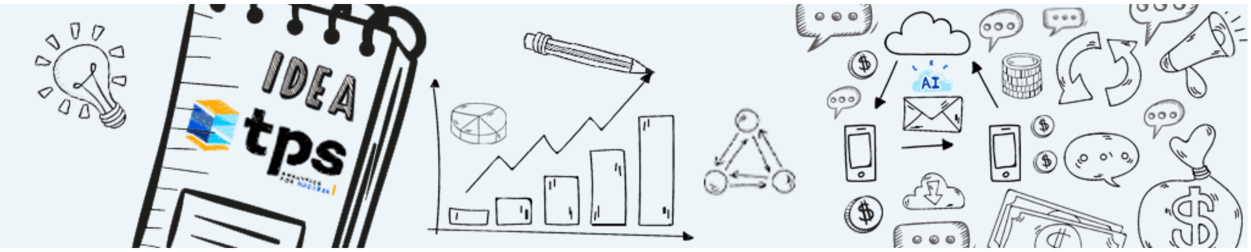


Zero Contact Branches

Incidence relationship between types of business vs COVID indexes (Number of Negative Cases, number of confirmed cases and number of deaths)

Alejandro Alvarado

November 2020

1. Introduction

1.1 Background

With the arrival of the pandemic, many businesses have required to accelerate their digital transformation processes, however, they have realized that this transformation is not only focused on digitizing work or reconverting their businesses, but also analyzing in depth the opportunities they have. and make an optimal management of investment resources.

For this, and especially the financial sector, has put to work seeking to better understand its environment, adapt / improve its digital processes, as well as approach vulnerable areas derived from the pandemic with the possibility of opening a new type of branch called Zero Contact.

Zero Contact consist on a Virtual Branch where common people can ask for a loan or cash from a branch sponsor or branch executive, this sponsor will receive money from our bank, this will happen using encrypted terminal that helps on onboarding process and manage their own information.

1.2 Problem Statement

It has started with the exploration of a new branch expansion process based on Data Science. To do this, and due to the pandemic, you want to know how much relationship exists between the various **types of businesses** located in a geographical area vs the **Number of Negative Cases, number of confirmed cases, number of deaths** caused by COVID-19? With this, the correlation between the types of businesses and the index mentioned can be determined to determine possible suitable areas for said new branches.

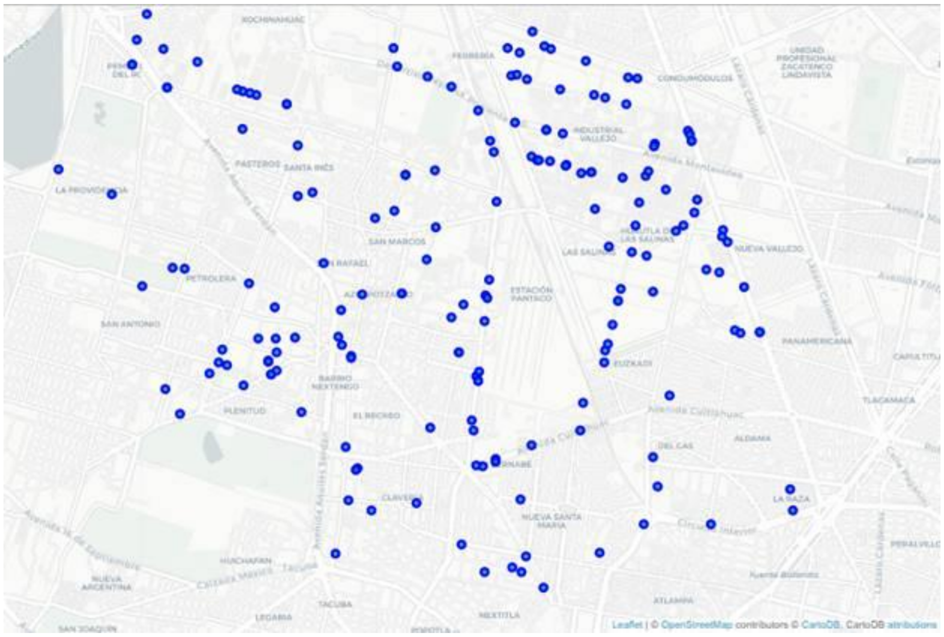
1.3 Data

To consider the problem we can list the data as below (I originally considered only 2 sources of data):

