Capstone Project: The Battle of Neighbourhoods in Puebla City: Electric Chargers

24-June-2021

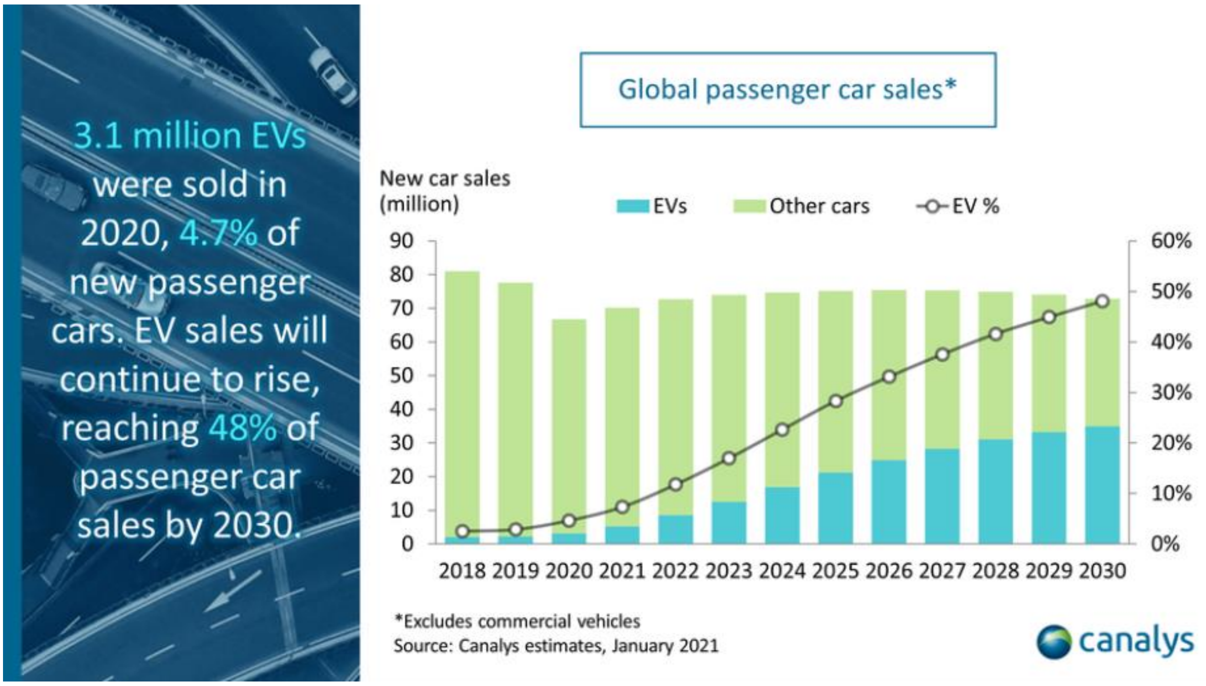
Diego Raul Martinez Velarde, Mechatronic Eng.

https://www.linkedin.com/in/diegormtzvelarde/
https://github.com/DiegoRaulMtz/Coursera_Capstone

1. Introduction

The electric car sales will be increased in the following years. As a following effect, for every place a car can reach, electric chargers will be required.

Nowadays just few brands from electrics cars can be seen on the streets. In Mexico, electric cars market is just starting. Currently is quite strange to see those cars unless you are located in wealthy neighbourhoods. On the other side, there are already electric cars production projects in most of OEM's. Therefore, it is a fact, that electric cars will be mass-produced.



Source: http://canalys-com-public-prod.s3.eu-west-2.amazonaws.com/cosi/campaign/1935/vKlFQ43feFy9cC_VMpHFXVQkJ0rid7A9.png

We know this is a huge challenge for every government indeed. However, we do not expect to update infrastructure for electrics cars in a random way.

As well as gas stations, these places should be located strategically. As well of different bunch of suppliers ready to bite a piece of cake from this market altogether with government.

1.1 Problem description

Let say I am businessperson interested to invest some of my money on this incoming business in the city of Puebla, Mexico. The question is...

Where are the best places to build them?

