New Taco Place in CDMX
Angelica Martinez
May 23, 2020

1. Introduction
1.1.   Background
Nowadays, if a person wanted to open a new business, be it a restaurant, a cafeteria, a flower shop, a hairdresser, a dental office, a pharmacy, a gym or any business that comes to mind, it is necessary to study the area where you would like your business to be located. From my point of view, many factors need to be considered. Mexico City is one of the most overpopulated cities in the world, just in my neighborhood there is a grocery store on every corner, for every 4 streets there is a gym and so many similar cases. Opening a new business in Mexico City is a little more complicated than it seems, opening a new business implies a bit of luck, that the product you offer pleases people more to prefer your business than the competition. All my life I have lived in the CDMX, I know which are the most dangerous colonies, the colonies with the largest population, the richest colonies, or the colonies where there are more companies. Taking advantage of the fact that this project is free, I would like to see how feasible it would be to put some "tacos" in one of the most popular areas in Mexico City. Considering that in Mexico there are some "tacos" on every street.

1.2.   Problem
It sounds pretty intrepid to want to open a new taco place within Mexico City. However, although it may not seem credible, there are not enough taco places within Mexico City, some of these are quite bad or are located in very dangerous areas (high crime), usually in residential areas within the CDMX there are not this class of restaurants, it is quite common to find Chinese, Italian, French or other less Mexican restaurants.

1.3.   Interest
The intention of this project is to be able to locate a new taco place for the public, in one of the most popular neighborhoods of Mexico City, considering the competition that could be found around it is probable that a person continues to prefer the taco place that you already know instead of ours. So, it is possible that we will be successful if we locate it in a popular area within Mexico City but there is no possible competition around it.

2. Data acquisition
2.1.   Data sources
My first tool to use would be foursquare, to be able to determine the businesses near the points where I would like to open my new "tacos", considering only for this first point the places that could be considered as my competition, since at the end of the day the idea It is not just opening a business to be opened, you have to be smart and consider that people are often based on routines. So even if my "tacos" are the richest in the world, it

would be ideal if I wasn't close to other businesses or restaurants that sell the products that I want to sell. For example, everyone in Mexico City knows that a taco stand has a better product than those bigger restaurants, which I don't want to say names, but that is a fact here in Mexico City and many other places.

The reason is that the taco stands are usually small, where a maximum of 30 people can be accommodated, therefore the meat that is bought is per day. I mean that it is a fresh product. First I would like to determine where a good location for my "tacos" could be, so the first thing I would do is investigate where there is more population in the delegations(is the way in which mexico city is divided) the name of the zone inside of Mexico City, for this I will use information from wikipedia and a dataset provided by the INEGI (Instituto National Statistics and Geography) in Mexico, has a website where we can download these datasets are for public use.

Determining which are the delegations with the largest population, we could also determine which are the most popular neighborhoods within those delegations. In the same way, I can find this information on wikipedia, on various pages of the government of Mexico there are data for public use. I would make the necessary scrapers to be able to extract the information to delimit more where I would put my "tacos". Having already defined the area by neighborhood, I would use foursquare to see what businesses are in that geographic point. Finally, I could cluster to secure the ideal point within that neighborhood where I could have some advantage in selling my product.

2.2.   Data cleaning

As a first approach we will determine which are the popular 'colonies' according to some popular websites, such as wikipedia, viahero and culture trip. For this we will carry out web scraping to extract this information.

Part of data cleaning is that we consider removing special characters such as the tilde in vowels, and we will handle all the information in capital letters in order to standardize this information.

One of the official websites of Mexico is INEGI, INEGI has several sites, one of these is https://datos.cdmx.gob.mx/explore/dataset/coloniascdmx/ from where a file with .geojson extension was obtained, this is open documentation so this contained the delegations, colonies and polygons of its geographical area, this dataset is already really clean, perhaps in the future what should be done is to eliminate the columns that do not really provide us with information, or in its opposite case add more columns for a more in-depth analysis.

2.3.   Feature selection

As mentioned in the previous point, our main dataset is the one that was acquired from the official page of Mexico

Making a merge between the main or most popular 'colonias' in the CDMX and with this dataset, we will only stay with the dataset of these most common places.