Because I live in CDMX (formerly Mexico City) in the town hall of Azcapotzalco it comes from Nahuatl ***azcaputzalli*** --> anthill, and ***co***--> locative: meaning "en el hormiguero" *(spanish)* --> "in the anthill" *(english)* it's a pre-hispanic town.

1. **Data from Foursquare**

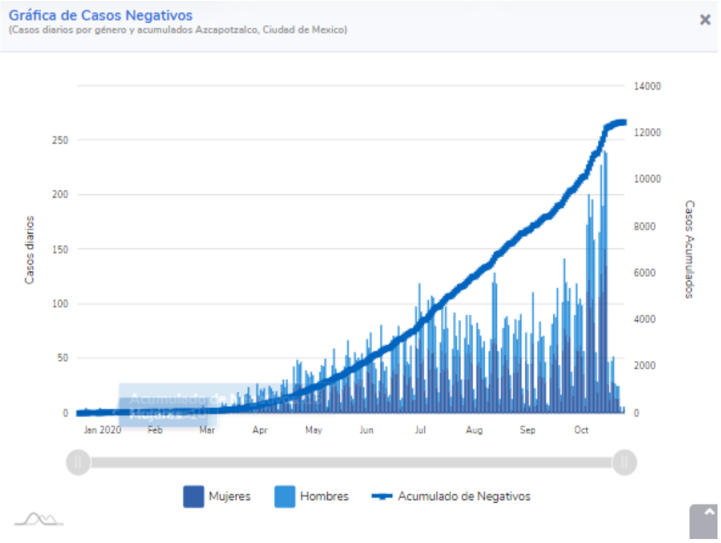
(https://foursquare.com/developers/apps) what type of venues exist in the surroundings of a geographic area [Azcapotzalco].



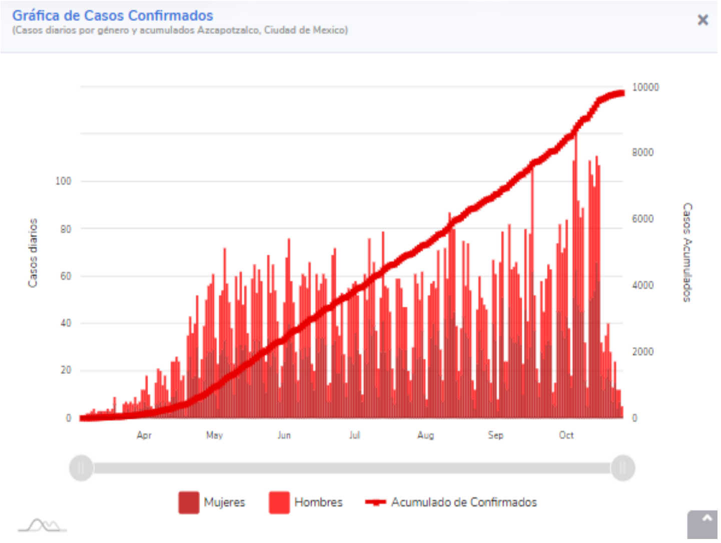
2. **Data from GDE (Dirección General de Epidemiología) in México**

Covid19 open data (Datos COVID-19 , DGE Open Data Mexico), general public information from México Government. Filter by [Azcapotzalco] on where the growth rates have been in three main metrics.

