Segmenting and Clustering Tesla Superchargers in the United States

Derek Carter, April 3, 2020

Introduction

Electric cars are the future of ground transportation. Aside from being better for the environment, they are faster, quieter, less expensive to power, accelerate more smoothly, and are more reliable than their internal combustion engine (ICE) counterparts. There is, however, a nontrivial drawback to electric vehicles (EVs): charging time. This is a non-issue for daily commuting because charging can occur overnight at the owner's home. Charging time can be an issue for trips requiring a charge before returning home though. For example, a 700-mile trip that took me 11 hours in my ICE vehicle now takes 12.5 hours in my EV, due to the extra time it takes to charge the EV as compared to filling up the ICE vehicle's gas tank.

Taking a road trip in an EV requires a mindset shift. If completing the road trip as fast as possible is paramount, then the EV will probably leave you frustrated. If you don't mind spending a little more time traveling and are able to leverage the charging breaks as an opportunity to stretch out, grab a snack or beverage, have lunch, chat with fellow EV travelers, or check out some of the nearby shops or attractions, then you will likely find road trips more enjoyable and reach your destination more relaxed in an EV than in an ICE vehicle.

In order to mitigate the inconvenience of charging time and further encourage the adoption of electric vehicles, this project will categorize fast charging stations according to nearby locations (e.g., fast food, coffee, groceries, park, …). Such a categorization will be useful to those planning a road trip, enabling selection of routes or of charging stations along the route which have amenities of interest. Further, such a categorization may be useful to entrepreneurs interested in opening a business near chargers which do not already have a similar business nearby.

Data

DC fast chargers include CCS, CHAdeMO, and Tesla Superchargers. Level 2 chargers are slower chargers, typically found in the home or at other final destinations (such as a hotel or shopping mall). A Level 2 charger typically provides 20-40 miles of range per hour of charging. This is ample for an overnight charge, but far too slow when on a road trip with the primary objective of charging the vehicle as quickly as possible. DC fast chargers, on the other hand, are able to charge up to 80% of the battery's capacity in roughly 30 minutes and would be the only feasible option for charging between the origin and final destination on a road trip [1]. Because the intended beneficiaries of this project are EV drivers on a road trip, or potentially entrepreneurs wishing to cater to them, the data will be limited to DC fast chargers. In order to keep the project to a reasonable scope, data will be further limited to Tesla Superchargers within the United States.

Charging Station Data

I considered the following options for obtaining charging stations data:

- National Renewable Energy Laboratory (NREL) Developer Network Alternative Fuel Stations API [2][3]
- Open Charge Map API [4]
- ChargeHub API [5]
- PlugShare API [6]

NREL and Open Charge Map are free, whereas ChargeHub and PlugShare are not, so I eliminated ChargeHub and PlugShare. Ultimately I chose NREL because it is the source of Open Charge Map data for the US and Canada and this project is limited to charging stations in the US.

The NREL API returns a total of 63 features per charging station (e.g., id, cards_accepted, open_date, station_name, station_phone, …). For this project, I will use only the following 8 features for each station (shown below with sample data from one station):

```
'id': 101972
'station_name': 'FAIRFIELD INN - Tesla Supercharger'
'latitude': 34.785416
'longitude': -86.942864
'city': 'Athens'
'state': 'AL'
'street_address': '21282 Athens-Limestone Blvd.'
'zip': '35613'
```

Location Data

The source of the location data will be the Foursquare API [7]. The Foursquare API returns a total of 27 features per venue (e.g., id, name, city, state, …). For this project, I will use only the following 4 features for each venue (shown below with sample data from one venue):

```
'name': 'Fairfield Inn by Marriott Athens'
'lat': 34.78587188328971
'lng': -86.9428607460327
('categories')'name': 'Hotel'
```

