Clustering neighborhoods in Mexico City according to venue type, population density and income level for a CPG retail business

Rogelio Antonio Melo Juarez
Mexico City, Mexico
2019 – 11 – 09

Table of Contents

1.0 Introduction

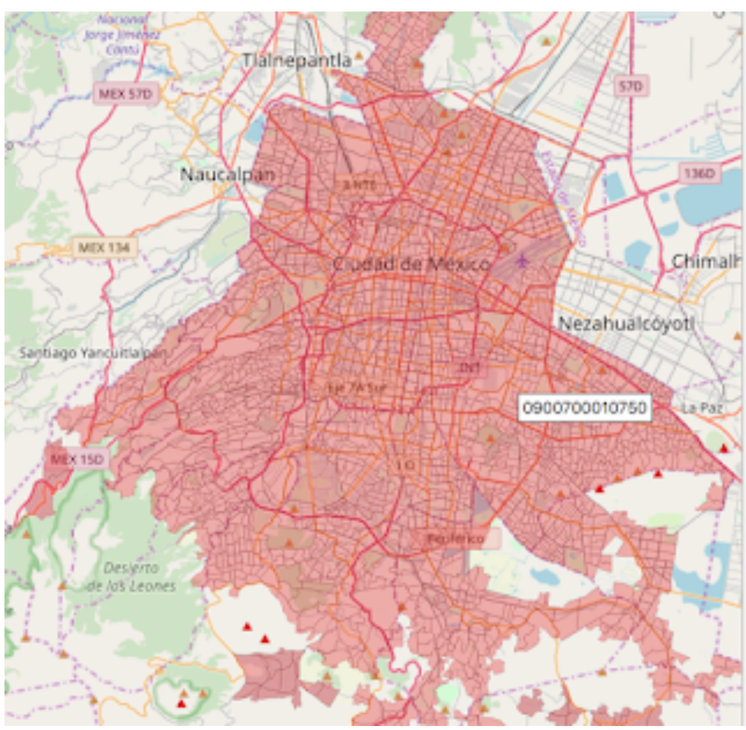
Opening a business in a city with which you're not familiar with can be a tricky decision. However if a product or market is not exploited in this new location, it could represent a profitable opportunity for the people opening it. Leveraging Foursquare data, the objective of this investigation is to cluster neighborhoods in the district of Álvaro Obregón in Mexico City in order to produce a trustworthy guide for people wanting to open a consumer packaged goods retail business in one of Mexico City's richest districts. We will match public geo referenced income data published by the INEGI, which is Mexico's statistics and geography institute, with the information extracted from the Foursquare API to obtain lists of Mexico City's neighborhoods in Álvaro Obregón which are similar to each other.

2.0 Data

2.1 I used Foursquare API's location data for Álvaro Obregón District in Mexico City regarding to the retail venues present in each neighborhood. These could be all types of consumer packaged goods distributors, such as convenience stores, bakeries, butcher's, supermarkets, etc.

2.2

In order to establish a relation with socioeconomic indicators I used a shape file provided by Mexico's Government which segments Mexico City in polygons named AGEBs. Each AGEB has a code which can be matched with several databases containing demographic, economic, social, etc. indicators.



2.3

Sample of the data containing each ageb key with its total population and socioeconomic indicator.

ageb_key	ageb_folium	total_population	socioeconomic_indicator
12	901000000000	1775	125.4609283
27	901000000000	2565	121.4501244
31	901000000000	3604	183.2671467
46	901000000000	5634	248.0198202
50	901000000000	3894	205.6854103
65	901000000000	5788	217.5270042
84	901000000000	2844	182.0973258
99	901000000000	4239	211.7606613
101	901000000000	1531	122.530567
116	901000000000	4258	223.0811856
135	901000000000	10874	116.8173039
14987	090100001014A	4571	179.3843867
169	901000000000	3008	205.038428
173	901000000000	4751	185.1927778
188	901000000000	666	92.01609724
192	901000000000	742	137.0144152

2.4

I used another database which lists the coordinates of each neighborhood in Mexico City. The next image is a sample of the data containing each of Mexico City's neighborhood with its geo point.