2.1. Number of negative cases



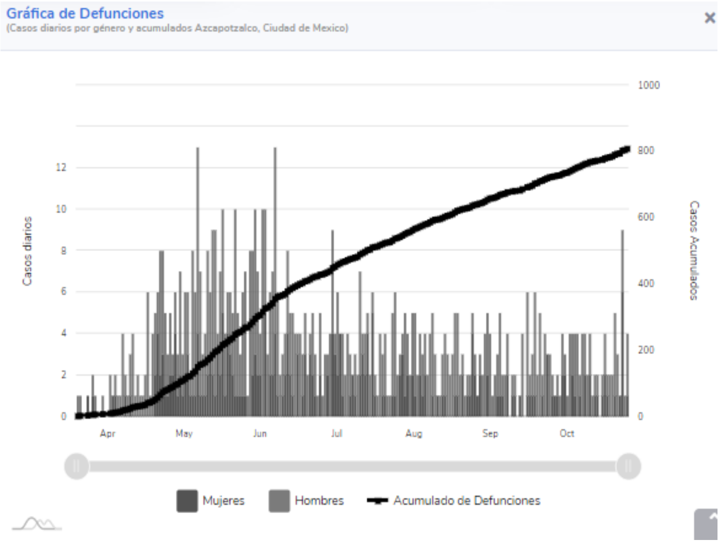
2.2. Number of confirmed cases



Gráfica de Casos Confirmados

(Casos diarios por género y acumulados Azcapotzalco, Ciudad de Mexico)

2.3. Number of deaths



Gráfica de Defunciones

(Casos diarios por género y acumulados Azcapotzalco, Ciudad de Mexico)

Here I found a little problem !!!

This DGE info, have no detailed data for **suburb**; by this moment only **mayor** is available, also the data of **Four Square** is on **suburb** level.

So, it's not same granularity, then I have to dismiss this data. *Yeah, it’s my foult*

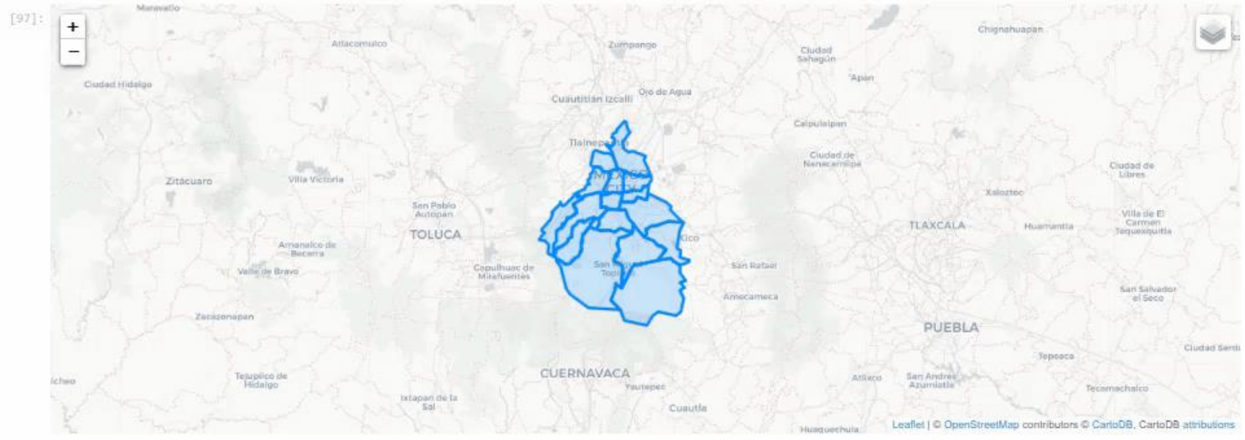
3. **Data** Cases associated with COVID in CDMX Mexico City data portal

GOBIERNO DE LA CIUDAD DE MÉXICO Datos Abiertos Ciudad de México

In this site, I found valued data! like locations shapes at different levels [Mexico City, Mexico City w/mayors, Suburbs of an Azcapotzalco Mayor], also many kinds of metrics, from COVID, Security, and others.