An extra activity in order to generate a new feature with which we will later be working, is to extract the latitude and longitude of the centroid of each of the polygons that reflect the geographical area of each of these 'colonias'.

	entidad	cve_alc	alcaldia	secc_com	secc_par	nombre	cve_col	geometry
0	9.0	16	MIGUEL HIDALGO	4924, 4931, 4932, 4935, 4936, 4940, 4987	4923, 4937, 4938, 4939, 4942	LOMAS DE CHAPULTEPEC	16-042	POLYGON ((-99.22017 19.42803, -99.22009 19.428...
1	9.0	16	MIGUEL HIDALGO	4963	4964	LOMAS DE REFORMA (LOMAS DE CHAPULTEPEC)	16-044	POLYGON ((-99.22967 19.41406, -99.22970 19.413...
2	9.0	16	MIGUEL HIDALGO	None	4918, 4919	DEL BOSQUE (POLANCO)	16-026	POLYGON ((-99.20821 19.43282, -99.20813 19.432...
3	9.0	3	COYOACAN	433, 500, 431, 513, 501	424, 425, 426, 430, 499	PEDREGAL DE SANTA URSULA I	03-135	POLYGON ((-99.14587 19.31979, -99.14579 19.319...
4	9.0	3	COYOACAN	376, 377, 378, 379, 404, 493, 498	374	AJUSCO I	03-128	POLYGON ((-99.15854 19.33038, -99.15785 19.329...

The previous image we can observe the data that this geopandas contains.
Then we can see our resulting geopandas after the creation of these two new features.

|	alcaldia	cve_alc	cve_col	entidad	geometry	nombre	secc_com	secc_par	latitude	longitude
0	MIGUEL HIDALGO	16.0	16-042	9.0	POLYGON ((-99.22017088373187 19.42803250649744...	LOMAS DE CHAPULTEPEC	4924, 4931, 4932, 4935, 4936, 4940, 4987	4923, 4937, 4938, 4939, 4942	19.422841	-99.215794
1	MIGUEL HIDALGO	16.0	16-044	9.0	POLYGON ((-99.22967474076427 19.4140557307484,...	LOMAS DE REFORMA (LOMAS DE CHAPULTEPEC)	4963	4964	19.410616	-99.226249
2	MIGUEL HIDALGO	16.0	16-026	9.0	POLYGON ((-99.2082100184801 19.4328156500052, ...	DEL BOSQUE (POLANCO)	None	4918, 4919	19.434219	-99.209404
7	VENUSTIANO CARRANZA	17.0	17-073	9.0	POLYGON ((-99.12511135172929 19.42919898252439...	CENTRO II	5261, 5264, 5265, 5266	5263	19.425714	-99.122660
13	COYOACAN	3.0	03-013	9.0	POLYGON ((-99.11866329441514 19.30664276633733...	CAMPESTRE COYOACAN (FRACC)	None	650, 651, 675	19.308823	-99.117055