NEIGHBORHOOD	ENTITY	Geo Point	DISTRICT
LOMAS DE REFORMA	9	19.4016815485	MIGUEL HIDALGO
DANIEL GARZA (AMPL)	9	19.4092184712	MIGUEL HIDALGO
IGNACIO MANUEL ALTAMIRANO	9	19.4631440514	MIGUEL HIDALGO
LEGARIA	9	19.4555305454	MIGUEL HIDALGO
LEGARIA (U HAB)	9	19.450020036	MIGUEL HIDALGO
ADOLFO LOPEZ MATEOS	9	19.4200520013	VENUSTIANO CARRANZA
ADOLFO RUIZ CORTINES I	9	19.3209669132	COYOACAN
PEDREGAL DE STO DOMINGO III	9	19.3314325978	COYOACAN
PASEOS DE TAXQUEA I	9	19.3482555644	COYOACAN
PROGRESISTA	9	19.4360921322	VENUSTIANO CARRANZA
SEVILLA	9	19.4107331904	VENUSTIANO CARRANZA
MIGUEL HIDALGO	9	19.4359898482	VENUSTIANO CARRANZA
TORRE BLANCA	9	19.4574907732	MIGUEL HIDALGO
UN HOGAR PARA NOSOTROS	9	19.4484020023	MIGUEL HIDALGO
PENSIL SUR	9	19.4465001414	MIGUEL HIDALGO
ALVARO OBREGON	9	19.4127921222	VENUSTIANO CARRANZA
JAMAICA	9	19.4069270279	VENUSTIANO CARRANZA
EMILIO CARRANZA	9	19.4493647246	VENUSTIANO CARRANZA
CTM VIII CULHUACAN (U HAB)	9	19.3161921597	COYOACAN
INFONAVIT CULHUACAN ZONA 2 (U HAB)	9	19.3298781623	COYOACAN
CROC CULHUACAN SECC 6 (U HAB)	9	19.3275552317	COYOACAN
LAS CAMPANAS	9	19.3094554092	COYOACAN
IMAN	9	19.3079335787	COYOACAN
LOS OLIVOS (FRACC)	9	19.3136569749	COYOACAN
PRESIDENTES EJIDALES PRIMERA SECCION	9	19.3268052641	COYOACAN
MEDIA LUNA	9	19.3018815416	COYOACAN
SAN DIEGO CHURUBUSCO	9	19.3557844724	COYOACAN
SANTA MARTHA DEL SUR	9	19.341178323	COYOACAN

2.5

This table contains a merge which includes neighborhood location, venue location and ageb data.