3.1. **Mayors / Districts** limits *(Delimitación de las alcaldías)*
Mayors / Districts limits

3.2. **Suburb** limits *(Delimitación de colonias)*
Suburb [Azcapotzalco] limits



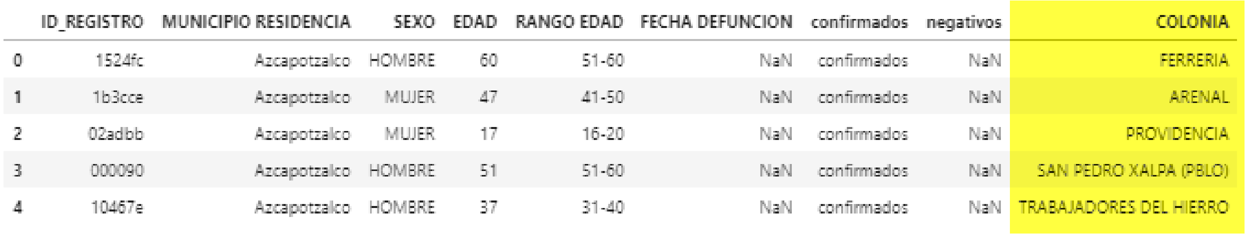
3.3. **Suburb** active cases of COVID-19 in Mexico City *(Casos activos de Covid-19 en Ciudad de México a nivel colonia)*
COVID-19 active cases

3.4. **Suburb** crime DATA in Azcapotzalco *(Crime cases in Azcapotzalco by suburb)*
Crime DATA by Azcapotzalco *This is only if I don't find anything in COVID-19 Data, please don't judge me =)*

2. Methodology

I start downloading and validating the data from CDMX, I found **111** suburbs from **Azcapotzalco**, and **31867** row cases in **COVID** data for Azcapotzalco (data updated on Dec 1st).

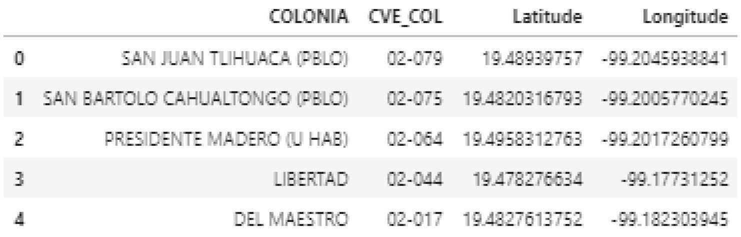
The COVID data does not have SUBURB [COLONIA] (*which is a low-level info, like Four Square*), so, I create a function to FAKE and randomize the SUBURB [COLONIA] in order to be used in a cluster classification.



	ID_REGISTRO	MUNICIPIO RESIDENCIA	SEXO	EDAD	RANGO EDAD	FECHA DEFUNCION	confirmados	negativos	COLONIA
0	1524fc	Azcapotzalco	HOMBRE	60	51-60	NaN	confirmados	NaN	FERRERIA
1	1b3cce	Azcapotzalco	MUJER	47	41-50	NaN	confirmados	NaN	ARENAL
2	02adbb	Azcapotzalco	MUJER	17	16-20	NaN	confirmados	NaN	PROVIDENCIA
3	000090	Azcapotzalco	HOMBRE	51	51-60	NaN	confirmados	NaN	SAN PEDRO XALPA (PBLO)
4	10467e	Azcapotzalco	HOMBRE	37	31-40	NaN	confirmados	NaN	TRABAJADORES DEL HIERRO

Sample of kind of data that we can use after the fix.

Here is the information used to locate Suburb [COLONIA] in Azcapotzalco.

