**New Taco Place in CDMX**

Angelica Martinez

May 23, 2020

1. **Introduction**
   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. **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. **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.

1. **Data acquisition**
   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.

* 1. **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.

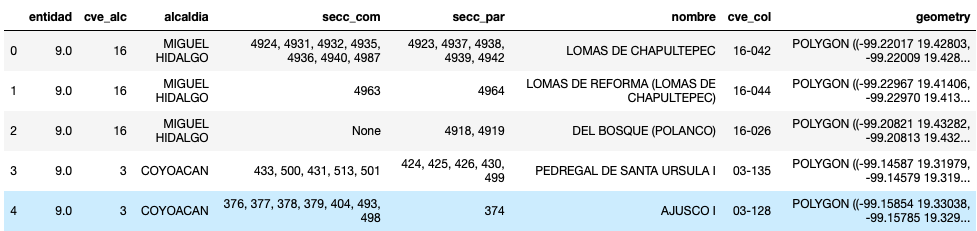
One of the official websites of Mexico is INEGI, INEGI has several sites, one of these is <https://datos.cdmx.gob.mx/explore/dataset/coloniascdmx/> from where a file with .geojson extension was obtained, this is open documentation so this contained the delegations, colonies and polygons of its geographical 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.

* 1. **Feature selection**

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

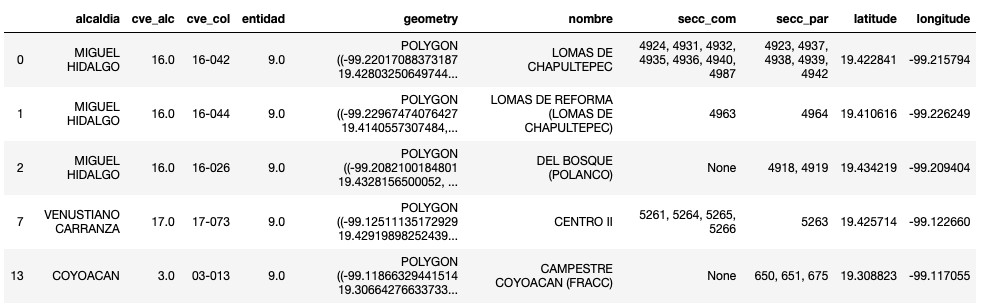
Making a merge between the main or most popular 'colonias' in the CDMX and with this dataset, we will only stay with the dataset of these most common places.

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'.

****

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.



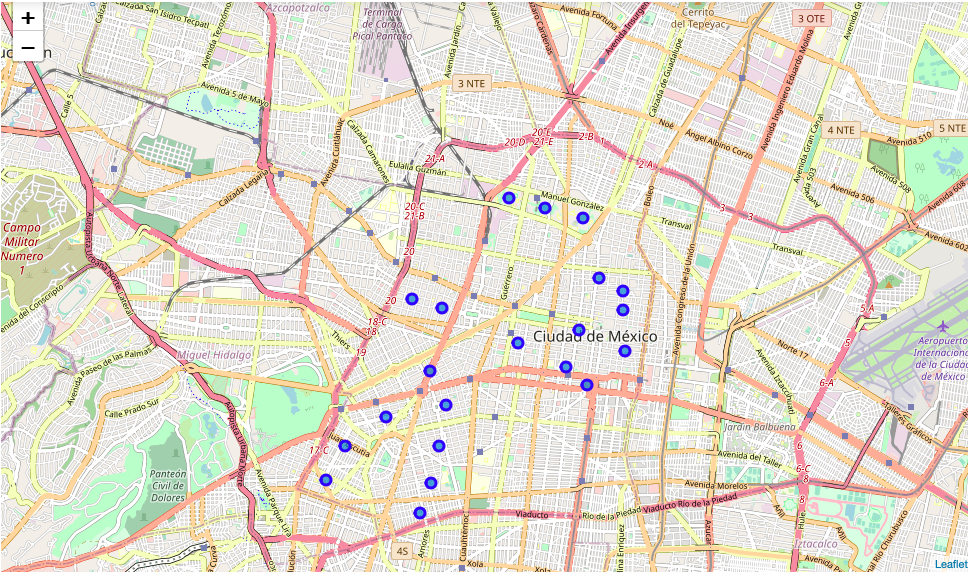
1. **Exploratory data analysis**
   1. **Analysis**

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.