Neighborhood	Latitude	Longitude	Venue	Venue Lati	Venue Long	Venue Cate	CVEGEO	CVE_AGEB
2DA JALALPA TEPITO (AMPL)	19.3756547081	-99.2332071969	Total Body fitness center	19.3772446033117	-99.2311347188103	Gym	901000010440	440
2DA JALALPA TEPITO (AMPL)	19.3756547081	-99.2332071969	Tlacoyos "polo"	19.3777238626049	-99.2311435112724	Fast Food Restaurant	901000010440	440
2DA JALALPA TEPITO (AMPL)	19.3756547081	-99.2332071969	Campo De Futbol Jalalpa	19.3746241284926	-99.2371767225752	Soccer Field	901000010440	440
2DA JALALPA TEPITO (AMPL)	19.3756547081	-99.2332071969	Carnitas Nacho	19.371821	-99.230726	Taco Place	901000010440	440
ESTADO DE HIDALGO_	19.3644436827	-99.2416810868	Starbucks	19.3650022013788	-99.2439625947752	Coffee Shop	901000010667	667
ESTADO DE HIDALGO_	19.3644436827	-99.2416810868	Carnitas Tamaulipas 1190	19.3652835056182	-99.2438875787432	Taco Place	901000010667	667
ESTADO DE HIDALGO_	19.3644436827	-99.2416810868	La Salle, Unidad Deportiva Santa Lucia	19.3662590790739	-99.2403639551873	Soccer Field	901000010667	667
ESTADO DE HIDALGO_	19.3644436827	-99.2416810868	El Lago de los Patos	19.3645214045616	-99.2453450266458	Taco Place	901000010667	667
ESTADO DE HIDALGO_	19.3644436827	-99.2416810868	La Barda	19.3646028373147	-99.2452345105371	Burger Joint	901000010667	667
ESTADO DE HIDALGO_	19.3644436827	-99.2416810868	Booth Nespresso (Centro Comercial Santa Fe)	19.3652243175189	-99.2441032967219	Café	901000010667	667
SAN CLEMENTE	19.3493092254	-99.2262178725	Santa Clara	19.3490669711245	-99.2264168425186	Ice Cream Shop	090100001181A	181A
SAN CLEMENTE	19.3493092254	-99.2262178725	Hipico Las Águilas	19.3494198208458	-99.2241947098318	Farm	090100001181A	181A
SAN CLEMENTE	19.3493092254	-99.2262178725	Little Caesars Pizza	19.3498849711743	-99.2228822080092	Pizza Place	090100001181A	181A
SAN CLEMENTE	19.3493092254	-99.2262178725	Barbacoa De Santiago	19.3490051968051	-99.2264293426409	Taco Place	090100001181A	181A
SAN CLEMENTE	19.3493092254	-99.2262178725	El Globo	19.3499497298817	-99.2277820056267	Bakery	090100001181A	181A
SAN CLEMENTE	19.3493092254	-99.2262178725	Farmacia del Ahorro	19.3497132128812	-99.2236601141008	Drugstore	090100001181A	181A
SAN CLEMENTE	19.3493092254	-99.2262178725	Alberca La Cuesta	19.345806002625	-99.225236833381	Pool	090100001181A	181A
SAN CLEMENTE	19.3493092254	-99.2262178725	Mercado De Los Domingos	19.3496535897365	-99.2233376117897	Food Court	090100001181A	181A
SAN CLEMENTE	19.3493092254	-99.2262178725	BBVA Bancomer	19.3481166408783	-99.2307628371884	ATM	090100001181A	181A
SAN CLEMENTE	19.3493092254	-99.2262178725	Kazoku Maki	19.3489309397443	-99.2267890713537	Sushi Restaurant	090100001181A	181A
SAN CLEMENTE	19.3493092254	-99.2262178725	Las Criminales	19.3500074180227	-99.22188942675	Burger Joint	090100001181A	181A
SAN CLEMENTE	19.3493092254	-99.2262178725	Sky-Gym Family Fitness Center Las Aguilas	19.3487534954304	-99.2283452160863	Gym	090100001181A	181A
SAN CLEMENTE	19.3493092254	-99.2262178725	7- Eleven	19.3498291148765	-99.2230700951931	Convenience Store	090100001181A	181A
SAN CLEMENTE	19.3493092254	-99.2262178725	7- Eleven	19.3483602769489	-99.2300453400378	Convenience Store	090100001181A	181A
SAN CLEMENTE	19.3493092254	-99.2262178725	Plaza Carpali	19.3491127495023	-99.2264265633497	Shopping Mall	090100001181A	181A
SAN CLEMENTE	19.3493092254	-99.2262178725	Steren Las Aguilas	19.3481973700373	-99.2298889160156	Electronics Store	090100001181A	181A
SAN CLEMENTE	19.3493092254	-99.2262178725	Sky Gym - Aguilas	19.3488294083718	-99.228401446626	Gym	090100001181A	181A