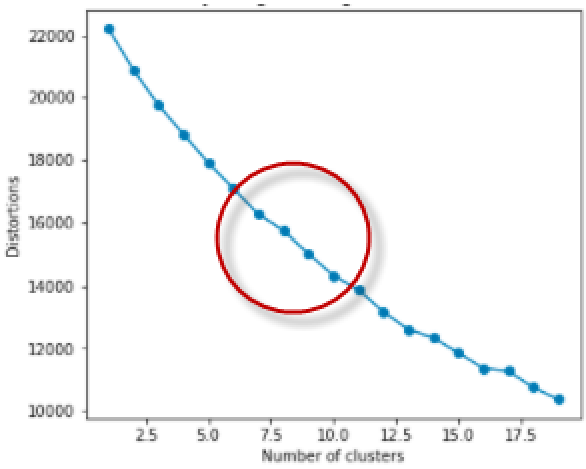


	COLONIA	CVE_COL	Latitude	Longitude
0	SAN JUAN TLIHUACA (PBLO)	02-079	19.48939757	-99.2045938841
1	SAN BARTOLO CAHUALTONGO (PBLO)	02-075	19.4820316793	-99.2005770245
2	PRESIDENTE MADERO (U HAB)	02-064	19.4958312763	-99.2017260799
3	LIBERTAD	02-044	19.478276634	-99.17731252
4	DEL MAESTRO	02-017	19.4827613752	-99.182303945

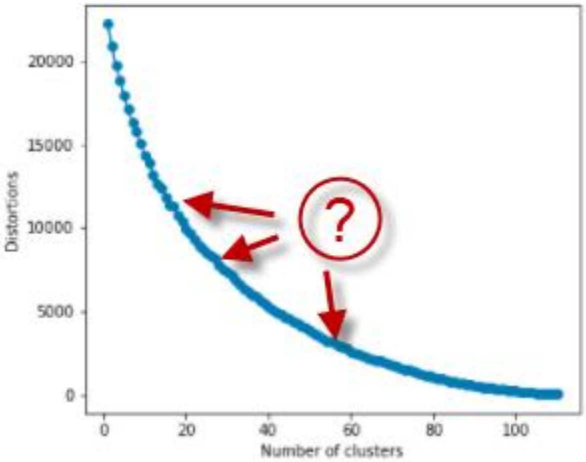
There are many models for **clustering** out there. In this project I used the model that is considered one of the simplest models amongst them. Despite its simplicity, the **K-means** is vastly used for clustering in many data science applications, especially useful if you need to quickly discover insights from **unlabeled data**. I use k-Means for suburbs segmentation.

Calculate distortion for a range of number of cluster in a density approach.

K=20



K=111 (Total of Suburbs [COLONIAS])

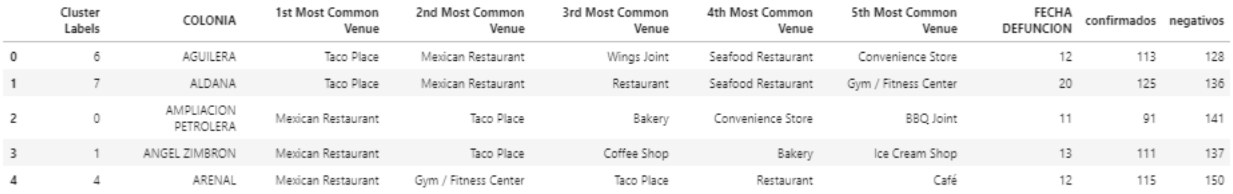


I ran K-Means to cluster the suburbs [COLONIAS] into 20 clusters because when I analyze the K-Means with elbow method it ensured me the 20 degree for optimum k of the K-Means. Results are not very clear (*for me*) so I ran into 111 clusters to see difference (*all data*), quiet or minor different! between them, **but** I see different points where to check!.

Was hard to decide the number of clusters … so, I let the ML to work!