****

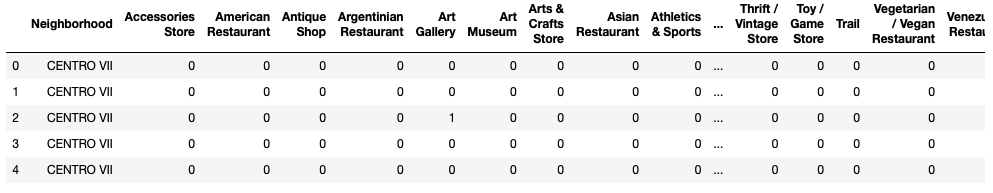
Let's concentrate on one of the "delegaciones" where the most popular "colonias" are located: Cuauhtemoc This is one of the "delegations" famous for different factors, however as it is popualr it is certain that there will be a greater amount of population. Below we can see the result only of the "colonias" 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.



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



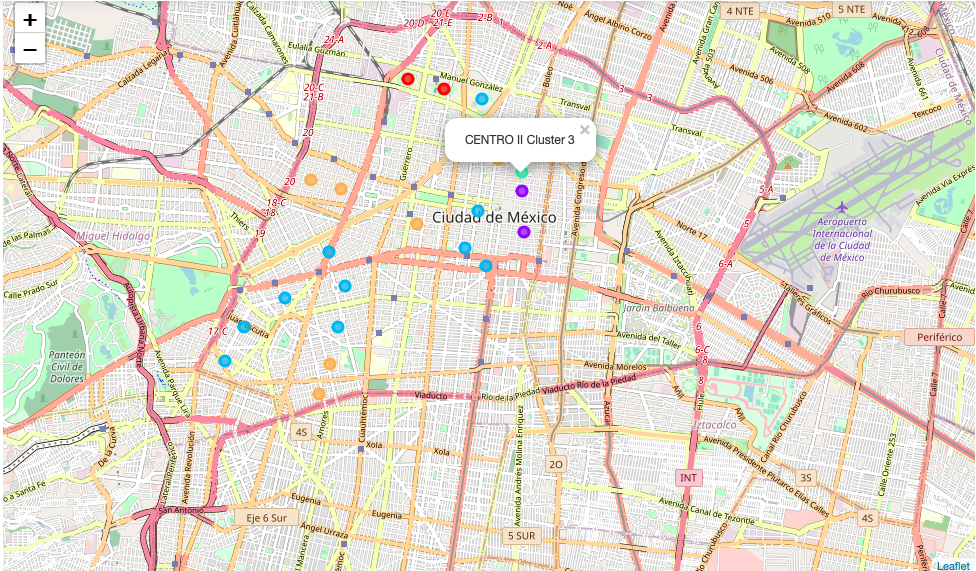
Now doing a reset\_index and grouping the dataframe by word frequency, we are left with the following result.



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.



Now we can visualize it within the 'delegacion' Cuauhtemoc within the city of Mexico



