Mama’s Bakery Franchise

INTRODUCTION / BUSINESS PROBLEM

The group of inversion BakeryArmy has interest in making a franchise out of the products of a local bakery. Its main products are cakes and cookies that have a great reception at a local level.

BakeryArmy is willing to invest in opening several stores all around Mexico City's suburbs. In order to assure its success, we have to do an analysis of the best locations around the center of the 1812 suburbs of the city considering the nearest direct competitors.

DATA

One of the best indicators to assure the success of a local business it's to check for local competitors.

First of all, we will use a public file of the locations of all the suburbs of Mexico City as reference. This file is available at: https://datos.cdmx.gob.mx/explore/dataset/coloniascdmx/download/?format=csv&timezone=America/Mexico_City&lang=es&use_labels_for_header=true&csv_separator=%2C

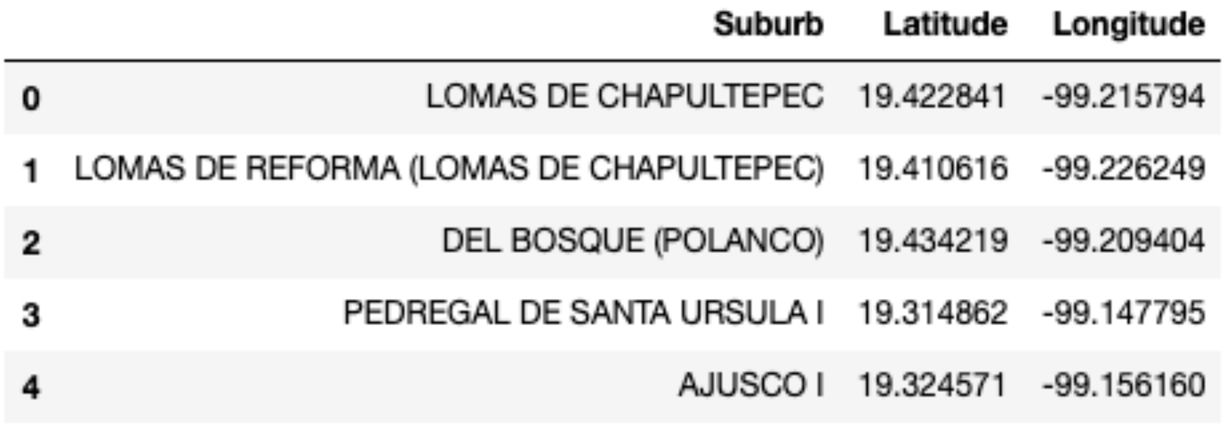
This file will help us perform an analysis with geo data using Foursquare. We will use the center of each suburb to check for the number of establishments that represent a direct competition to our franchise.

As from the possible Foursquare Venue Categories (documented at: https://developer.foursquare.com/docs/build-with-foursquare/categories/) we will consider as direct competition the categories: Bakery, Cafeteria, Creperie and Donut Shop.

The final objetive of the data analysis is to provide a quantitative evaluation of the best suburbs (considering them from their geographical center) to have insight on where to open franchise establishments.