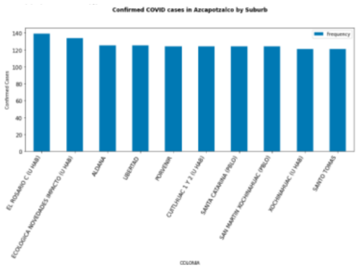
I have some common venue categories in suburbs [COLONIAS]. This reason I used unsupervised learning **K-means** algorithm to cluster the suburbs [COLONIAS]. **K-Means** algorithm is one of the most common cluster methods of unsupervised learning.

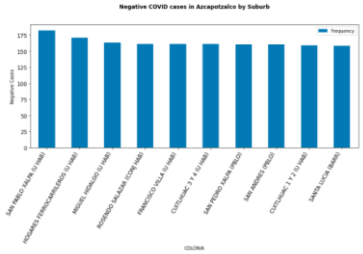
Here is my merged table with cluster labels for each suburb [COLONIA] and COVID data.

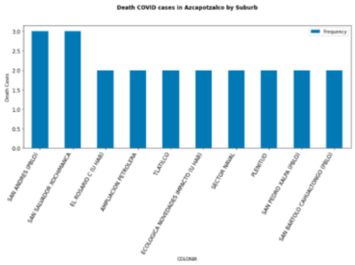


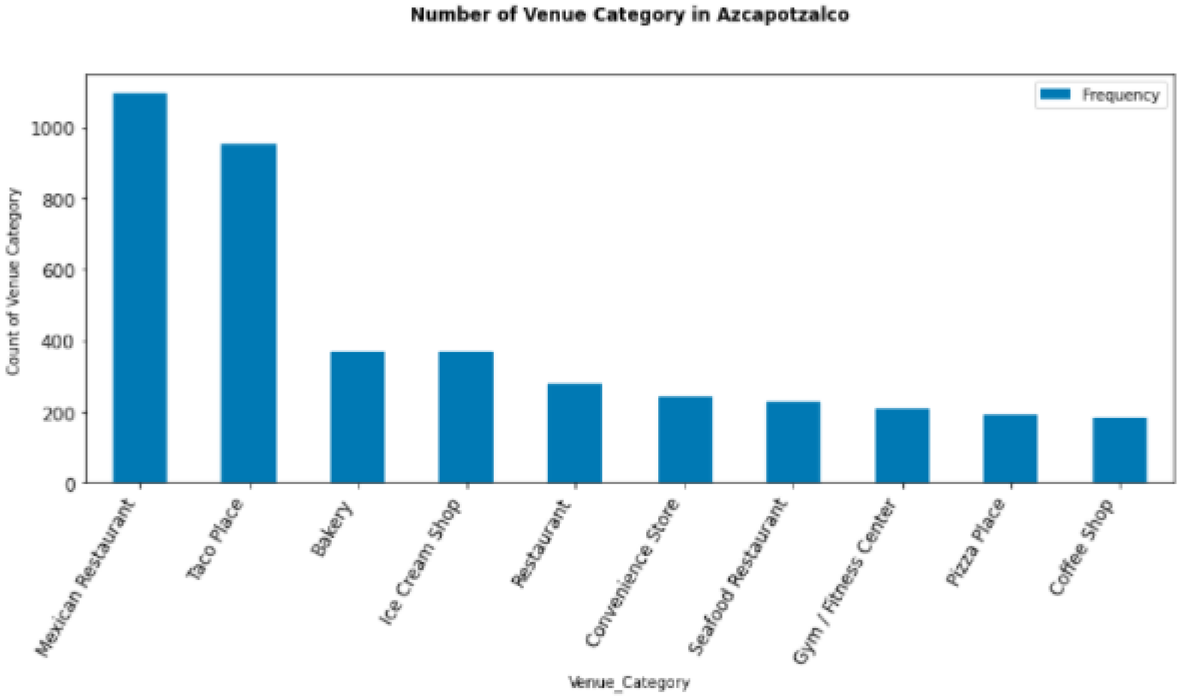
	Cluster Labels	COLONIA	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	FECHA DEFUNCION	confirmados	negativos
0	6	AGUILERA	Taco Place	Mexican Restaurant	Wings Joint	Seafood Restaurant	Convenience Store	12	113	128
1	7	ALDANA	Taco Place	Mexican Restaurant	Restaurant	Seafood Restaurant	Gym / Fitness Center	20	125	136
2	0	AMPLIACION PETROLERA	Mexican Restaurant	Taco Place	Bakery	Convenience Store	BBQ Joint	11	91	141
3	1	ANGEL ZIMBRON	Mexican Restaurant	Taco Place	Coffee Shop	Bakery	Ice Cream Shop	13	111	137
4	4	ARENAL	Mexican Restaurant	Gym / Fitness Center	Taco Place	Restaurant	Café	12	115	150