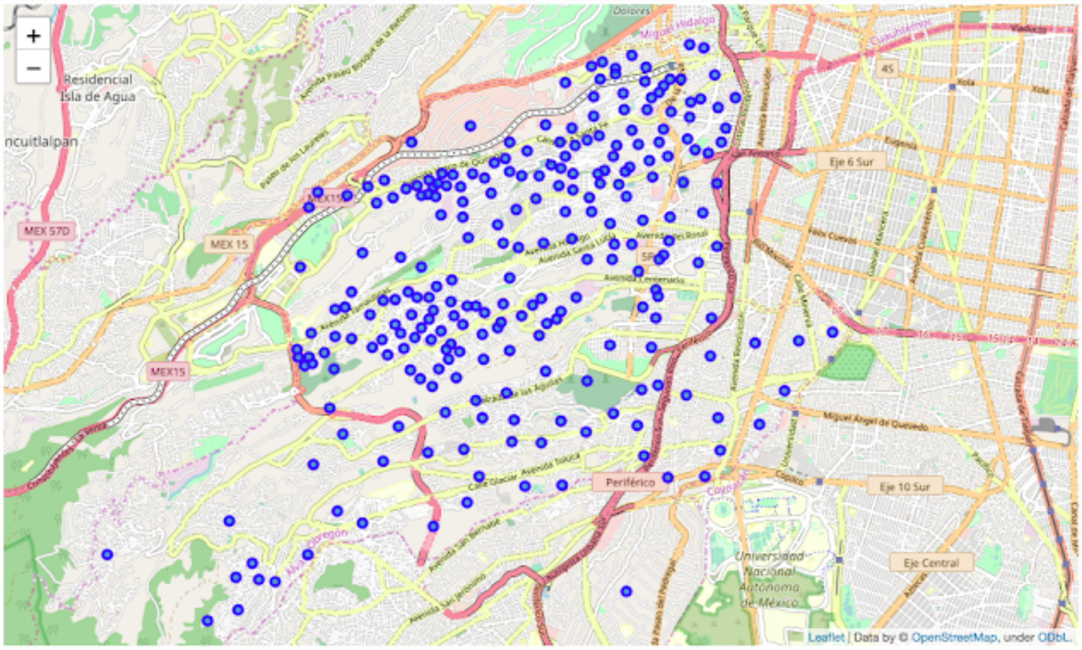
3.0 Methodology

3.1 Foursquare data

First, I used the Mexico City's neighborhoods table and did a visual analysis of the locations of interest. In my case, I chose to narrow my exercise to the district of Álvaro Obregón. The sample of the table looks as follows.

	District	Neighborhood	Latitude	Longitude
0	ALVARO OBREGON	2DA JALALPA TEPITO (AMPL)	19.375655	-99.233207
1	ALVARO OBREGON	ESTADO DE HIDALGO_	19.364444	-99.241681
2	ALVARO OBREGON	SAN CLEMENTE	19.349309	-99.226218
3	ALVARO OBREGON	VILLA SOLIDARIDAD	19.383402	-99.219292
4	ALVARO OBREGON	TEPOPOTLA	19.362128	-99.251763

Afterwards, I used the Folium python library to represent the geo points in the table indicating each neighborhood.



Once I concluded the first visual analysis, I used the Foursquare API to get the venues located in each neighborhood. I defined the limit of results to **100 venues** and at a distance not greater than **500 meters** of each geo point.

The following dataset containing the venue's latitude and longitude data, the venue category and name and the neighborhood in which it is located.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	2DA JALALPA TEPITO (AMPL)	19.375655	-99.233207	Campo De Futbol Jalalpa	19.374624	-99.237177	Soccer Field
1	2DA JALALPA TEPITO (AMPL)	19.375655	-99.233207	La Michoacana	19.371645	-99.232556	Ice Cream Shop
2	2DA JALALPA TEPITO (AMPL)	19.375655	-99.233207	El Duende	19.371489	-99.231661	Scenic Lookout
3	2DA JALALPA TEPITO (AMPL)	19.375655	-99.233207	Carnitas Nacho	19.371821	-99.230726	Taco Place
4	ESTADO DE HIDALGO_	19.364444	-99.241681	Starbucks	19.365002	-99.243963	Coffee Shop
5	ESTADO DE HIDALGO_	19.364444	-99.241681	Carnitas Tamaulipas 1190	19.365284	-99.243888	Taco Place
6	ESTADO DE HIDALGO_	19.364444	-99.241681	La Salle, Unidad Deportiva Santa Lucia	19.366259	-99.240364	Soccer Field

I found that in the whole data there where 245 unique categories found. Next, I created a table which sorts each neighborhood's most common venues.

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	19 DE MAYO	Convenience Store	Burger Joint	Taco Place	Donut Shop	Garden
1	1RA VICTORIA	Taco Place	Mexican Restaurant	Burger Joint	Park	Pharmacy
2	1RA VICTORIA SECCION BOSQUES	Taco Place	Park	Gym Pool	Garden	Bar
3	26 DE JULIO	Coffee Shop	Convenience Store	Pharmacy	Taco Place	Mexican Restaurant
4	2DA JALALPA TEPITO (AMPL)	Soccer Field	Ice Cream Shop	Taco Place	Scenic Lookout	Fabric Shop

3.2 Mexico City's socioeconomic and demographic data

To match the venues and neighborhood's data with the socioeconomic indicators I first used an open source tool called Qgis, which lets you intersect shapefiles with geo points. That's how I obtained a dataset which matches each venue with the AGEB it's located in. When I have the AGEB for every venue I do another merge with the indicators database which has the key of all the AGEBs in Mexico City.

First, I matched each venue with the AGEB in which it is located.