3. Exploratory data analysis
3.1.   Analysis

Considering the most popular neighborhoods, we will begin to graph these within Mexico City. Remember that geographical coordinate of CDMX are 19.4326296, -99.1331785.
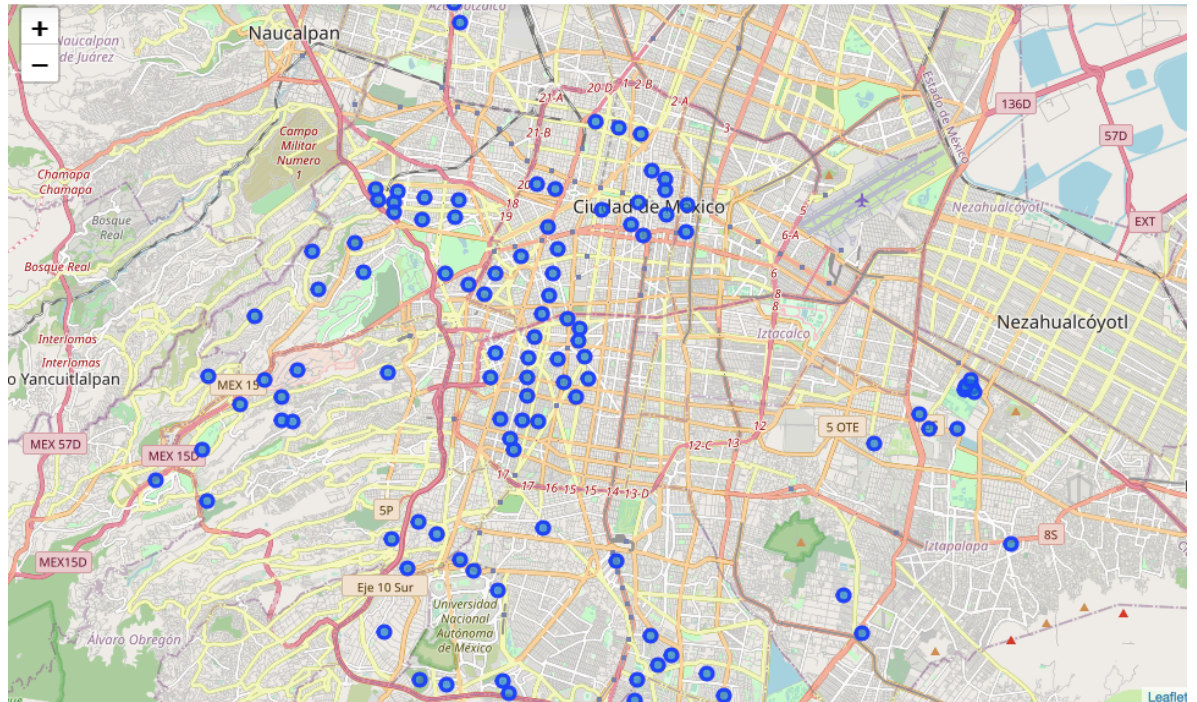
Let's concentrate on one of the "delegaciones" where the most popular "colonias" are located: Cuauhtemoc This is one of the "delegations" famous for different factors, however as it is popualr it is certain that there will be a greater amount of population. Below we can see the result only of the "colonias" within the Cuauhtemoc delegation.
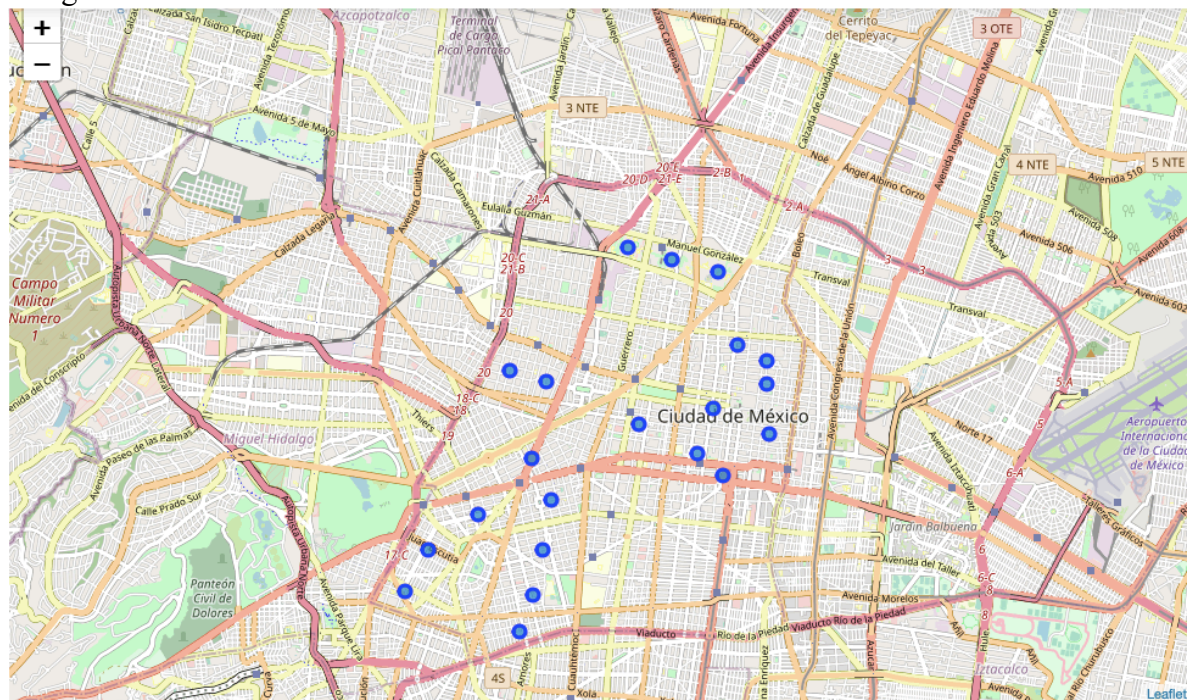
Making use of the Foursquare API to see which are the most common venues and their frequency within those points of the CDMX.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	CENTRO VII	19.430225	-99.128141	El Antiguo Edhen	19.430340	-99.129250	Falafel Restaurant
1	CENTRO VII	19.430225	-99.128141	Ehden	19.430328	-99.129244	Middle Eastern Restaurant
2	CENTRO VII	19.430225	-99.128141	Casa Talavera	19.428149	-99.127677	Art Gallery
3	CENTRO VII	19.430225	-99.128141	Al Andalus	19.427881	-99.129224	Middle Eastern Restaurant
4	CENTRO VII	19.430225	-99.128141	Chilli-Aquilli	19.428940	-99.126986	Restaurant