I examine what is the frequency of average COVID data in different ranges. Bars can help to visualization (so, lets check top[10] for each metric):



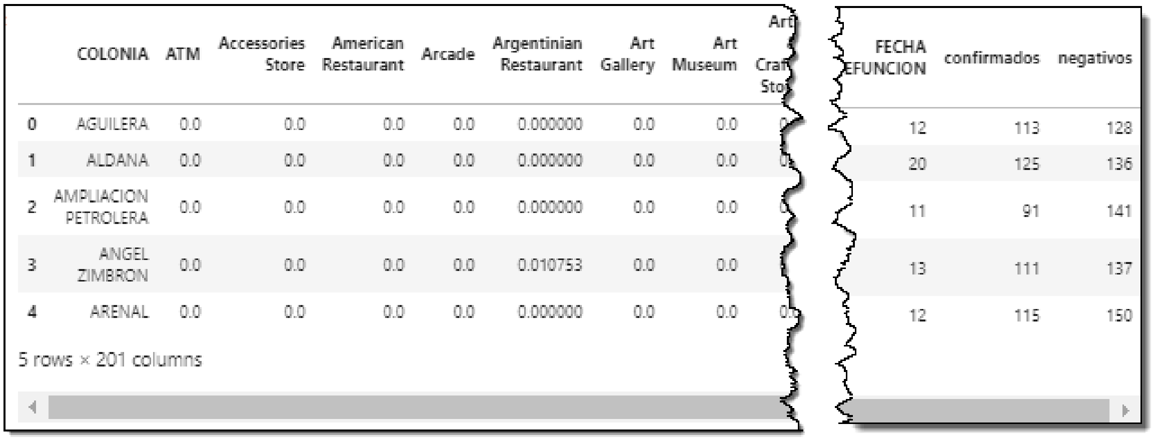






3. Results

I merge all variables with related cluster information’s in one main master table.