Thinking a little bit by walking on the customer shoes: Why should I go to other side of the country, if there is no way to charge my car properly and risk my trip back home? What benefit do I have by buying a electric car if there now way to charge or give it proper maintenance?

In order to make it a profitable business, not only a charge station should be build, but a complete net around the city. Therefore, we need to choose a strategically places to locate these charges spots.

Here we can approach the problem in different perspectives:

- By income (How much does a neighbourhood earns in average)
- By main streets
- By most frequently visited places
- By law requirements
- By the already installed charge spots.
- Between others

On this analysis, we will choose "By most frequently visited places"focus and use the Foursquare applications in order to find those places.

The main hypothesis comes as:

It is more likely that people recharge their cars in crowded places and where the recurrence is high. Not only for business is better but also for accessibility of people.

As follows, the clusters generated by K-means will helps us to identify these group of places, where people visit the most and where is bigger variety of venues.

2. Data to be manipulated

1. List of boroughs from Puebla City.

nte.mx

Códigos postales de Puebla Puebla, México

Información de asentamientos de Puebla Puebla México

Asentamiento	Nombre Asentamiento	Municipio	Estado	Código Postal
Colonia	15 de Septiembre	Puebla	Puebla	72227
Colonia	16 de Septiembre Norte	Puebla	Puebla	72230
Colonia	16 de Septiembre Sur	Puebla	Puebla	72474
Colonia	18 de Marzo	Puebla	Puebla	72595
Colonia	2 de Marzo	Puebla	Puebla	72227
Colonia	2 de Octubre	Puebla	Puebla	72498
Colonia	2a. Ampliación Unión Antorchista	Puebla	Puebla	72490
Colonia	2a. Sección 18 de Marzo	Puebla	Puebla	72595

We will use this zip code list to extract neighbourhoods from Puebla City.

2. Locations through Geocoder. In order to locate the location of each neighbourhood

3. Foursquare venues. In order to get the most visited places in the city of Puebla.

3. Methodology

3.1 Find a link where to get Puebla City boroughs and filter the relevant data. As well as correct labelling for our analysis:

	Asentamiento	Nombre Asentamiento	Municipio	Estado	Codigo Postal
0	Colonia	15 de Septiembre	Puebla	Puebla	72227
1	Colonia	16 de Septiembre Norte	Puebla	Puebla	72230
2	Colonia	16 de Septiembre Sur	Puebla	Puebla	72474
3	Colonia	18 de Marzo	Puebla	Puebla	72595
4	Colonia	2 de Marzo	Puebla	Puebla	72227

	Neighborhood	Borough	Codigo Postal
0	15 de Septiembre	Puebla	72227
1	16 de Septiembre Norte	Puebla	72230
2	16 de Septiembre Sur	Puebla	72474
3	18 de Marzo	Puebla	72595
4	2 de Marzo	Puebla	72227