METHODOLOGY

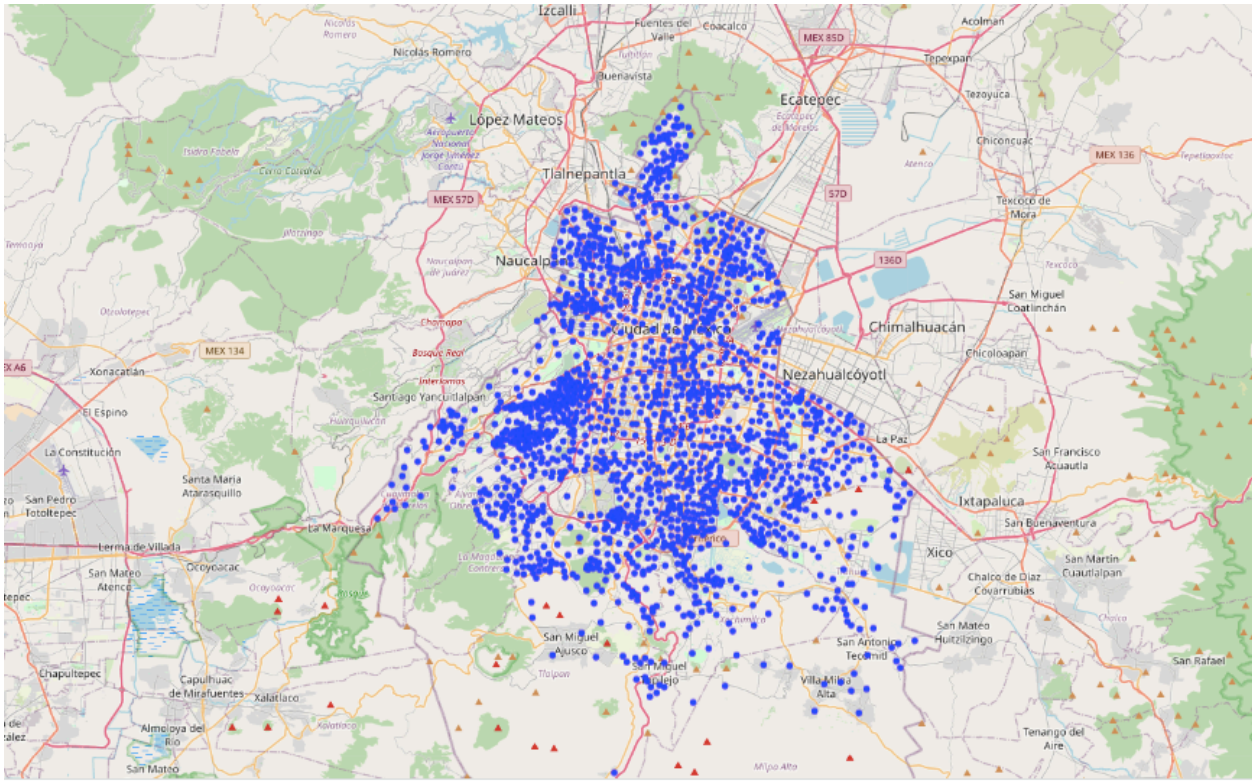
After getting the suburbs’ data into a dataframe, we had to change the column names to English, split the geo data in order to have a ‘Latitude’ and a ‘Longitude’ column, and drop the unnecessary columns. As a result we have a table with a structure like the one below.



	Suburb	Latitude	Longitude
0	LOMAS DE CHAPULTEPEC	19.422841	-99.215794
1	LOMAS DE REFORMA (LOMAS DE CHAPULTEPEC)	19.410616	-99.226249
2	DEL BOSQUE (POLANCO)	19.434219	-99.209404
3	PEDREGAL DE SANTA URSULA I	19.314862	-99.147795
4	AJUSCO I	19.324571	-99.156160

Also, it was needed to drop some information with NaNs. This information did not represent a considerable loss of data.

We use this dataframe to obtain a Mexico's City map with these points.

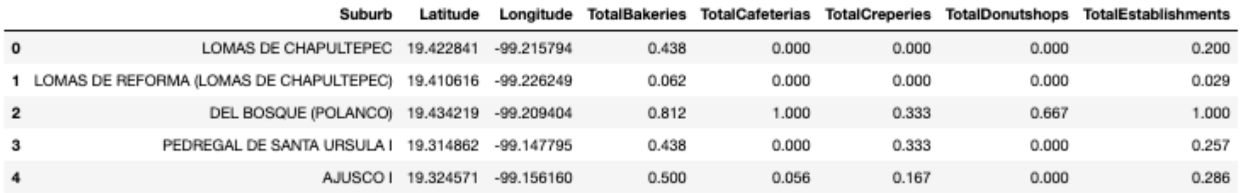


FOURSQUARE DATA ACQUISITION

First of all, we needed to set up the Foursquare credentials in order to make use of their API. After we done this, we use the SEARCH Endpoint to search for the number 4 specific competitors around a 1km radius using their VenueId provided by Foursquare documentation.

