

Finding feasible locations to open a sea food restaurant in Miguel Hidalgo borough in Mexico City

By Marco Hernández, Data Science IBM / Coursera Student from

Business problem

Mexico City zone is considered one of the most multicultural places in Latino America. You may find people from diverse nationalities that make business, study, or go for vacations. Restaurants in Mexico City are full of competitors that offer different types of food and are usually concentrated in specific neighborhoods, especially those where Business offices from multinational companies are settled.

Miguel Hidalgo borough is one of the most important zones in Mexico City, because there are important venues as museums, touristic attractions, corporate offices, theaters, and different places to eat. Many restaurant franchises use to look for rental spaces in Miguel Hidalgo Borough to take part of the marketplace.

Sea food* is one of the most attractive types of food that customer looks for in Miguel Hidalgo Borough. This data science project looks to provide information to franchises or entrepreneurs looking to understand how the different types of restaurants are distributed in Miguel Hidalgo borough neighborhoods. The research will allow to get different clusters that allow us to identify the most common venues in the borough and we expect to identify if there are neighborhoods where sea food and different types of restaurants are located.

Description of data to be used for the project

Data Source: We will use the data from Mexican Postal service to get the Postal codes, boroughs, and neighborhoods in Mexico City. The data are in a CSV file that was loaded to Jupiter lab. This file required data wrangling tasks that we did by dropping data without information or dropping those rows with duplicated information.

We installed **Geocoder** to get the latitudes and longitudes in our notebook in **Jupyter Lab**.

Once we had the latitudes and longitudes for each neighborhood, we created a *dataframe* that was converted into a **CSV file**. This file was the cornerstone that we use as input to identify venues in Miguel Hidalgo borough.

To get information from venues we used **Foursquare App** that being used in combination with our **Data Science code in Python** allow us to get the relevant places in every neighborhood as well as their categories.

We used **K means clustering** a machine learning algorithm to segment the venues in Miguel Hidalgo borough. The last step allowed us to identify how the venues are located and as well as those were classified in clusters. Now we are able to understand if there's a cluster of restaurants and where are those located, more specifically now we can identify if *there are* clusters where Seafood restaurants are located, as well as we can now identify clusters where *there are not* Seafood restaurants.

Interest: From the point of view of business(entrepreneurs and franchises) by having a clear reference of the venues in specific clusters will allow to have a perspective of which are the more feasible neighborhoods to open a Seafood restaurant from the location perspective.