Methodology

Using the NREL API, I obtained a JSON-formatted list of all US Tesla Supercharger stations. I then created a DataFrame with only the 8 aforementioned features of interest. Following is a sample of the first 5 charging stations:

	id	station_name	latitude	longitude	street_address	city	state	zip
0	101972	FAIRFIELD INN - Tesla Supercharger	34.785416	-86.942864	21282 Athens-Limestone Blvd.	Athens	AL	35613
1	101973	Auburn Mall - Tesla Supercharger	32.627837	-85.445105	1627 Opelika Road	Auburn	AL	36830
2	101974	Uptown Entertainment District - Tesla Supercha...	33.525826	-86.807072	2221 Richard Arrington Junior Blvd.	Birmingham	AL	35203
3	101975	Hampton Inn Greenville - Tesla Supercharger	31.855989	-86.635765	219 Interstate Drive	Greenville	AL	36037
4	101976	The Bel Air Mall - Tesla Supercharger	30.671556	-88.118644	3201 Airport Blvd	Mobile	AL	36606

Using the Foursquare API, I obtained a list of all venues within 400 meters of each charging station. Following is a sample of the first 5 nearby venues:

	Station ID	Station Latitude	Station Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	101972	34.785416	-86.942864	Fairfield Inn by Marriott Athens	34.785872	-86.942861	Hotel
1	101972	34.785416	-86.942864	Starbucks	34.784197	-86.945789	Coffee Shop
2	101972	34.785416	-86.942864	Chick-fil-A	34.785090	-86.945393	Fast Food Restaurant
3	101972	34.785416	-86.942864	Panera Bread	34.784683	-86.943670	Bakery
4	101972	34.785416	-86.942864	Bojangles' Famous Chicken 'n Biscuits	34.783776	-86.944886	Fast Food Restaurant

One hot encoding was used to create a DataFrame suitable to k-means clustering. In particular, the DataFrame contains one column per venue category, in addition to the Station ID column, and one row per charging station. Each venue category column contains the percentage of all the venues within 400 meters of the charging station in question which are of that venue category. Following is a sample of the first 5 rows and the first few columns:

	Station ID	ATM	Accessories Store	Adult Boutique	Afghan Restaurant	African Restaurant	Airport	Airport Service	Airport Terminal	American Restaurant	...
0	101972	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	0.0	0.235294	...
1	101973	0.0	0.031250	0.0	0.0	0.0	0.0	0.0	0.0	0.031250	...
2	101974	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	...
3	101975	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	0.0	0.250000	...
4	101976	0.0	0.035714	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	...

A new DataFrame was then created, containing columns for the first through 10th most common venues near each charging station. Following is a sample of the first 5 rows:

	Station ID	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	101972	Fast Food Restaurant	American Restaurant	Hotel	Intersection	Ice Cream Shop	Breakfast Spot	Coffee Shop	Pizza Place	Bakery	Food Stand
1	101973	Department Store	Clothing Store	Women's Store	Lingerie Store	Mobile Phone Shop	Discount Store	Fast Food Restaurant	Shopping Mall	Seafood Restaurant	Chinese Restaurant
2	101974	Hotel	Pizza Place	Coffee Shop	Gym / Fitness Center	Brazilian Restaurant	Pub	Fried Chicken Joint	Museum	Burger Joint	Mexican Restaurant
3	101975	Hotel	Gym / Fitness Center	American Restaurant	Mobile Phone Shop	Flea Market	Falafel Restaurant	Farm	Farmers Market	Fast Food Restaurant	Filipino Restaurant
4	101976	Clothing Store	Shoe Store	Department Store	Bakery	Lingerie Store	Kids Store	Discount Store	Cosmetics Shop	Convenience Store	Fast Food Restaurant

Next k-means clustering was used on this DataFrame in order to group the stations into clusters based on the most common venue types near that station. First, though, the optimal number of clusters to use with k-means was determined. For this, we used the elbow method.