Using the one hot encoding technique we are going to create a new dataframe with the venues more frequently in those geographical points.

|	Neighborhood	Accessories Store	American Restaurant	Antique Shop	Argentinian Restaurant	Art Gallery	Art Museum	Arts & Crafts Store	Asian Restaurant	Athletics & Sports	...	Thrift / Vintage Store	Toy / Game Store	Trail	Vegetarian / Vegan Restaurant	Venezu Restau
0	CENTRO VII	0	0	0	0	0	0	0	0	0	...	0	0	0	0	
1	CENTRO VII	0	0	0	0	0	0	0	0	0	...	0	0	0	0	
2	CENTRO VII	0	0	0	0	1	0	0	0	0	...	0	0	0	0	
3	CENTRO VII	0	0	0	0	0	0	0	0	0	...	0	0	0	0	
4	CENTRO VII	0	0	0	0	0	0	0	0	0	...	0	0	0	0	

Now doing a reset_index and grouping the dataframe by word frequency, we are left with the following result.

	Neighborhood	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	CENTRO I	Taco Place	Mexican Restaurant	Bridal Shop	History Museum	Deli / Bodega	Record Shop	Museum	Market	Food Stand	Spanish Restaurant
1	CENTRO II	Taco Place	Jewelry Store	Toy / Game Store	Science Museum	Restaurant	Fast Food Restaurant	Bar	History Museum	Historic Site	Hostel
2	CENTRO III	History Museum	Historic Site	Art Museum	Jewelry Store	Mexican Restaurant	Museum	Cosmetics Shop	Restaurant	Bookstore	Concert Hall
3	CENTRO IV	Mexican Restaurant	Department Store	Hotel	Ice Cream Shop	Art Museum	Arts & Crafts Store	Boutique	Bakery	Coffee Shop	Bar
4	CENTRO V	Clothing Store	Mexican Restaurant	Boutique	Hotel	Bakery	Café	Pharmacy	Pedestrian Plaza	Spanish Restaurant	Pizza Place

At this point our dataframe is ready to cluster. Making use of k-means clustering considering k the number of clusters equal to 5 we would obtain the following dataframe.

secc_com	secc_par	latitude	longitude	...	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
4749, 4750, 4751, 4752, 4753, 4754, 4756	None	19.430225	-99.128141	...	Taco Place	Museum	Art Museum	Mexican Restaurant	Middle Eastern Restaurant	Arts & Crafts Store	Restaurant	History Museum	Falafel Restaurant	Music Venue
4594, 4595, 4596, 4597, 4675, 4676, 4712	None	19.453315	-99.141769	...	Movie Theater	Mexican Restaurant	Historic Site	Burger Joint	Pizza Place	Taco Place	Park	Coffee Shop	Diner	Sandwich Place
4852, 4865, 4871, 4872, 4873, 4882, 4883, 48...	4870	19.427004	-99.161605	...	Coffee Shop	Bakery	Art Gallery	Cosmetics Shop	Hotel	Italian Restaurant	Restaurant	Donut Shop	Comfort Food Restaurant	Men's Store
4734, 4735, 4736, 4737, 4738, 4739	None	19.439850	-99.128518	...	Taco Place	Jewelry Store	Toy / Game Store	Science Museum	Restaurant	Fast Food Restaurant	Bar	History Museum	Historic Site	Hostel
4526, 4527, 4528, 4529	4530, 4531, 4535, 4537	19.419419	-99.169162	...	Seafood Restaurant	Coffee Shop	Bistro	Restaurant	Bakery	Pizza Place	Gym / Fitness Center	Taco Place	Tea Room	Café