Methodology

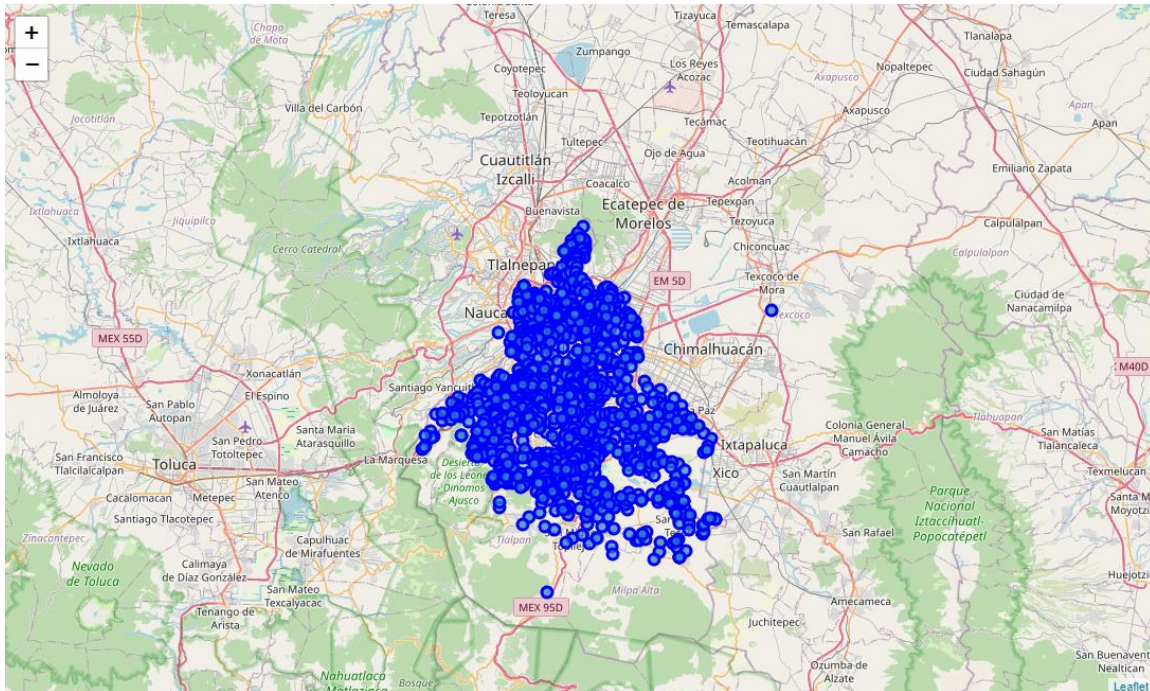
Exploratory Data Analysis

Identification of neighborhoods and venues in Mexico City

As we discussed before, we create a file with the boroughs, neighborhoods, latitudes, and longitudes in Mexico City. Since we have this data, we create a table that allow us to identify that **we have 16 boroughs and 1367 neighborhoods in Mexico City.**

	PostalCode	Borough	Neighborhood	Latitude	Longitude
count	1515.000000	1515	1515	1515.000000	1515.000000
unique	NaN	16	1367	NaN	NaN
top	NaN	Álvaro Obregón	San Miguel	NaN	NaN
freq	NaN	222	8	NaN	NaN
mean	8543.928053	NaN	NaN	19.376632	-99.150321
std	4909.859519	NaN	NaN	0.355640	0.500013
min	1000.000000	NaN	NaN	16.286210	-116.635340
25%	4260.000000	NaN	NaN	19.306595	-99.200615
50%	8910.000000	NaN	NaN	19.364240	-99.143810
75%	13099.500000	NaN	NaN	19.441475	-99.101855
max	16900.000000	NaN	NaN	31.890930	-93.096720

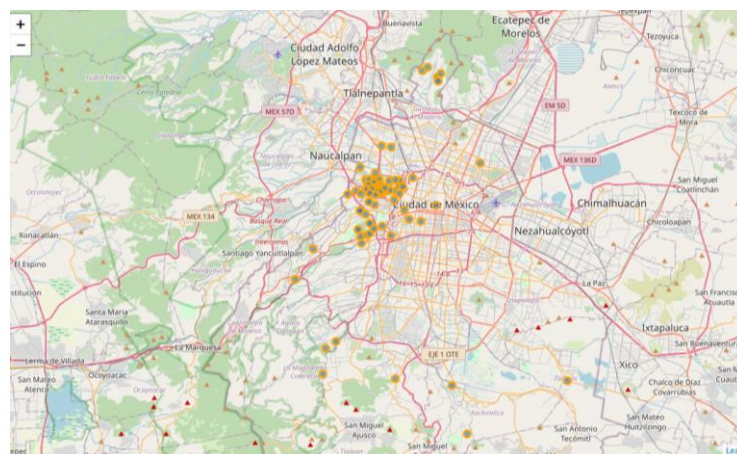
Taking the original database we can deploy a map to observe the neighborhoods in Mexico City



The borough of our interest is Miguel Hidalgo which have 85 neighborhoods

	PostalCode	Borough	Neighborhood
count	85.000000	85	85
unique	NaN	1	85

We can see **the 85 neighborhoods from Miguel Hidalgo borough** in the next map



We have identified the main venues near the neighborhoods as well as its category, we show a sample of the table that we created for any venue and its category in any neighborhood for Miguel Hidalgo Borough.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Lomas de Chapultepec I Sección	19.36216	-99.27344	Lego Store México	19.361742	-99.272794	Toy / Game Store
1	Lomas de Chapultepec I Sección	19.36216	-99.27344	Mexicano (Liverpool Gourmet)	19.361587	-99.274490	Mexican Restaurant
2	Lomas de Chapultepec I Sección	19.36216	-99.27344	Centro Santa Fe	19.361314	-99.273878	Shopping Mall
3	Lomas de Chapultepec I Sección	19.36216	-99.27344	Liverpool	19.361531	-99.274371	Department Store
4	Lomas de Chapultepec I Sección	19.36216	-99.27344	Distroller Santa Fé	19.361183	-99.273713	Toy / Game Store