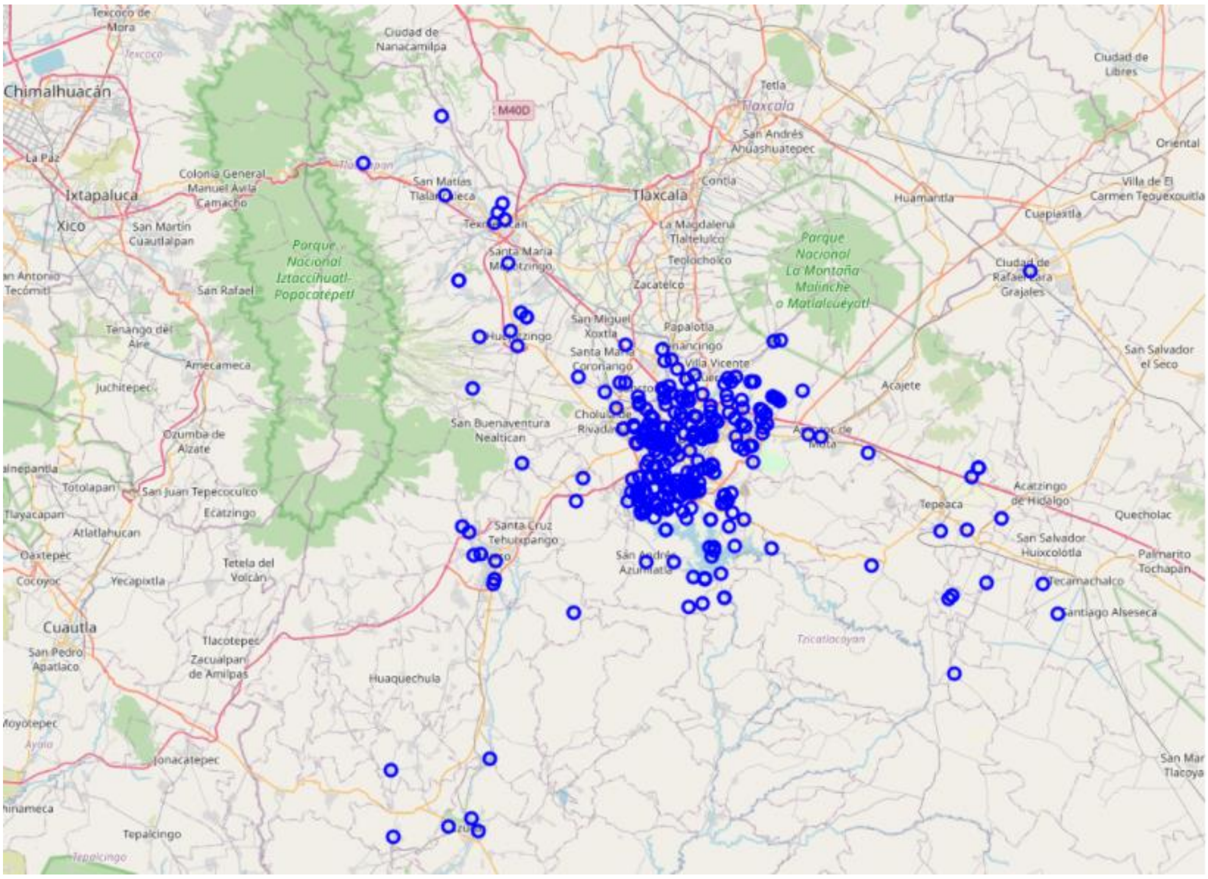
3.2 We will find the coordinates of each neighbourhood's using Geocoder.

3.3 Next step is to limit the radius of analysis around Puebla City. Everything out of a radius of ~50 km will be dropped. We consider that 1' Degree global coordinate is around 111 km.

	Unnamed: 0	Neighborhood	Borough	Codigo Postal	Latitude	Longitude
0	0	15 de Septiembre	Puebla	72227	19.019110	-98.220416
1	1	16 de Septiembre Norte	Puebla	72230	24.051496	-104.592982
2	2	16 de Septiembre Sur	Puebla	72474	18.994701	-98.219232
3	3	18 de Marzo	Puebla	72595	18.968889	-98.160278
4	4	2 de Marzo	Puebla	72227	19.051995	-98.254796

	Unnamed: 0	Neighborhood	Borough	Codigo Postal	Latitude	Longitude
0	2	16 de Septiembre Sur	Puebla	72474	18.994701	-98.219232
1	3	18 de Marzo	Puebla	72595	18.968889	-98.160278
2	4	2 de Marzo	Puebla	72227	19.051995	-98.254796
3	11	6 de Junio	Puebla	72227	19.111389	-98.147222
4	12	8 de Diciembre	Puebla	72227	19.085461	-98.140916

As tool of visualization, we´ll get support on Folium library to look on the locations:



3.4 Now we want to get the venues of this area and relate them to their proper neighbourhood:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	16 de Septiembre Sur	18.994701	-98.219232	El Taco Hidalguense	18.995973	-98.218067	Taco Place
1	16 de Septiembre Sur	18.994701	-98.219232	Burrito Loco	18.996484	-98.220291	Burrito Place
2	16 de Septiembre Sur	18.994701	-98.219232	Oxxo	18.997288	-98.222876	Convenience Store
3	16 de Septiembre Sur	18.994701	-98.219232	El Tiburón	18.995349	-98.220820	Seafood Restaurant
4	16 de Septiembre Sur	18.994701	-98.219232	Comex	18.996424	-98.218716	Outdoor Supply Store

3.5. There are some venues we will not consider on this analysis due to the following facts:

3.5.1 We remove drinking places from charge spots. We do want to promote drinking while driving. Safety first.

3.5.2 It does not make sense to consider those places, where people spend few time to get a benefit. For example:

- Convenience Stores
- Places "Take n' go"
- Small shops businesses
- Snacks stores
- Between others.

