_ev_analytics. GitHub. Retrieved April 13, 2022, from https://github.com/Sheldon-Zhang/Gatech_EV_Analytics
[5] March 2014; pp. 1–8.42. Lloyd, S. Least squares quantization in PCM. IEEE Trans. Inf. Theory 1982, 28, 129–137.
[6] Robert L. Thorndike (December 1953). ”Who Belongs in the Family?”. Psychometrika. 18 (4): 267–276. doi:10.1007/BF02289263.
[7] Shukla, A.; Verma, K.; Kumar, R. Consumer perspective based placement of electric vehicle charging stations by clustering techniques. In Proceedings of the 2016 National Power Systems Conference (NPSC), Bhubaneswar, India, 19–21 December 2016; pp. 1–6.
[8] Shukla, A.; Verma, K.; Kumar, R. Consumer perspective based placement of electric vehicle charging stations by clustering techniques. In Proceedings of the 2016 National Power Systems Conference (NPSC), Bhubaneswar, India, 19–21 December 2016; pp. 1–6.
[9] Varga, B.O.; Sagoian, A.; Mariasiu, F. Prediction of Electric Vehicle Range: A Comprehensive Review of Current Issues and Challenges. Energies 2019, 12, 946.
[10] Patel, E., Kushwaha, D. S. (2020). Clustering cloud workloads: K-means vs gaussian mixture model. Procedia Computer Science, 171, 158–167. https://doi.org/10.1016/j.procs.2020.04.017
[11] B. L. McDonald, “Forecast: 2021 US EV sales to increase 70 year over year,” CleanTechnica, 30-Oct-2020. [Online]. Available:https://cleantechnica.com/2020/10/30/forecast-2021-us-ev-sales-to-increase-70-year-over-year/. [Accessed: 14-May-2022]