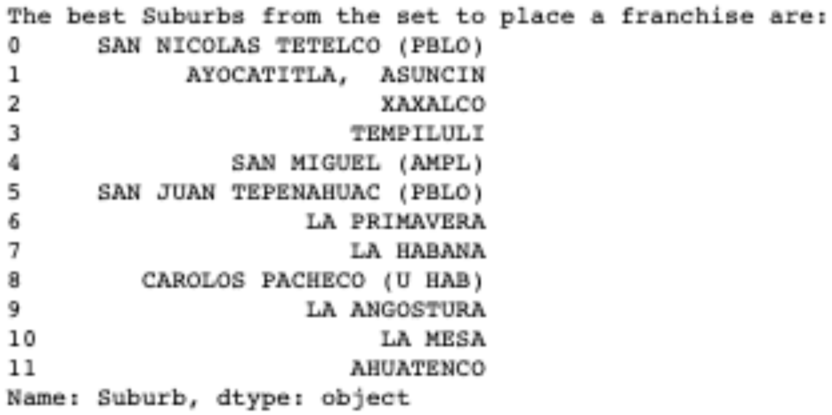
At this point the project was limited to the analysis of 100 suburbs because of the max requests limit of the free subscription. Nevertheless, the project is scalable if we count with a Premium account.

After processing the results and putting them into a dataframe, we sum every kind of establishment obtained per suburb in order to create a 'Total Establishment'. After we have done this, we will scale the values of every kind of stablishments in the table using a MinMaxScaler in order to have a better understanding of the data. We get a table with a structure like the one below.