3.6 One hot coding

In order to continue with math operations, our next approach is to one hot the data, so we can work with numbers. After this process we will get the most repeated venues per area according to it's weight on the values:

```
----16 de Septiembre Sur----
               venue  freq
0  Seafood Restaurant   0.5
1       Garden Center   0.5
2     Motorcycle Shop   0.0
3           Multiplex   0.0
4              Museum   0.0
```

```
----2 de Marzo----
                   venue  freq
0           Soccer Field   0.4
1  Gym / Fitness Center   0.2
2                    Gym   0.2
3                 Lounge   0.2
4  Yucatecan Restaurant   0.0
```

```
----Acocota----
                venue  freq
0  Mexican Restaurant  0.40
1               Hotel  0.08
2                Park  0.04
3      Farmers Market  0.04
4              Church  0.04
```

3.7 Finally we will get the most common venues per neighbourhood, as well as the location of it:

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	16 de Septiembre Sur	Garden Center	Seafood Restaurant	Yoga Studio	Food	Fast Food Restaurant
1	2 de Marzo	Soccer Field	Gym / Fitness Center	Gym	Lounge	Yoga Studio
2	Acocota	Mexican Restaurant	Hotel	Farmers Market	Italian Restaurant	City Hall
3	Adolfo López Mateos	Mexican Restaurant	Japanese Restaurant	Burger Joint	Seafood Restaurant	Baseball Field
4	Alcanfores	Mexican Restaurant	Event Service	Flea Market	Fast Food Restaurant	Farmers Market

3.8 Here we can look on the most repeated venues:

Mexican Restaurant	403
Taco Place	346
Convenience Store	213
Coffee Shop	129
Restaurant	123
Seafood Restaurant	104
Pharmacy	92
Hotel	86
Pizza Place	82
Café	79
Bar	70
Gym / Fitness Center	60
Bakery	54
Italian Restaurant	49
Burger Joint	49
Ice Cream Shop	47
Park	45
Gym	44
Steakhouse	40

Total: 257 Categories

Relevant: 161

We will apply the criteria previously mentioned. And dropped all those categories, in which we have drinking places and "take n' go" venues (Examples highlighted in yellow).

3.9 Getting the most repeated venue type per neighbourhood. We will focus on the 5 most repeated ones:

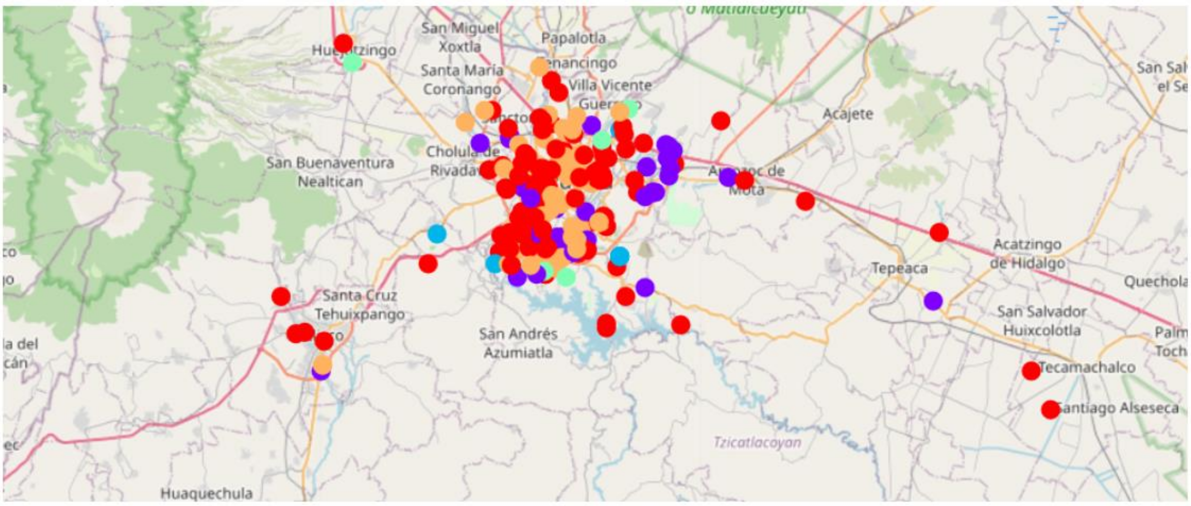
	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	15 de Septiembre	Seafood Restaurant	Mexican Restaurant	Restaurant	Burger Joint	Coffee Shop
1	16 de Septiembre Sur	Seafood Restaurant	Pool	Garden Center	Yucatecan Restaurant	Outdoor Sculpture
2	2 de Marzo	Soccer Field	Lounge	Japanese Restaurant	Gym / Fitness Center	Gym
3	8 de Diciembre	Park	Yucatecan Restaurant	Music Venue	Nail Salon	Nature Preserve
4	Acocota	Mexican Restaurant	Hotel	Seafood Restaurant	Garden	Gym / Fitness Center