	Neighborhood	Latitude	Longitude	Venue	Venue Lati	Venue Long	Venue Cate	CVEGEO	CVE_AGEB
0	2DA JALALPA TEPITO (AMPL)	19.375655	-99.233207	Total Body fitness center	19.377245	-99.231135	Gym	0901000010440	0440
1	2DA JALALPA TEPITO (AMPL)	19.375655	-99.233207	Tlacoyos "polo"	19.377724	-99.231144	Fast Food Restaurant	0901000010440	0440
2	2DA JALALPA TEPITO (AMPL)	19.375655	-99.233207	Campo De Futbol Jalalpa	19.374624	-99.237177	Soccer Field	0901000010440	0440
3	2DA JALALPA TEPITO (AMPL)	19.375655	-99.233207	Carnitas Nacho	19.371821	-99.230726	Taco Place	0901000010440	0440
4	ESTADO DE HIDALGO_	19.364444	-99.241681	Starbucks	19.365002	-99.243963	Coffee Shop	0901000010667	0667

Second, using the socioeconomic indicators database I matched [CVE_AGEB = ageb_key].

	ageb_key	ageb_folium	total_population	socioeconomic_indicator
0	12	9.01E+11	1775	125.460928
1	27	9.01E+11	2565	121.450124
2	31	9.01E+11	3604	183.267147
3	46	9.01E+11	5634	248.019820
4	50	9.01E+11	3894	205.685410

Finally, I obtained a dataset that looks like this:

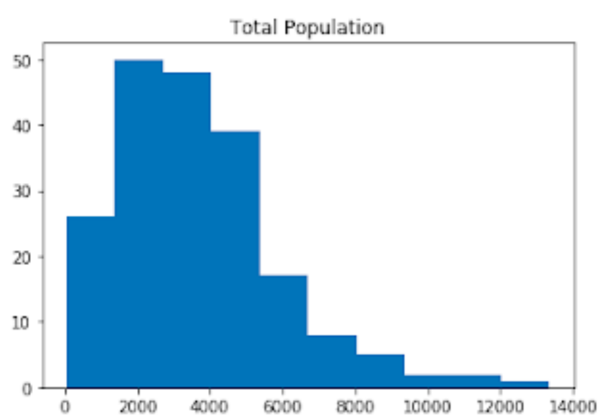
	total_population	socioeconomic_indicator	Neighborhood	Latitude	Longitude	Venue	Venue Lati	Venue Long	Venue Cate
0	1775	125.460928	COVE	19.401334	-99.197524	Paleteria y Neveria Lupita	19.401818	-99.199235	Ice Cream Shop
1	1775	125.460928	COVE	19.401334	-99.197524	Las Bolas de Pepe	19.403010	-99.197648	Beer Bar
2	1775	125.460928	COVE	19.401334	-99.197524	Casa Ikeda	19.402740	-99.198812	Furniture / Home Store
3	1775	125.460928	COVE	19.401334	-99.197524	El Surtidor	19.402280	-99.200130	Furniture / Home Store
4	1775	125.460928	COVE	19.401334	-99.197524	Quecas De Lucha	19.400765	-99.197172	Mexican Restaurant

Since I wanted to make my data on income and population more understandable, I transformed it to dummy variables which indicate a certain segment of population density and income level.

The dataset looks something like this:

	total_population	socioeconomic_indicator	Neighborhood	Latitude	Longitude	Venue
0	low_pop_density	high_socio_econ_lvl	COVE	19.401334	-99.197524	Paleteria y Neveria Lupita
1	low_pop_density	high_socio_econ_lvl	COVE	19.401334	-99.197524	Las Bolas de Pepe
2	low_pop_density	high_socio_econ_lvl	COVE	19.401334	-99.197524	Casa Ikeda
3	low_pop_density	high_socio_econ_lvl	COVE	19.401334	-99.197524	El Surtidor
4	low_pop_density	high_socio_econ_lvl	COVE	19.401334	-99.197524	Quecas De Lucha

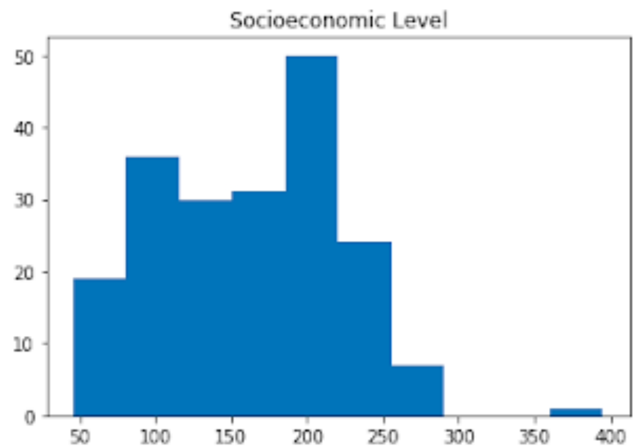
I performed two more visual analysis regarding the distributions of income and population across neighborhoods. I found that the most neighborhoods have a population between 2000 and 5000 inhabitants and that most neighborhoods have a medium socioeconomic level.