With the whole table we were able to define 253 categories of venues in the neighborhoods from the borough Miguel Hidalgo. One interesting insight we found in this table is that the Top category in one of the neighborhoods **was Mexican restaurant**.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
count	2956	2956.000000	2956.000000	2956	2956.000000	2956.000000	2956
unique	85	NaN	NaN	1690	NaN	NaN	253
top	Campo Militar 1	NaN	NaN	Starbucks	NaN	NaN	Mexican Restaurant

We created too a table to identify the Top 10 venues for each neighborhood. We showed a sample of such a table as well as a table with descriptive statistics. Once again we found that in the top od the 10 most common venues the Mexican Restaurant Category appear as the 1st and 2nd most common venue, followed by Café, Taco Place, Pharmacy, Burger Joint, Pizza Place , Ice Cream Shop and Bakery. So We clearly can stated that Miguel Hidalgo Neighborhoods can be considered as a borough where food sector is quite relevant

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 10 de Abril	Convenience Store	Mexican Restaurant	Taco Place	Pharmacy	Food Truck	Burger Joint	Park	Pizza Place	Café	Sandwich Place
1 16 de Septiembre	Fast Food Restaurant	Farmers Market	Health & Beauty Service	Pool	Lounge	Dry Cleaner	Yoga Studio	Electronics Store	Donut Shop	Dine-in Theater
2 5 de Mayo	Mexican Restaurant	Taco Place	Burger Joint	Gym	Coffee Shop	Breakfast Spot	Dessert Shop	Snack Place	Food & Drink Shop	Metro Station
3 Agricultura	Restaurant	Breakfast Spot	Ice Cream Shop	Coffee Shop	Food Truck	Gym	Snack Place	Electronics Store	Mexican Restaurant	Burger Joint
4 Ahuehuetes Aduhuac	Taco Place	Mexican Restaurant	Convenience Store	Bakery	Breakfast Spot	Farmers Market	Bar	Coffee Shop	Shipping Store	Seafood Restaurant

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
count	85	85	85	85	85	85	85	85	85	85
unique	85	20	27	36	43	43	48	45	49	55
top	Ampliación Popo	Mexican Restaurant	Mexican Restaurant	Café	Taco Place	Pharmacy	Burger Joint	Pizza Place	Mexican Restaurant	Ice Cream Shop
freq	1	27	17	7	10	7	6	6	6	7