3.10 Applying K-Means Method to group neighbourhoods into 5 clusters:

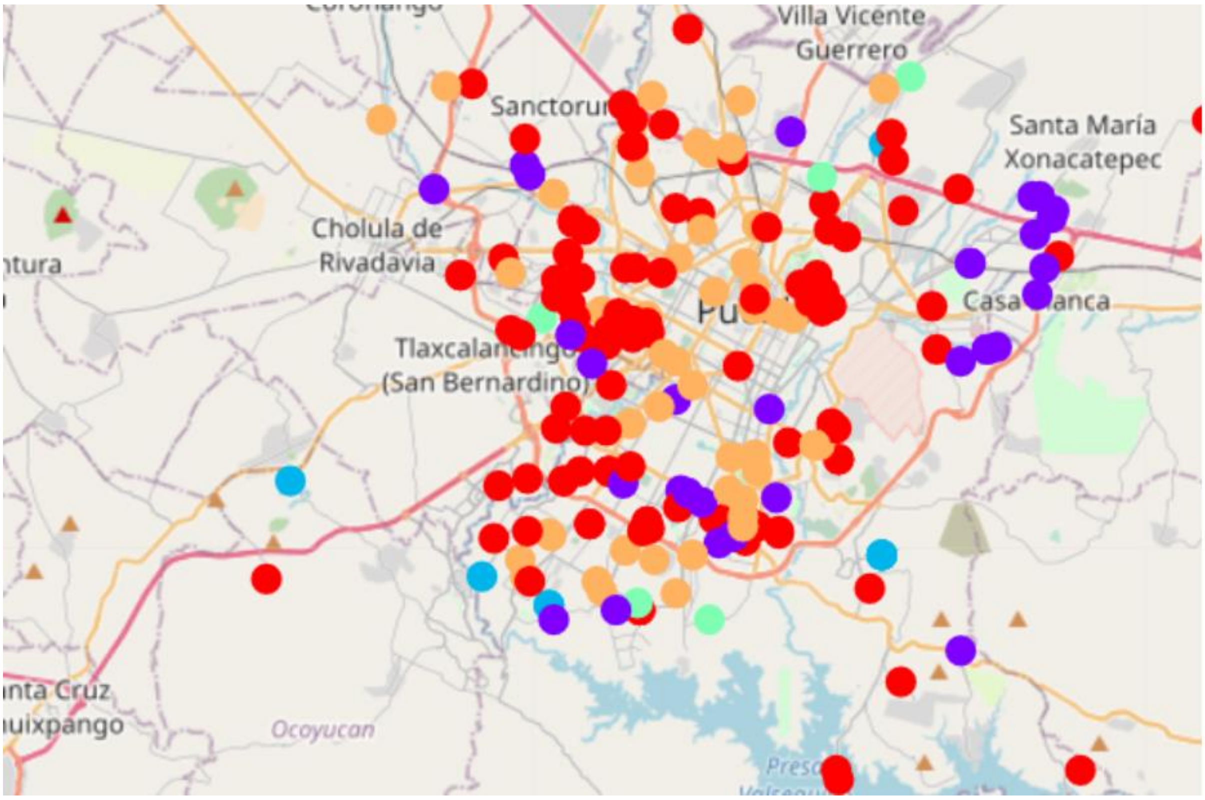
```
kclusters = 5

Puebla_grouped_clustering = Puebla_grouped.drop('Neighborhood', 1)

# run k-means clustering
kmeans = KMeans(n_clusters=kclusters, random_state=0).fit(Puebla_grouped_clustering)

# check cluster labels generated for each row in the dataframe
kint = kmeans.labels_
kint
```

```
5]: array([0, 0, 4, 1, 3, 3, 3, 4, 4, 0, 4, 0, 4, 4, 4, 0, 4, 4, 3, 4, 0, 4,
           3, 3, 0, 0, 4, 4, 4, 0, 0, 0, 3, 0, 4, 4, 4, 3, 3, 4, 3, 4, 4, 4,
           4, 4, 4, 4, 4, 0, 0, 4, 4, 4, 3, 4, 3, 3, 3, 4, 3, 3, 0, 4, 0, 4,
           1, 0, 3, 4, 4, 3, 4, 4, 3, 4, 0, 0, 4, 4, 4, 3, 3, 3, 3, 3, 4, 4,
           4, 0, 3, 0, 4, 0, 4, 2, 3, 4, 0, 4, 3, 0, 4, 4, 4, 4, 0, 3, 4, 4,
           3, 1, 4, 4, 4, 2, 4, 4, 4, 4, 4, 4, 3, 3, 0, 4, 4, 0, 3, 4, 3, 4,
           4, 4, 4, 4, 4, 3, 4, 4, 4, 4, 3, 4, 0, 4, 4, 4, 4, 3, 3, 3, 4, 4,
           4, 3, 0, 3, 4, 2, 3, 2, 2, 3, 4, 4, 4, 3, 2, 3, 4, 1, 3, 0, 4, 4,
           0, 0, 3, 3, 4, 4, 4, 0, 3, 1, 4, 4, 0, 3, 3, 3, 4, 4, 4, 4, 0, 4,
           3, 3, 3, 4, 0, 3, 4, 4, 4, 4, 3, 3, 4, 4, 4, 1, 1, 0, 3, 4, 4, 3,
           4, 4], dtype=int32)
```