To analyze the data with categorical variables, I divided the population quantities into bins:

- 0 - 5000 inhabitants: “Low population density"
- 5000–10,000 inhab. : “Medium population density”

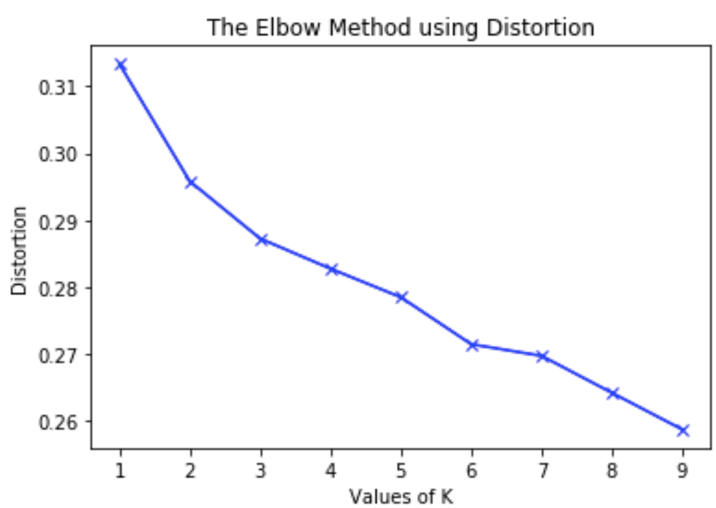
- 10,000–15,000 inhab. : “High population density”



- 0 - 150: “High socioeconomic level"
- 150 – 200 : “Medium socioeconomic level"
- 250 – 400 : “Low socioeconomic level"

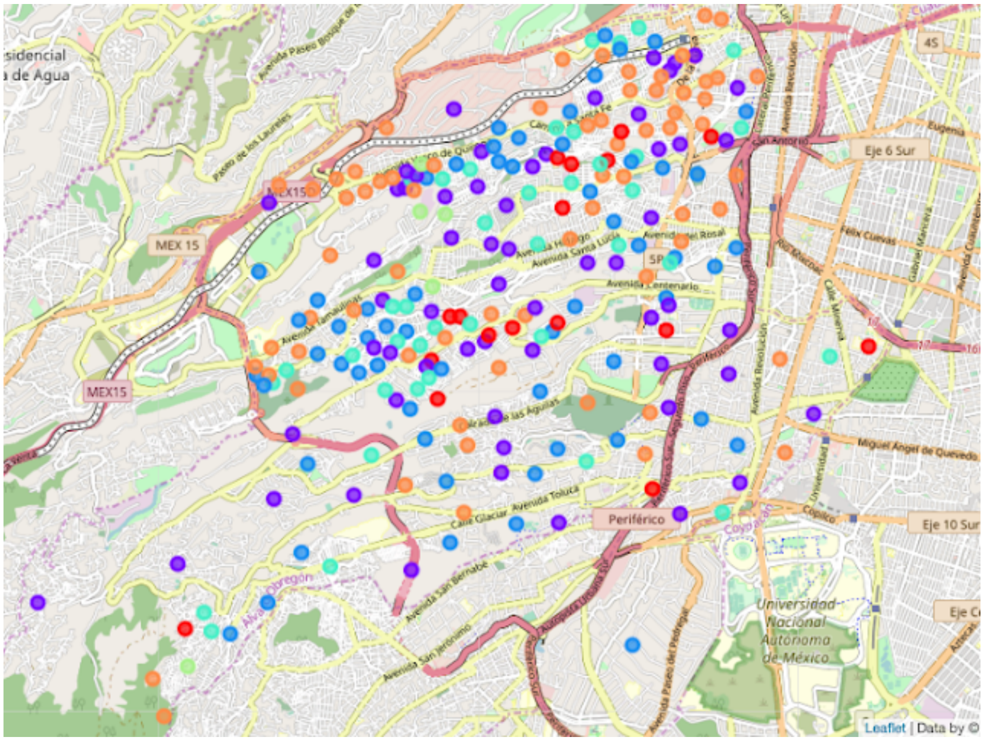
3.3 Clustering

Since neighborhoods share venue categories and indicators I decided to use K-means clustering but first I ran a validation to select the optimal K using the elbow method. I decided to use k=6 because it's a point where the slope of the curve tragically flattens, even though it keeps decreasing.