1. **Results**
   1. **Examine**

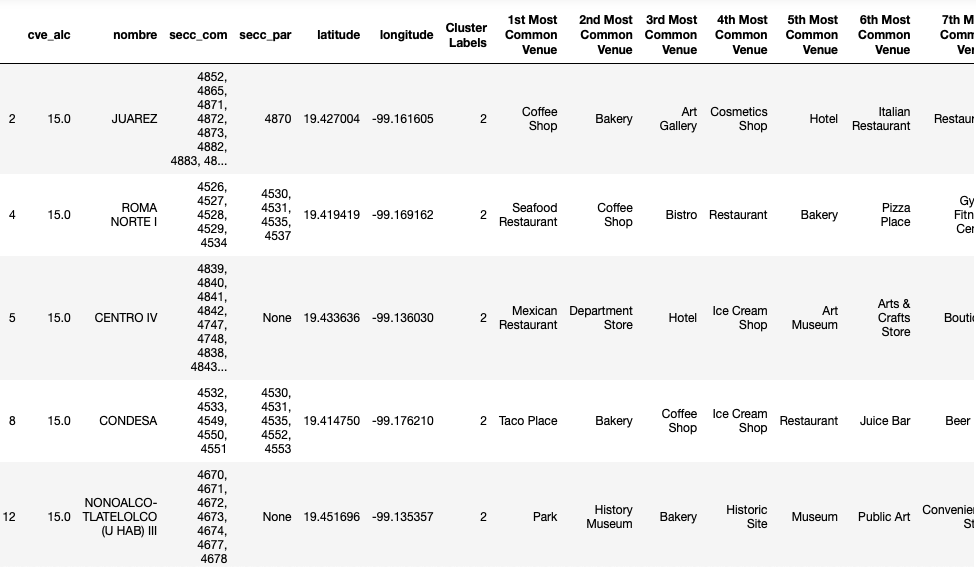
Analyzing the result of cluster 1, we can see that if we wanted to open a new taco place in this geographical area, we would have enough competition.



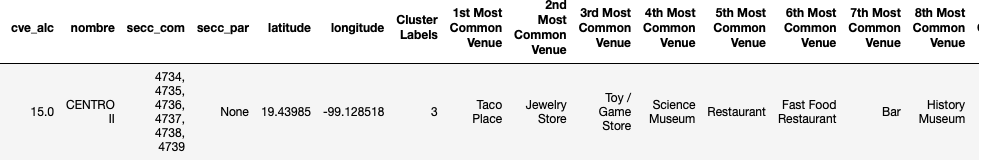
Analyzing the result of cluster 2, we can see that if we wanted to open a new taco place in this geographical area, we would have enough competition.

****

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.

****

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.