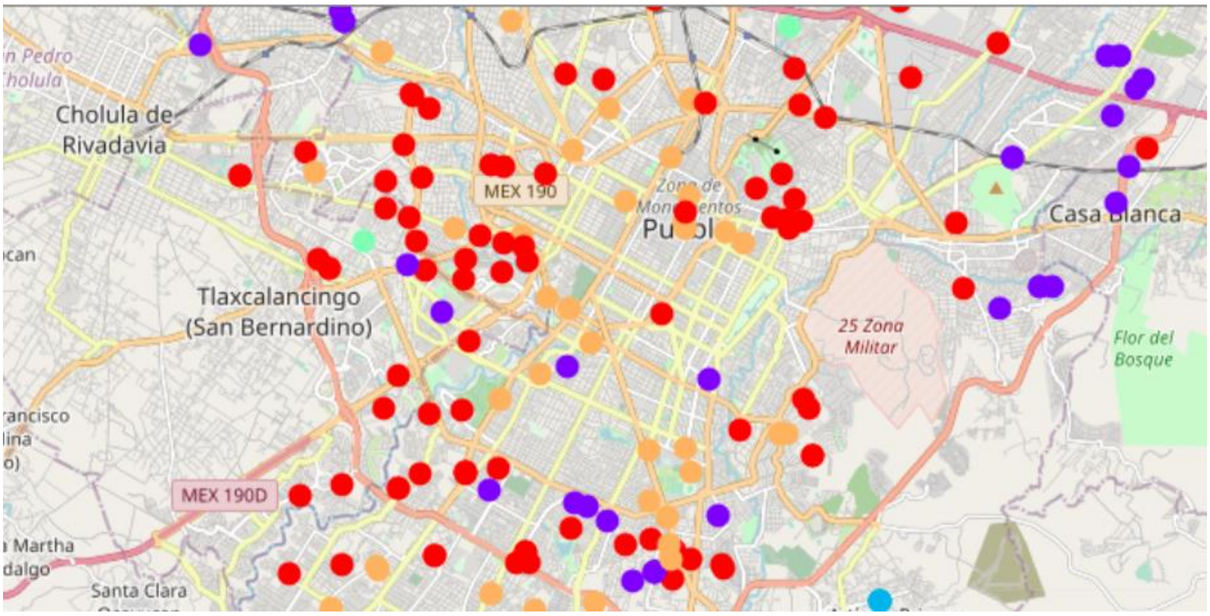
3.8 Clustering



Zoom: 10



Zoom 11



Zoom12

Cluster1: Purple

	Borough	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Puebla	0	Seafood Restaurant	Mexican Restaurant	Restaurant	Burger Joint	Coffee Shop
1	Puebla	0	Seafood Restaurant	Pool	Garden Center	Yucatecan Restaurant	Outdoor Sculpture
13	Puebla	0	Event Space	Restaurant	Pet Service	Nail Salon	Nature Preserve
15	Puebla	0	Hot Spring	Lounge	Restaurant	Nail Salon	Nature Preserve
23	Puebla	0	Café	Steakhouse	Burger Joint	Garden	Seafood Restaurant

Cluster 2: Blue

	Borough	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
5	Puebla	1	Park	Yucatecan Restaurant	Music Venue	Nail Salon	Nature Preserve
85	Puebla	1	Park	Restaurant	Yucatecan Restaurant	Music Venue	Nail Salon
86	Puebla	1	Park	Restaurant	Yucatecan Restaurant	Music Venue	Nail Salon
141	Puebla	1	Park	Tennis Court	Paintball Field	Music Venue	Nail Salon
227	Puebla	1	Park	Grocery Store	Music Venue	Nail Salon	Nature Preserve

Cluster 3: Cyan