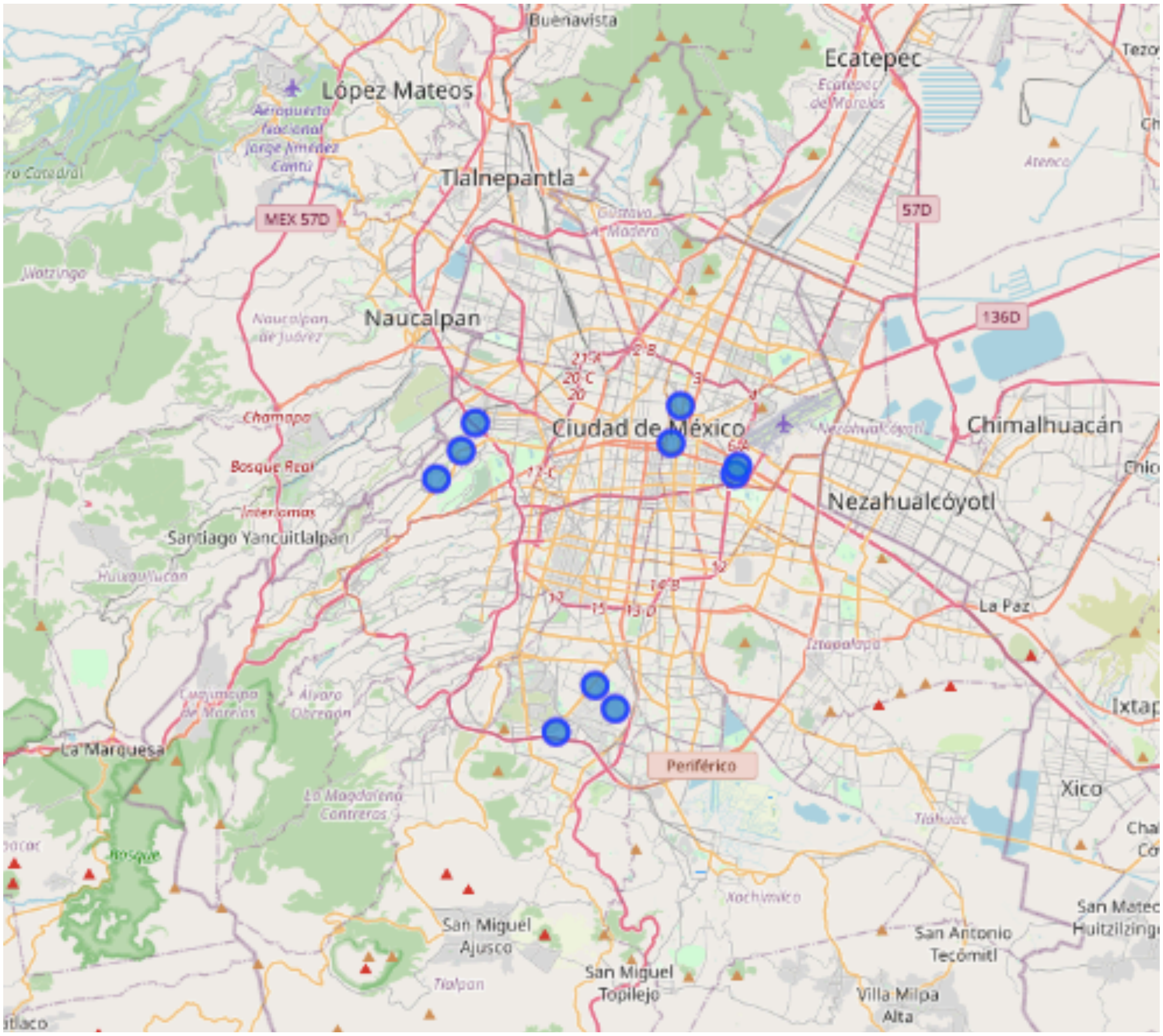
	Suburb	Latitude	Longitude	TotalBakeries	TotalCafeterias	TotalCreperies	TotalDonutshops	TotalEstablishments
0	LOMAS DE CHAPULTEPEC	19.422841	-99.215794	0.438	0.000	0.000	0.000	0.200
1	LOMAS DE REFORMA (LOMAS DE CHAPULTEPEC)	19.410616	-99.226249	0.062	0.000	0.000	0.000	0.029
2	DEL BOSQUE (POLANCO)	19.434219	-99.209404	0.812	1.000	0.333	0.667	1.000
3	PEDREGAL DE SANTA URSULA I	19.314862	-99.147795	0.438	0.000	0.333	0.000	0.257
4	AJUSCO I	19.324571	-99.156160	0.500	0.056	0.167	0.000	0.286

We use the 'TotalEstablishments' column as the main characteristic to determine the best locations to place our franchise, being 0 the best value and 1 the worst. With this consideration we get our top 10 places:



```
The best Suburbs from the set to place a franchise are:
0     SAN NICOLAS TETELCO (PBLO)
1           AYOCATITLA,  ASUNCIN
2                        XAXALCO
3                      TEMPILULI
4               SAN MIGUEL (AMPL)
5     SAN JUAN TEPENAHUAC (PBLO)
6                   LA PRIMAVERA
7                      LA HABANA
8         CAROLOS PACHECO (U HAB)
9                   LA ANGOSTURA
10                       LA MESA
11                     AHUATENCO
Name: Suburb, dtype: object
```