Now we can visualize it within the 'delegacion' Cuauhtemoc within the city of Mexico
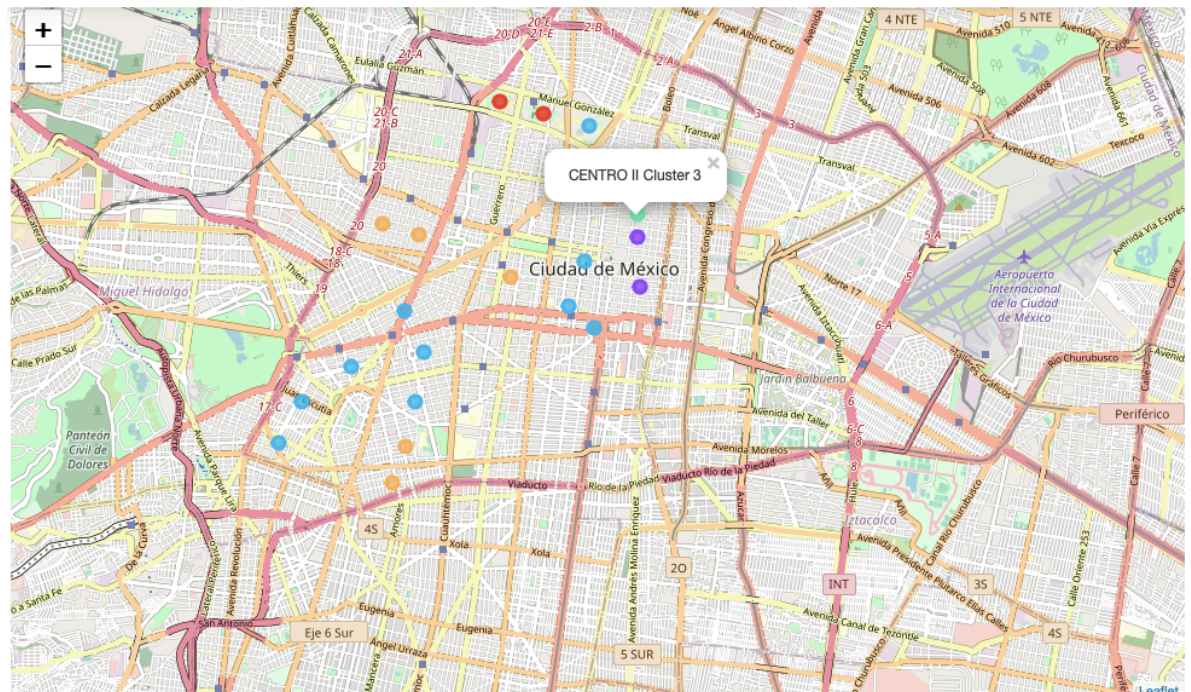
4. Results
4.1.   Examine

Analyzing the result of cluster 1, we can see that if we wanted to open a new taco place in this geographical area, we would have enough competition.

	cve_alc	nombre	secc_com	secc_par	latitude	longitude	Cluster Labels	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	8 C
1	15.0	NONOALCO-TLATELOLCO (U HAB) II	4594, 4595, 4596, 4597, 4675, 4676, 4712	None	19.453315	-99.141769	0	Movie Theater	Mexican Restaurant	Historic Site	Burger Joint	Pizza Place	Taco Place	Park	
11	15.0	NONOALCO-TLATELOLCO (U HAB) I	4598, 4599, 4600, 4601, 4602, 4603, 4604	None	19.454913	-99.148085	0	Pizza Place	Restaurant	Mexican Restaurant	Music Venue	Sushi Restaurant	Park	Diner	

Analyzing the result of cluster 2, we can see that if we wanted to open a new taco place in this geographical area, we would have enough competition.