4.0 Results:

Using the Folium library, I made an Álvaro Obregón District map showing its neighborhoods grouped by cluster labels.



Cluster 0 (Red dots) :

I called cluster 0 the "Italian cluster" since the most common venue in almost all of its neighborhoods is a "Pizza place". Also, it's characterized by a predominantly low population density in the Álvaro Obregón district and by a balanced medium - high socioeconomic level.

Most common venues:

Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
ARCOS DE CENTENARIO	Pizza Place	Mexican Restaurant	Burger Joint	Stationery Store	Coffee Shop
AVE REAL (AMPL)	Burger Joint	Gym	Flea Market	Pizza Place	Scenic Lookout
CANUTILLO 3RA SECCION	Pizza Place	Mexican Restaurant	Burger Joint	Stationery Store	Coffee Shop
CANUTILLO PREDIO LA PRESA	Pizza Place	Taco Place	Pharmacy	Mexican Restaurant	Fabric Shop
CANUTILLO(AGUASCALIENTES)	Pizza Place	Burger Joint	Mexican Restaurant	Zoo Exhibit	Embassy / Consulate

Most common population density level:

	Neighborhood	1st Most Common pop level	2nd Most Common pop level
17	ARCOS DE CENTENARIO	medium_pop_density	high_pop_density
24	AVE REAL (AMPL)	low_pop_density	high_pop_density
40	CANUTILLO 3RA SECCION	low_pop_density	high_pop_density
41	CANUTILLO PREDIO LA PRESA	low_pop_density	high_pop_density
42	CANUTILLO(AGUASCALIENTES)	low_pop_density	high_pop_density
47	COLINAS DEL SUR	low_pop_density	high_pop_density

Most common socioeconomic level:

Neighborhood	1st Most Common sociecon level	2nd Most Common sociecon level	3rd Most Common sociecon level
ARCOS DE CENTENARIO	high_socio_econ_lvl	low_socioecon_level	medium_socio_econ_lvl
AVE REAL (AMPL)	medium_socio_econ_lvl	low_socioecon_level	high_socio_econ_lvl
CANUTILLO 3RA SECCION	high_socio_econ_lvl	low_socioecon_level	medium_socio_econ_lvl
CANUTILLO PREDIO LA PRESA	high_socio_econ_lvl	low_socioecon_level	medium_socio_econ_lvl
CANUTILLO(AGUASCALIENTES)	medium_socio_econ_lvl	low_socioecon_level	high_socio_econ_lvl
COLINAS DEL SUR	high_socio_econ_lvl	low_socioecon_level	medium_socio_econ_lvl
EL PIRU (FRACC)	medium_socio_econ_lvl	low_socioecon_level	high_socio_econ_lvl
GOLONDRINAS	medium_socio_econ_lvl	low_socioecon_level	high_socio_econ_lvl
HERON PROAL	medium_socio_econ_lvl	low_socioecon_level	high_socio_econ_lvl
LA PERA XOCHINAHUAC (U HAB)	high_socio_econ_lvl	low_socioecon_level	medium_socio_econ_lvl

I repeated this tables and obtained the following results:

Cluster 1 (Purple dots) :

I called this cluster, the "Recreation Cluster" since it comprises several venue types related to different activities, you can find restaurants and coffee shops but also movie theaters, swimming pools, golf courses and zoos appear in the most common places of the neighborhoods. This cluster also is characterized by low population density and high socioeconomic level.

Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
AGUILAS PILARES	Mexican Restaurant	Bakery	Grocery Store	Dance Studio	Shopping Mall
ALCANTARILLA	Italian Restaurant	Ice Cream Shop	Restaurant	Bar	Recording Studio
ALPES	Coffee Shop	Mexican Restaurant	Taco Place	Sporting Goods Shop	Seafood Restaurant
ATLAMAYA	Music Venue	Bistro	Taco Place	Chinese Restaurant	Argentinian Restaurant

Cluster 2 (Blue dots) :

I called this cluster the "Practical Cluster", it seems that taco place's and convenience stores are very present here. This could be related to people who don't have much time to have a meal or tho go to the supermarket or grocery store. Also, medium socioeconomic level is predominant as well as low population density.

Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
19 DE MAYO	Restaurant	Taco Place	Mexican Restaurant	Garden	Donut Shop
26 DE JULIO	Coffee Shop	Convenience Store	Restaurant	Mexican Restaurant	Donut Shop
ACUEDUCTO	Taco Place	Coffee Shop	Food Truck	Bakery	Pharmacy
ACUILOTLA	Convenience Store	BBQ Joint	Playground	Basketball Court	Liquor Store
AGUILAS 3ER PARQUE	Taco Place	Convenience Store	Pizza Place	Bakery	Restaurant

Cluster 3 (Aqua dots) :

By all means this is the "Taco Cluster", the name is self explanatory and we could also add that theres other common venues present but without a doubt, tacos dominate the scene. It's also characterized by low population density and medium income level. You could say this is the average taco lover Joe cluster.

Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
ARTURO GAMIZ	Taco Place	Restaurant	College Gym	Trail	Mexican Restaurant
ARTURO MARTINEZ	Taco Place	Seafood Restaurant	Residential Building (Apartment / Condo)	Burger Joint	Market
BALCONES DE CEHUAYO	Taco Place	Park	Food Truck	Coffee Shop	Basketball Court
BELEN DE LAS FLORES (U HAB)	Taco Place	Bookstore	Gym	Park	Fast Food Restaurant

Cluster 4 (Light green dots) :

I called this cluster the "fitness cluster" since it's most common venue is a gym together with some soccer fields. Predominantly you find low population density and medium income.

Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
2DA JALALPA TEPITO (AMPL)	Fast Food Restaurant	Taco Place	Soccer Field	Gym	Zoo Exhibit
GARCIMARRERO	Gym	Burger Joint	Soccer Field	Electronics Store	Flower Shop
JALALPA (AMPL)	Gym	Bakery	Fast Food Restaurant	Mexican Restaurant	Ice Cream Shop
LOMAS DE CAPULIN	Gym	Zoo Exhibit	Food	Fondue Restaurant	Flower Shop
LOS CEDROS	Supermarket	Soccer Field	Gym	Zoo Exhibit	Electronics Store

Cluster 5 (Orange Dots) :

This is also a mexican cluster since it's most common venues are taco places too, but also other mexican food restaurants. You can also observe that medium socioeconomic level and low population density are present. We could say that compared to the "Taco Cluster" this one has more variety regarding to mexican food.

Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
1RA VICTORIA	Taco Place	Mexican Restaurant	Park	Pharmacy	Bar
1RA VICTORIA SECCION BOSQUES	Taco Place	Park	Mexican Restaurant	Food	Gym Pool
2DA EL PIRUL (AMPL)	Mexican Restaurant	Taco Place	Burger Joint	Restaurant	Food Truck
ABRAHAM GONZALEZ	Taco Place	Mexican Restaurant	Restaurant	Bar	Flea Market
ALFALFAR	Mexican Restaurant	Taco Place	Burger Joint	Farmers Market	Café
ALFONSO XIII	Taco Place	Seafood Restaurant	Mexican Restaurant	Farmers Market	Residential Building (Apartment / Condo)

5.0 Discussion

Even though the results seem pretty robust in some clusters, there's still significant variation inside others. To make decisions as a CPG business which will be translated into an upfront investment, this analysis might not be enough. Taking this into account, what we can do as next steps is to go deeper inside every cluster. The greatest benefit of this study is that it narrows our scope into which we have to look in order to make an informed decision.

Additionally, if more data is included in the study for example, for all the districts in Mexico City, it's possible that the results of the model will improve.

6.0 Conclusion

After running a K-means algorithm and obtaining a clustering analysis for the neighborhoods in Álvaro Obregón district of Mexico City, we found that for the objective of opening a CPG business we might need a deeper analysis but the most important result of this study is to narrow the scope of search which is also needed to start the business.

For example, if we take the "Italian cluster" we might not want to look into there if we want to open a Pizza parlour.

The results are straightforward to understand, but we recommend that further analysis is performed to reach the objective defined in the title of this investigation.

7.0 References

Agebs shapefile
https://datos.gob.mx/busca/dataset?tags=AGEBS

Neighborhoods in Mexico City
https://datos.cdmx.gob.mx/explore/dataset/coloniascdmx/information/

Githu repository
https://github.com/rogeliomj/CC_Final_Assignment