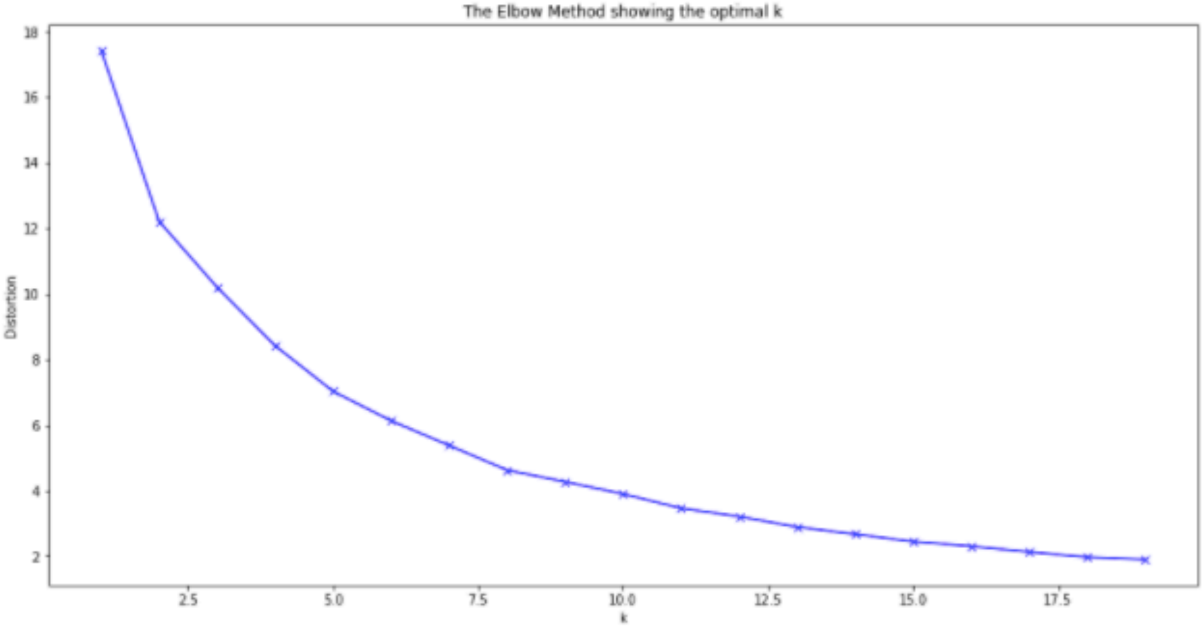
And we plot them in a map



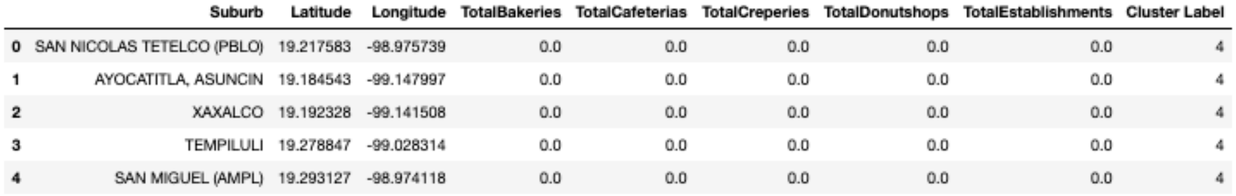
At this point, we have successfully found the best locations for our franchise using the density of local competitors in an area, but, what about if there is a particular kind of competitor that would represent a stronger or weaker competition?

We will use a Kmeans clustering method in order to classify the kind of suburbs based on their local competitors distribution.

We run an Elbow Method analysis using values for k from 1 to 20 and we get the graph below

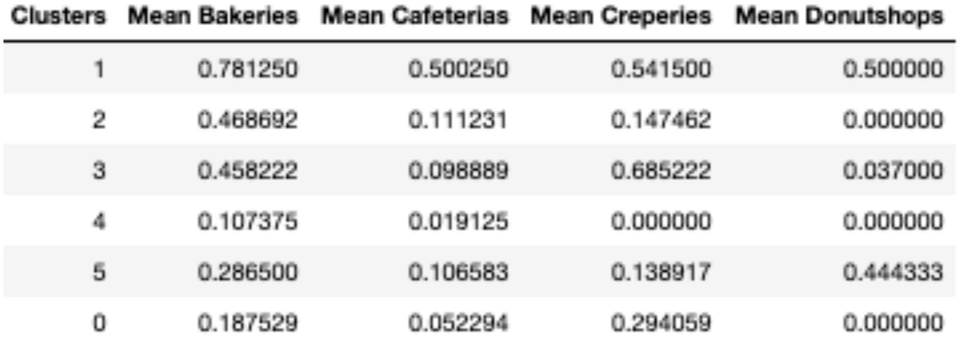


We can observe a good value of k at 6. We use this value to train our model and add the cluster label to our dataframe to obtain a table with the following structure