	cve_alc	nombre	secc_com	secc_par	latitude	longitude	Cluster Labels	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 Com V
0	15.0	CENTRO VII	4749, 4750, 4751, 4752, 4753, 4754, 4756	None	19.430225	-99.128141	1	Taco Place	Museum	Art Museum	Mexican Restaurant	Middle Eastern Restaurant	Arts & Crafts Store	Restaurant	Hi Mus
10	15.0	CENTRO III	4740, 4741, 4742, 4743, 4744, 4745	None	19.436771	-99.128452	1	History Museum	Historic Site	Art Museum	Jewelry Store	Mexican Restaurant	Museum	Cosmetics Shop	Resta

Analyzing the result of cluster 3, we can see that if we wanted to open a new taco place in this geographical area, we would have enough competition.

	cve_alc	nombre	secc_com	secc_par	latitude	longitude	Cluster Labels	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 M Comm Ver
2	15.0	JUAREZ	4852, 4865, 4871, 4872, 4873, 4882, 4883, 48...	4870	19.427004	-99.161605	2	Coffee Shop	Bakery	Art Gallery	Cosmetics Shop	Hotel	Italian Restaurant	Restaur
4	15.0	ROMA NORTE I	4526, 4527, 4528, 4529, 4534	4530, 4531, 4535, 4537	19.419419	-99.169162	2	Seafood Restaurant	Coffee Shop	Bistro	Restaurant	Bakery	Pizza Place	Gy Fitn Cer
5	15.0	CENTRO IV	4839, 4840, 4841, 4842, 4747, 4748, 4838, 4843...	None	19.433636	-99.136030	2	Mexican Restaurant	Department Store	Hotel	Ice Cream Shop	Art Museum	Arts & Crafts Store	Bouti
8	15.0	CONDESA	4532, 4533, 4549, 4550, 4551	4530, 4531, 4535, 4552, 4553	19.414750	-99.176210	2	Taco Place	Bakery	Coffee Shop	Ice Cream Shop	Restaurant	Juice Bar	Beer
12	15.0	NONOALCO-TLATELOLCO (U HAB) III	4670, 4671, 4672, 4673, 4674, 4677, 4678	None	19.451696	-99.135357	2	Park	History Museum	Bakery	Historic Site	Museum	Public Art	Convenien St

Analyzing the result of cluster 4, we can see that if we wanted to open a new taco place in this geographical area, we would have enough competition.

| cve_alc	nombre	secc_com	secc_par	latitude	longitude	Cluster Labels	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
15.0	CENTRO II	4734, 4735, 4736, 4737, 4738, 4739	None	19.43985	-99.128518	3	Taco Place	Jewelry Store	Toy / Game Store	Science Museum	Restaurant	Fast Food Restaurant	Bar	History Museum

5. Observations

Considering the results by clusters it looks like the most common venue per kluster, only located in cuauhtemoc neighborhood. Don't look like a really good option this neighborhood to place a new 'tacos' restaurant. Mexican Restaurant always contains 'tacos' in their menus and 'Taco place' the same name tell us it's a 'tacos' restaurant. In conclusion, Cuauhtemoc even when it's one of the most popular neighborhoods or 'colonias' in CDMX, is not a feasible option to open our new 'tacos'.