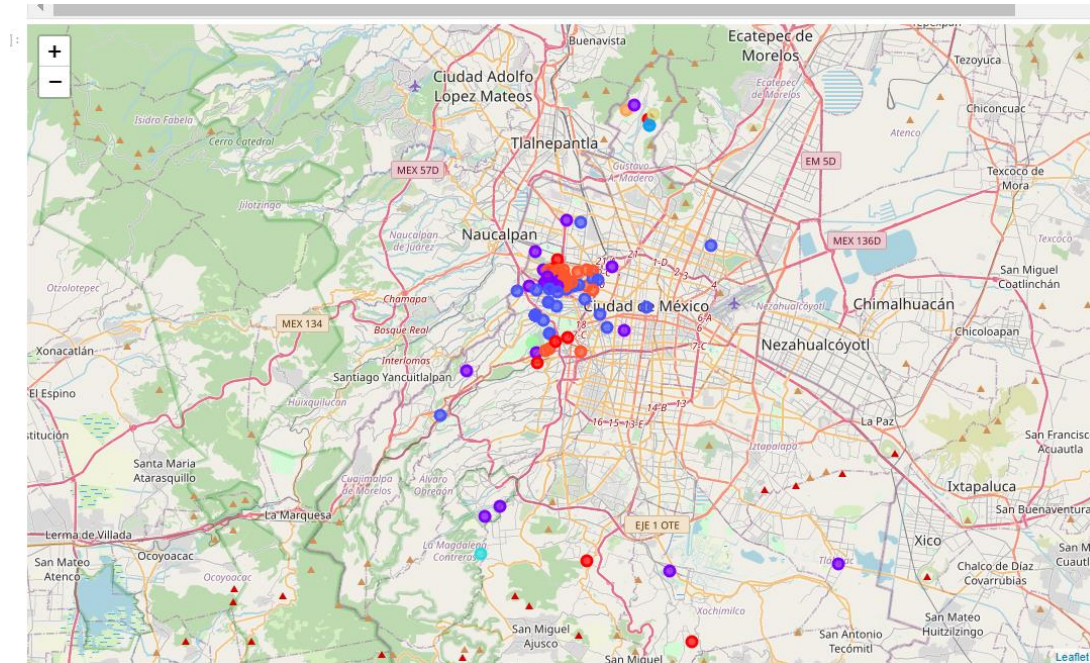
Machine Learning algorithm used

Because we are looking to understand how the neighborhoods in Miguel Hidalgo borough can be grouped by clusters, we will use the K means machine learning algorithm.

K-Means can group information/entities into **clusters** that have similarities and are dissimilar to the entities belonging to another cluster. The term 'K' refers to the number of clusters we want to create to our data. Given that we have 253 venues categories, we have defined K= 9.

Results section

Once we ran the K means ML algorithm, we got 9 clusters:



Cluster 1: 8 neighborhoods where the most common place where: Taco Place, Mexican Restaurant,

Cluster 2: 20 neighborhoods with Mexican Restaurant, Taco Place, Plaza, restaurant, cafe

Cluster 3: 28 neighborhoods with coffeeshops, **Seafood restaurants**, Tacos, ice creams, bakeries

Cluster 4: 2 neighborhoods with Farmers market, mountain, health and beauty services, pool, Falafel Restaurant, Eastern European restaurant

Cluster 5: 1 neighborhood with farm, creperie, pharmacies

Cluster 6: 2 neighborhoods with soccer fields, forests, **Seafood restaurant**, Casino, Yoga, Drive-in Theater

Cluster 7: 1 neighborhood with snack place and shopping mall, sake bar, yoga

Cluster 8: 1 neighborhood with Mountain, housing department, food court park, cycle studio

Cluster 9: 21 neighborhoods with Taco place, Mexican Restaurant, coffee shops, restaurants

Discussion

We have 3 possible recommendations

- 1) **Seafood cluster:** A feasible option is to open the restaurant in the **cluster 3** where Seafood restaurants are the 2nd most common place. In 8 of the neighborhoods there are seafood restaurants, so there is a potential location in the neighborhoods of this cluster. People in business might call this cluster as *Red Ocean*, where a fierce competency by price, quality, differentiation happens day by day. Of course, this implies to have a differentiation business strategy to overcome the competitors in the cluster.
- 2) **Mexican restaurants:** The **cluster 2** where most restaurants are Mexicans can be an opportunity to open a seafood restaurant, because there are only a few competitors with seafood, and it can represent an opportunity to find and to attract a potential market of customers who look for alternatives to Mexican food restaurants
- 3) **Clusters without restaurants or minimal food vendors:** This can be an interesting alternative given that **Cluster 9** does not have Seafood as a common place in its neighborhoods, however there are Taco place, Mexican Restaurant, coffee shops, restaurants. People in business discipline might take this cluster as a *Blue Ocean*, but it does imply further research to find if there are potential consumers .

Conclusion

Our project provides important input that can be integrated into Market feasibility study for a Seafood restaurant.

Marketing specialists now can focus their market analysis on the clusters of neighborhoods recommended in our project for Miguel Hidalgo borough.

Market are population as well as socioeconomic lifestyle analysis can be integrated with the clusters data we defined and find clusters of potential customers that can provide yet more information.

Further research should be done to complete a whole feasibility study and having a possible location from this project is a key step to start with a clear focus and perspective of the food sector in Miguel Hidalgo borough.