	Borough	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
118	Puebla	2	Soccer Field	Comfort Food Restaurant	Yucatecan Restaurant	Park	Nail Salon
145	Puebla	2	Soccer Field	Yucatecan Restaurant	Paintball Field	Music Venue	Nail Salon
208	Puebla	2	Soccer Field	Yucatecan Restaurant	Paintball Field	Music Venue	Nail Salon
211	Puebla	2	Soccer Field	Garden Center	Yucatecan Restaurant	Museum	Music Venue
214	Puebla	2	Burger Joint	Soccer Field	Yucatecan Restaurant	Museum	Music Venue

Cluster 4: Orange

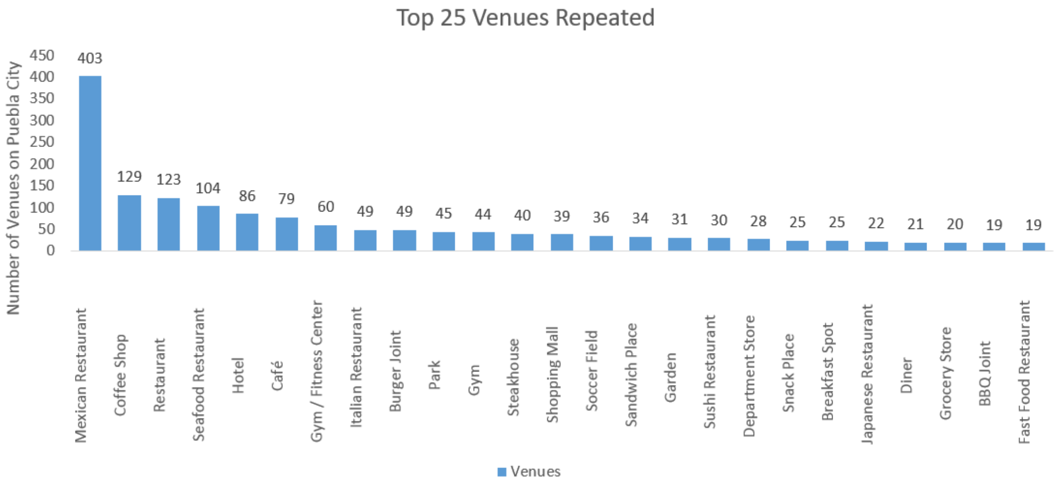
	Borough	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
6	Puebla	3	Mexican Restaurant	Hotel	Seafood Restaurant	Garden	Gym / Fitness Center
7	Puebla	3	Mexican Restaurant	Restaurant	Rock Climbing Spot	Health & Beauty Service	Flea Market
9	Puebla	3	Mexican Restaurant	Yucatecan Restaurant	Park	Nail Salon	Nature Preserve
26	Puebla	3	Pool Hall	Plaza	Mexican Restaurant	Tea Room	Yucatecan Restaurant
33	Puebla	3	Mexican Restaurant	Restaurant	Coffee Shop	Hotel	Vegetarian / Vegan Restaurant