After assigning each of the stations to a cluster, we then merge the clustering DataFrame with the original DataFrame so that in addition to the clustering and 10 most common venues for each station, we also have the station name, latitude, and longitude. We use this information to create a map to visualize the station clusters. Following is a sample of the first 5 rows of the merged table, along with the first several columns:

	id	station_name	latitude	longitude	street_address	city	state	zip	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue
0	101972	FAIRFIELD INN - Tesla Supercharger	34.785416	-86.942864	21282 Athens-Limestone Blvd.	Athens	AL	35613	1	Fast Food Restaurant	American Restaurant	Hotel	Intersection
1	101973	Auburn Mall - Tesla Supercharger	32.627837	-85.445105	1627 Opelika Road	Auburn	AL	36830	3	Department Store	Clothing Store	Women's Store	Lingerie Store
2	101974	Uptown Entertainment District - Tesla Supercha...	33.525826	-86.807072	2221 Richard Arrington Junior Blvd.	Birmingham	AL	35203	0	Hotel	Pizza Place	Coffee Shop	Gym / Fitness Center
3	101975	Hampton Inn Greenville - Tesla Supercharger	31.855989	-86.635765	219 Interstate Drive	Greenville	AL	36037	2	Hotel	Gym / Fitness Center	American Restaurant	Mobile Phone Shop
4	101976	The Bel Air Mall - Tesla Supercharger	30.671556	-88.118644	3201 Airport Blvd	Mobile	AL	36606	3	Clothing Store	Shoe Store	Department Store	Bakery

After creation of the map, we list the members of each cluster along with their 10 most common venue types. After analyzing each cluster to determine a pattern among the 10 most common nearby venue types, a bar chart is created to illustrate cluster sizes, colorizing each bar according the same corresponding cluster in the map and labeling each bar with the results of the cluster analysis.

Results

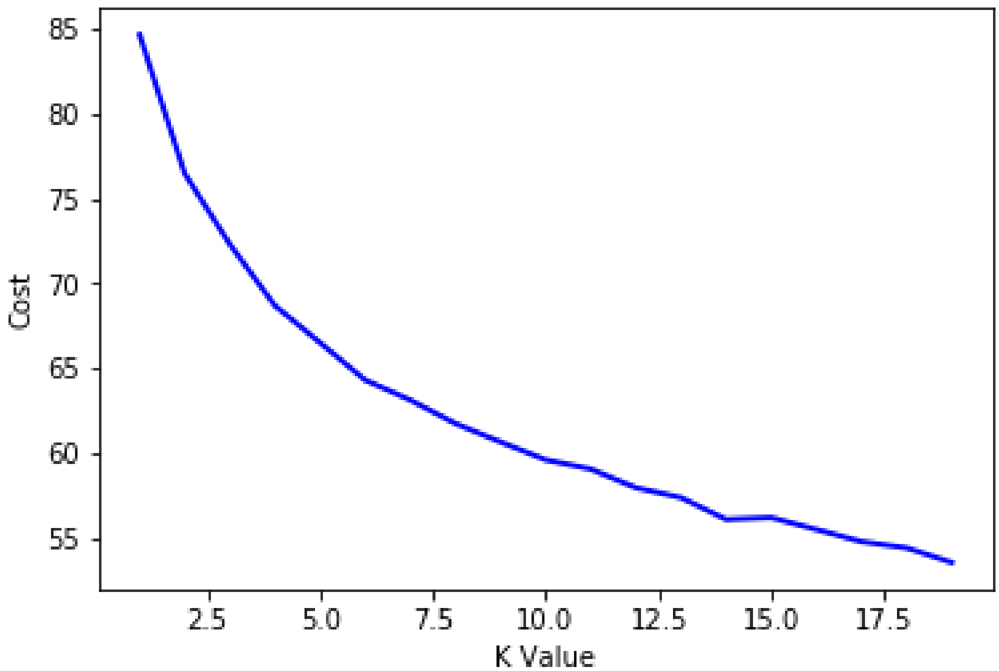
The NREL API returned 800 Tesla Supercharger stations in the United States. Calling the Foursquare API with each of those 800 charging stations yielded 19,562 venues within 400 meters of them. Among those 19,562 venues, there were 446 unique categories. Note that only 798 of the 800 charging stations had one or more venues within 400 meters, so 2 of the charging stations had no venues nearby. In particular:

	id	station_name	latitude	longitude	street_address	city	state	zip
554	122243	5R Travel Center - Tesla Supercharger	32.28138	-107.760286	1695 US-180 E351	Deming	NM	88030
718	150232	Irving 24 - Tesla Supercharger	45.60954	-68.522863	1941 Medway Road	Medway	ME	04460