However, we still have options within the popular neighborhoods of Mexico City, so we repeat the process for 3 different 'delegations': Cuajimalpa, Miguel Hidalgo and Coyoacan.
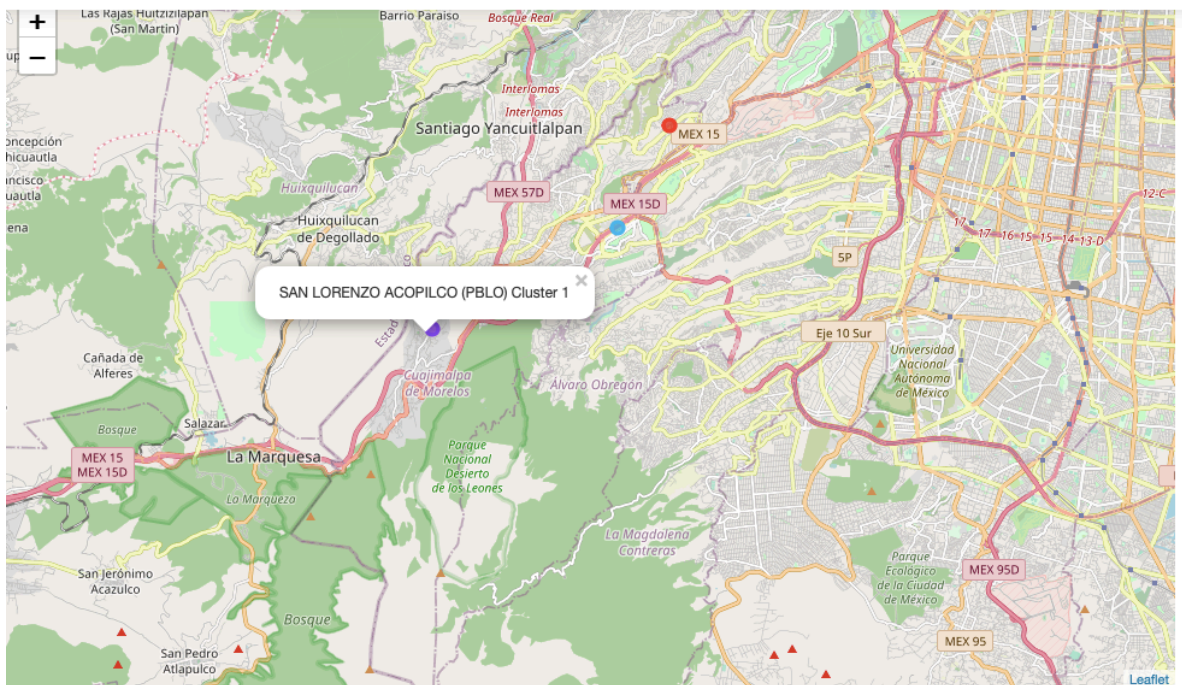
This image is the map with the results of clustering k = 5 in Coyoacan.
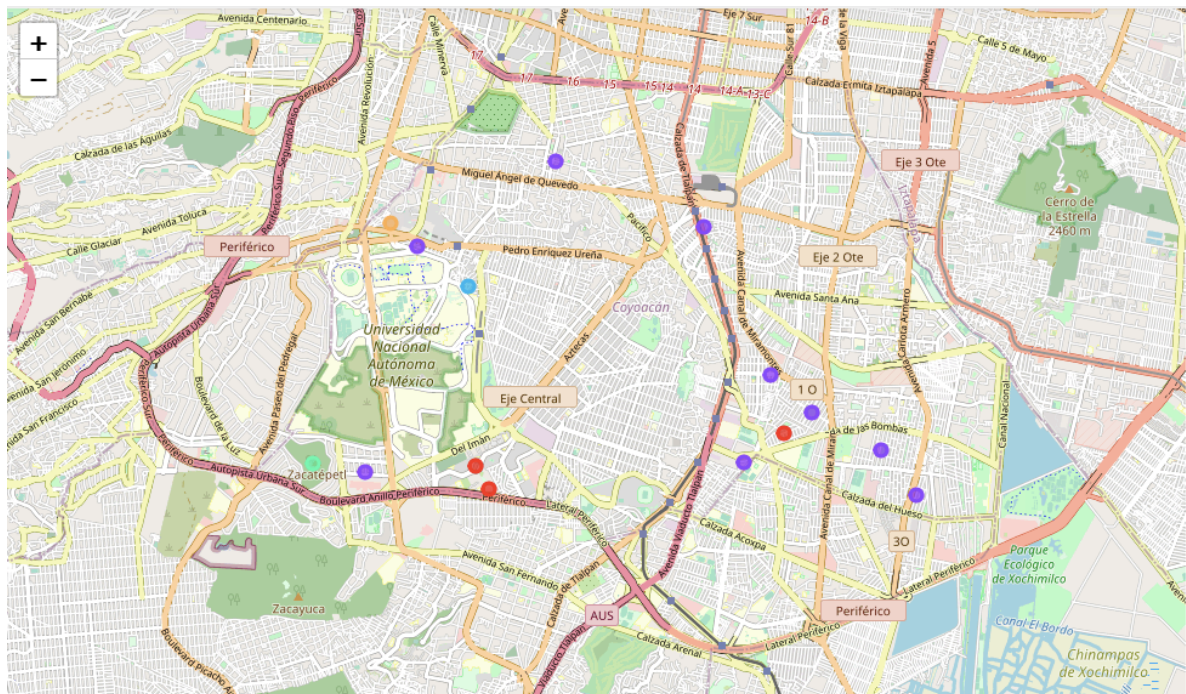
This image is the map with the results of clustering k = 5 in Miguel Hidalgo.
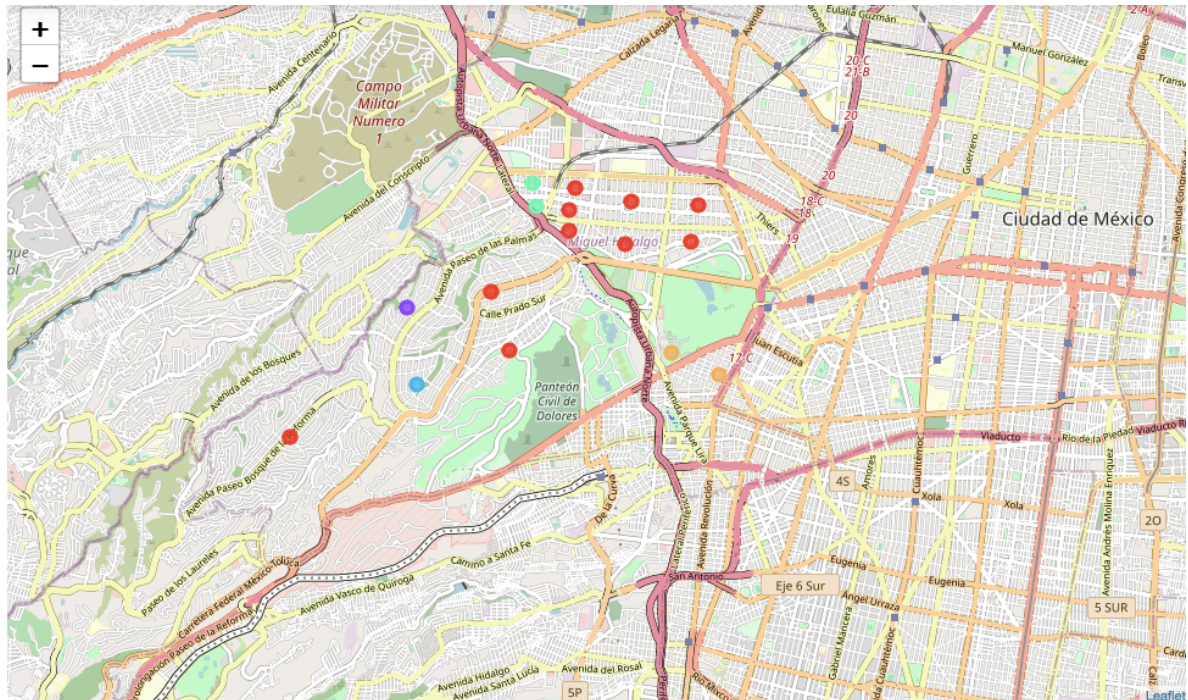
This image is the map with the results of clustering k = 3 in Cuajimalpa.

6. Conclusion

This analysis is based on two important phases, the extraction of information and its analysis. In order to extract information, we could consider several web pages to determine which are the most popular neighborhoods in Mexico City. It is intrepid to want to put some 'tacos' in Mexico City, knowing that it is the most popular Mexican food in the world. Similarly, it is well known that in Mexico City it is quite common to find some 'tacos' everywhere. However, this analysis is focused on being able to determine where we would run the least risk of bankruptcy to locate this business, the main element to consider was the competition to discard places or to consider possible places or areas where it could be successful.

From the scraping that was done for 3 popular web pages on the internet, we determined the most popular colonies. Considering the popular 'colonias' we were able to obtain the most popular 'delegaciones'. These 'colonies' are popular for different things, a deeper analysis would have to consider the population, the socioeconomic status by delegation and an important factor at least here in CDMX is the crime found in that 'delegation' or in that 'colony' since many times this would also be a factor in determining whether or not the business could be successful.

We can determine that the neighborhoods where there is no nearby 'taco' place would be in the POLANCO neighborhood of the 'delegation' Miguel Hidalgo, the second option would be SAN ANGEL of the 'delegation' COYOACAN, and an option where we apparently do not have almost no competition would be in practically the entire CUAJIMALPA 'delegation'.