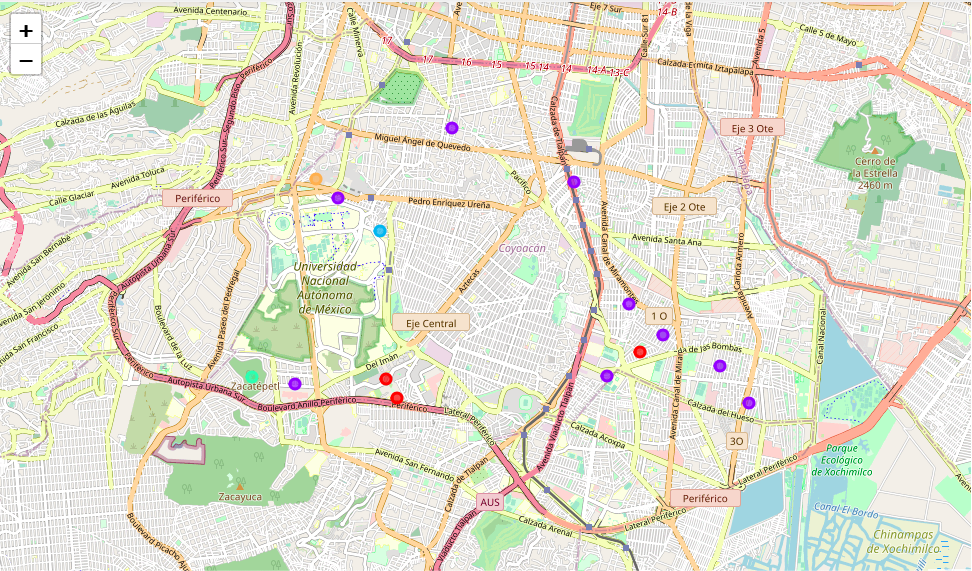
****

1. **Observations**

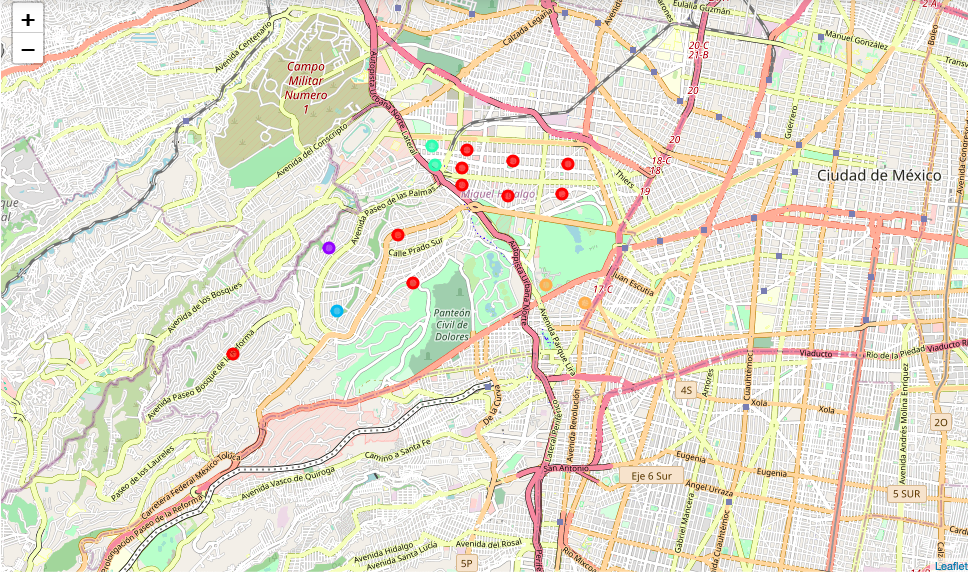
Considering the results by clusters it looks like the most common venue per kluster, only located in cuauhtemoc neighborhood. Don't look like a really good option this neighborhood to place a new 'tacos' restaurant. Mexican Restaurant always contains 'tacos' in their menus and 'Taco place' the same name 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.

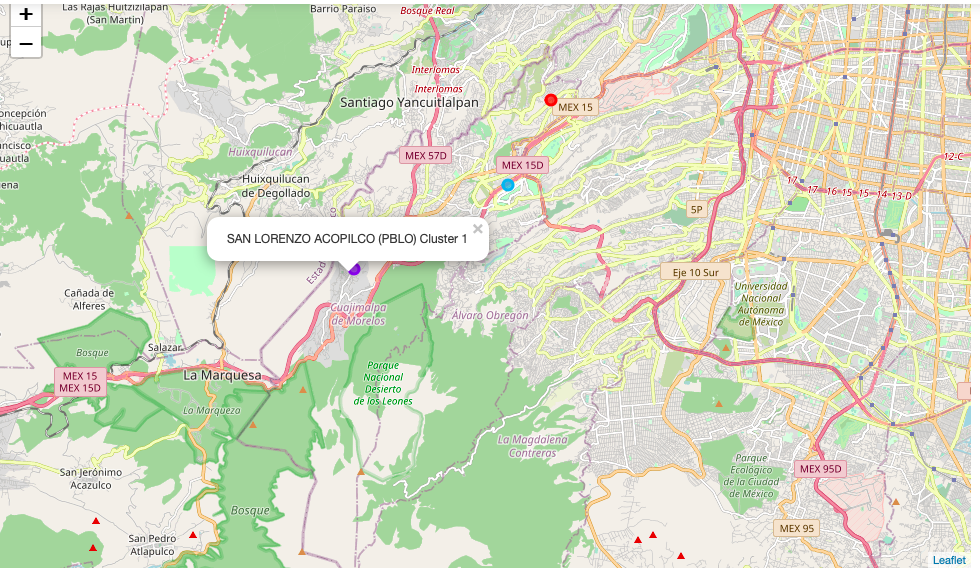
This image is the map with the results of clustering k = 5 in Coyoacan.



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.



1. **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'.