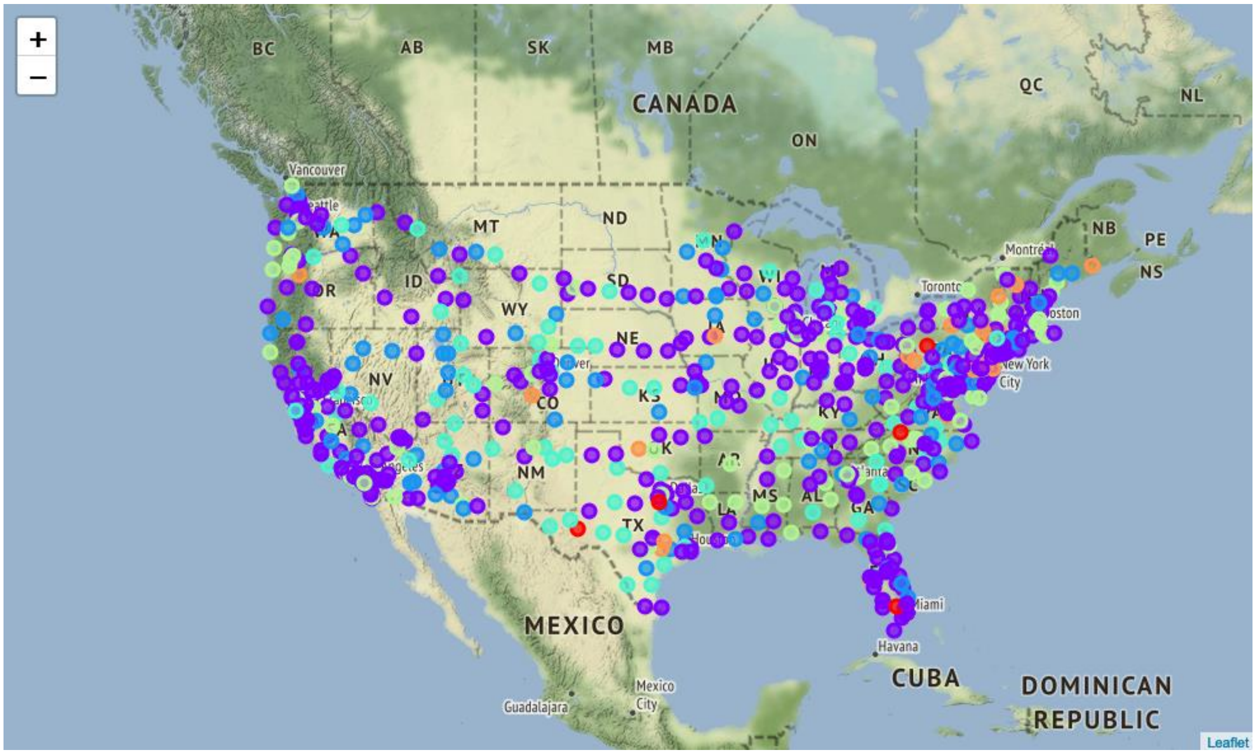
I expanded the search to a radius of 800 meters for the two stations which had no venues within 400 meters. This resulted in a total of only four venues being returned for these two stations. Given there are so few venues near these two stations and none of them are within 400 meters, they will be excluded from the analysis. Note that the 400-meter threshold is relevant because beyond that it becomes less feasible to walk to the venue, spend some time there, and walk back during the time it takes for the vehicle to complete charging. For completeness, following are the four venues between 400 and 800 meters from the two charging stations above:

	Station ID	Station Latitude	Station Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	122243	32.28138	-107.760286	Blake's Lotaburger	32.274692	-107.760069	Burger Joint
1	150232	45.60954	-68.522863	Gateway Inn	45.612364	-68.528674	Motel
2	150232	45.60954	-68.522863	Renally's Services, LLC. The Floor Safety Spec...	45.605643	-68.528651	Business Service
3	150232	45.60954	-68.522863	Circle K	45.612827	-68.531062	Convenience Store

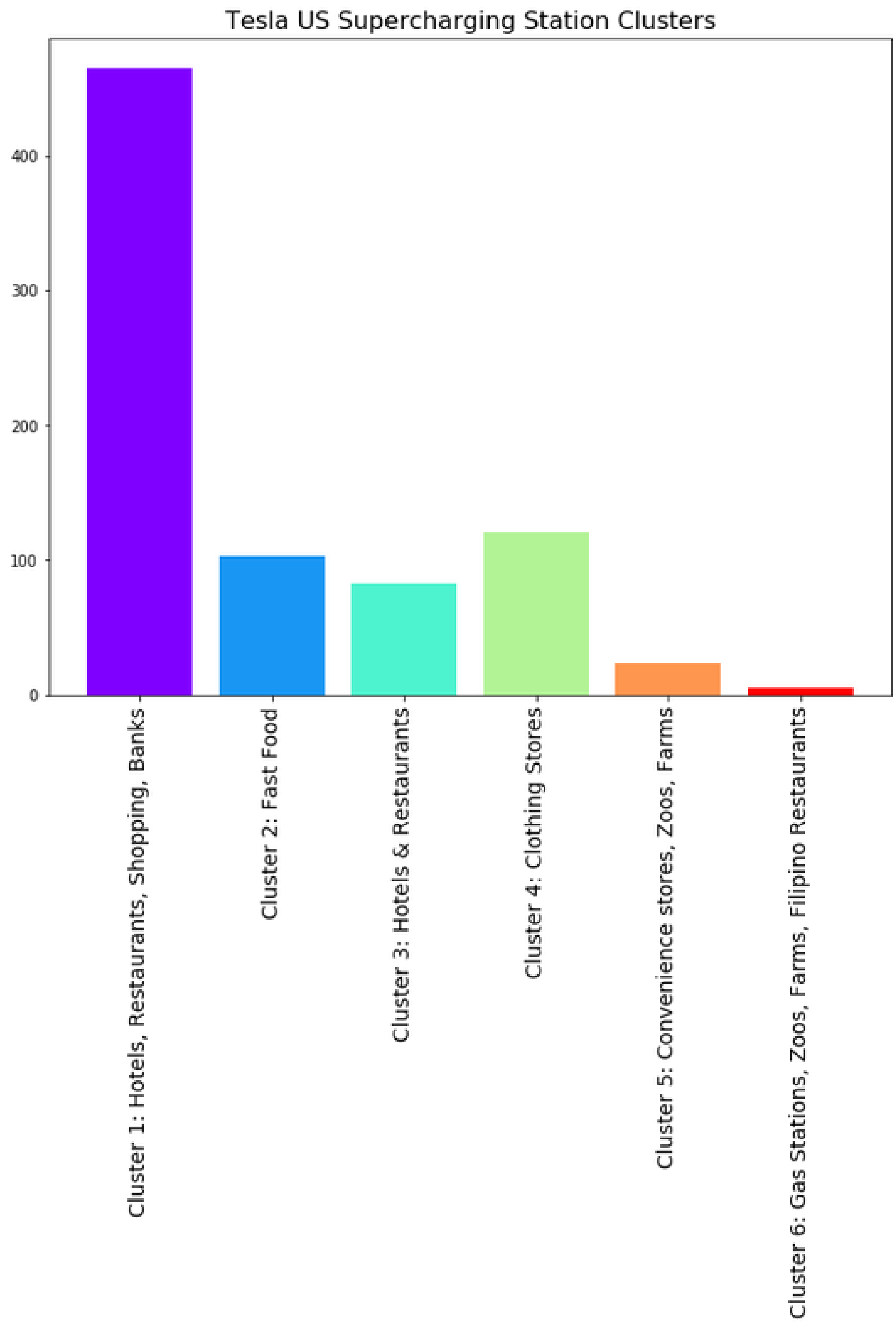
The elbow method did not yield an obvious elbow for determining the optimal number of k-means clusters. That said, the absolute slope (or rate of decrease in cost) does start to trend lower at a K value of 6, so I grouped the charging stations into 6 clusters.