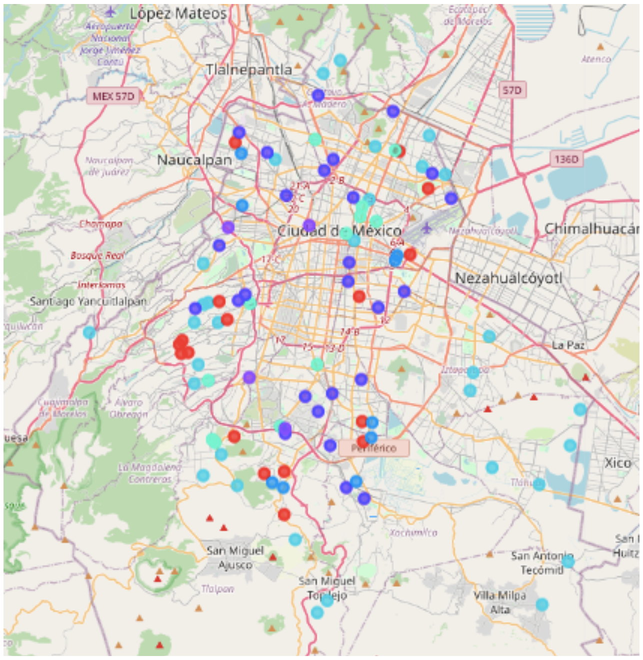
	Suburb	Latitude	Longitude	TotalBakeries	TotalCafeterias	TotalCreperies	TotalDonutshops	TotalEstablishments	Cluster Label
0	SAN NICOLAS TETELCO (PBLO)	19.217583	-98.975739	0.0	0.0	0.0	0.0	0.0	4
1	AYOCATITLA, ASUNCIN	19.184543	-99.147997	0.0	0.0	0.0	0.0	0.0	4
2	XAXALCO	19.192328	-99.141508	0.0	0.0	0.0	0.0	0.0	4
3	TEMPILULI	19.278847	-99.028314	0.0	0.0	0.0	0.0	0.0	4
4	SAN MIGUEL (AMPL)	19.293127	-98.974118	0.0	0.0	0.0	0.0	0.0	4

In order to better visualize the characteristic of each cluster, we created a descriptive dataframe:



Clusters	Mean Bakeries	Mean Cafeterias	Mean Creperies	Mean Donutshops
1	0.781250	0.500250	0.541500	0.500000
2	0.468692	0.111231	0.147462	0.000000
3	0.458222	0.098889	0.685222	0.037000
4	0.107375	0.019125	0.000000	0.000000
5	0.286500	0.106583	0.138917	0.444333
0	0.187529	0.052294	0.294059	0.000000

And plot each cluster in a map



• RESULTS AND DISCUSSION

Our analysis shows that there are some suburb centers in Mexico City that do not count with direct competitors (considering a 1km radius area). These are the areas that are consider to be the best to place our franchise. If, for some reason, there is the need to consider other areas, the general recommendation is to use the TotalEstablishments variable to refer other areas, the lower value, the better recommendation.

As the analysis was done, we notice that in the near future, we will notice that among these direct competitors, there would be some that represent a stronger competition than others, so as a base for future analysis, we created clusters based on the density per kind of competitors around the area in order to consider other suburbs for future franchises. In this part of the analysis we notice that the best k for clustering is 6. Through the visualization of the clusters' density of main competitors we can take future decisions on where to place other franchises.

- **CONCLUSION**

The purpose of this project was to identify the best suburbs in Mexico City to place a bakery, this objective was partially limited because the Foursquare API's free suscription didn't allow more requests. Nevertheless, the code supports the iteration of the complete dataset in case of having a Premium suscription.

Apart from this, the project allows us to take the decission on where we can place a franchise based on the density of local competitors around the area.