Cluster 5: Red

	Borough	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
3	Puebla	4	Soccer Field	Lounge	Japanese Restaurant	Gym / Fitness Center	Gym
10	Puebla	4	Coffee Shop	Shopping Mall	Restaurant	Mexican Restaurant	Seafood Restaurant
12	Puebla	4	Garden Center	Lounge	Yucatecan Restaurant	Museum	Music Venue
14	Puebla	4	Shopping Mall	Hotel	Italian Restaurant	Hookah Bar	Grocery Store
16	Puebla	4	Shopping Mall	Restaurant	Fast Food Restaurant	Spanish Restaurant	French Restaurant

4. Results

Here we get different results on our analysis. Our first insight is that the places with most repeated venues are Mexican Restaurants:



Regarding the clusters we have the following results:

Cluster 1: Purple

If we consider Puebla City as a circle, most of these venues are located in the surroundings. This let me know, there is a pattern of places before someone enter or leaves the city.

Cluster 2: Blue

There are just few elements clustered. Most of them on them on the south on the city. Mostly related to fields and parks with less urbanization.

Cluster 3: Cyan

We see a similarity to the purple ones. Most of them on the surroundings, but less venues on this category.

Cluster 4: Orange

Here are clustered the most repeated venue (Mexican Restaurant). Most of them are well distributed inside Puebla City. Somehow opposite to the Purple Cluster.

Cluster 5: Red

This cluster contains the most repeated venues on it. It has two main distributions: On one side, the red points are mostly on the surroundings of busy areas. Such as university, some historical places, and wealthier neighbourhoods. On the other side, we can see some small towns around the main, with a prevalent red color.

5. Discussions.

There are three main facts that should be considered in order to have a more accurate model:

1. Some places were not found by the geocoder location service. Either because did not recognize the place or because it confuses itself with another town around the world with a similar name. So we got a loose of data during the process.
2. Recurrence of people on sites. Something interesting is that there are lot of Mexican restaurant in comparison with other venues. Nevertheless, is hard to say if the place will have high or low recurrence. This can be seen with Mexican Restaurants. Many of them are small places and businesses run out by families, even in small streets or houses. So just local people go these places frequently. Another information such a recurrence of people per week in a venue might also help.
3. Incomes. Let's be honest, people with a bigger income will be affordable to buy electric cars earlier than peoples with smaller ones. This might also help to set a priority where to place a charge spot. As consequence, this variable would also impact the analysis and clustering.

6. Conclusions

With this insight, let us return to the businessperson position.

Cluster purple is not fitting our request due to few places considered inside the town. Cluster blue a cyan just have few places not enough to build a complete net of chargers.

Now the decision lays either orange or red. If orange is selected, a bigger net inside the city can be created strategically. However, we are limited outside the city. On the other side, the red cluster will allow us to travel to the appendant towns. Which can be good for tourism and industries logistics.

In my point view, I would choose orange cluster. First, I make a net, which people can travel from daily places such as home, restaurants, gyms and work. This cluster have a bigger density of points per km2. In order, they can transport themselves to daily activities. Once this infrastructure is installed, I would proceed to red cluster. So now, it's possible for drivers to travel bigger distances in the surrounding towns.