The map shows each of the 798 charging stations represented as a filled circle colorized according to the cluster to which it belongs. Clicking on a charging station circle will show the charging station name, the cluster to which it belongs, as well as its address.



Analysis of the 6 clusters resulted in the following cluster descriptions (x-axis titles). Bar colors correspond to cluster circles on the map, while bar lengths indicate number of members within cluster.



Note that the cluster descriptions are an indication of the predominant venue types near charging stations within that cluster; i.e., they are by no means an exhaustive list of all venue types near charging stations within that cluster.

Discussion

With the above clustering, a route can indeed be selected which includes charging stations within the cluster of interest, at least to a certain extent. Further, with long range batteries, it is often not necessary to charge at every charging station along the route. As such, charging stations along a selected route can even be chosen based on membership within the cluster of interest. For example, if I am traveling from Oklahoma City, Oklahoma to Waco, Texas and I notice that the Tesla SuperCharger in Alvarado, Texas is in Cluster 6, which has a lot of Filipino restaurants, I may plan my route so that I charge at that charging station rather than the also-en route Denton charging station, which is in Cluster 1, thereby affording me the opportunity to expand my culinary experience to include Filipino food.

Charging station cluster membership may even be used as a factor in deciding where to go on a weekend getaway. For example, an Elmira, New York resident grappling with whether to visit New York City or the state capital (Albany), may lean toward NYC if he also wanted to visit a zoo en route because the route to NYC includes 2 charging stations that are in Cluster 5, members of which often have zoos within walking distance; whereas the route to Albany has no such charging stations.

The recommendation would be, when planning a trip which requires supercharging along the way, observe the charging station clusters map to determine which route to take, depending on whether the charging stations are members of a cluster of interest. Similarly, the charging station clusters map can be consulted to determine which charging stations to stop at along the selected route.

Conclusion

I have provided a couple examples illustrating the potential value of this charging station cluster information, but there are undoubtedly additional use cases. For example, if caravanning with a group who are driving an ICE vehicle, a route with charging stations in Cluster 6, which contains a predominance of gas stations within walking distance of its charging station, would be convenient. Further, an entrepreneur wishing to sell merchandise catering to electric vehicle owners (e.g., Tesla t-shirt, coffee mug with EV slogan, …) may wish to approach convenience stores near Cluster 5, which has charging stations with numerous convenience stores nearby.

Note that I only included Tesla superchargers in this project. For further flexibility in selection of routes or selection of charging stations along a given route, this project could be expanded to also include non-Tesla DC fast chargers.

References

- [1] EV Charging Stations: https://www.caranddriver.com/news/a30031153/ev-charging-guide/
- [2] Alternative Fuels Data Center: https://afdc.energy.gov/
- [3] NREL Developer Network API: https://developer.nrel.gov/docs/transportation/alt-fuel-stations-v1/all/
- [4] Open Charge Map API: https://openchargemap.org/site/develop
- [5] ChargeHub API: https://www.mogiletech.com/apiaccess
- [6] PlugShare API: https://recargo.freshdesk.com/support/solutions/articles/29000015750-plugshare-charging-stations-api-documentation-access
- [7] Foursquare API: https://developer.foursquare.com/