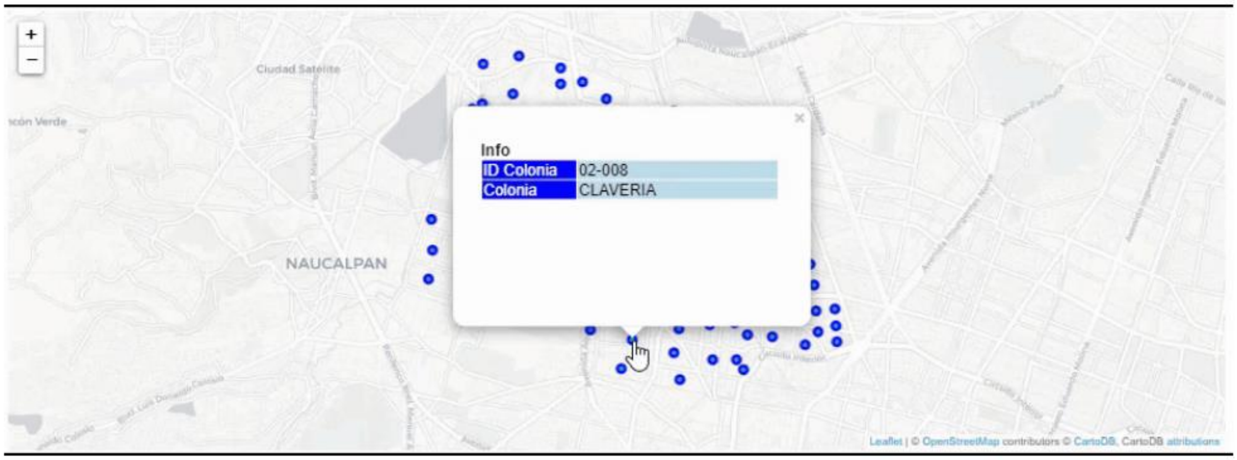


	COLONIA	ATM	Accessories Store	American Restaurant	Arcade	Argentinian Restaurant	Art Gallery	Art Museum	Art Craf Sto	FECHA EFUNCION	confirmados	negativos
0	AGUILERA	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0	12	113	128
1	ALDANA	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.	20	125	136
2	AMPLIACION PETROLERA	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0	11	91	141
3	ANGEL ZIMBRON	0.0	0.0	0.0	0.0	0.010753	0.0	0.0		13	111	137
4	ARENAL	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.	12	115	150

5 rows × 201 columns

You can also see a clustered map suburbs [COLONIAS] of Azcapotzalco in the below.

As it seems in above maps, we can check the **clusters** as below (*including new data*):



Info

ID Colonia	02-008
Colonia	CLAVERIA

+
−
Ciudad Satelite
ncón Verde
NAUCALPAN
Leaflet | © OpenStreetMap contributors © CartoDB, CartoDB attributions

In summary section, one of my aim was also visualize the COVID info for any suburb [COLONIA] with color cluster map.

4. Discussion

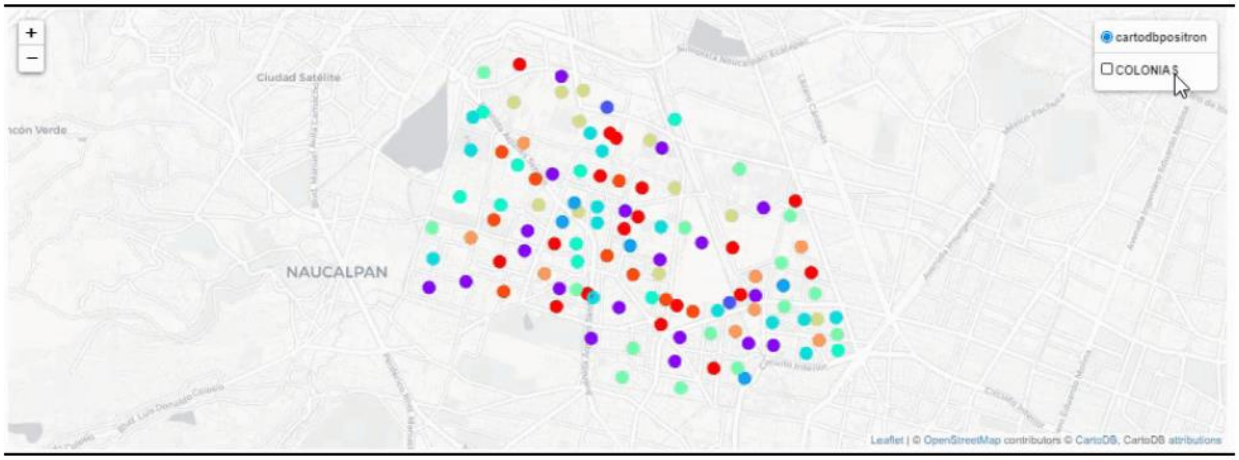
Azcapotzalco in one of the mayors of CDMX (formerly Distrito Federal). The total number of measurements and population densities of the 111 districts in total and can vary. As there is such a complexity, very different approaches can be tried in clustering and classification studies. Moreover, it is hard to not every classification method can yield the same high-quality results for this result.

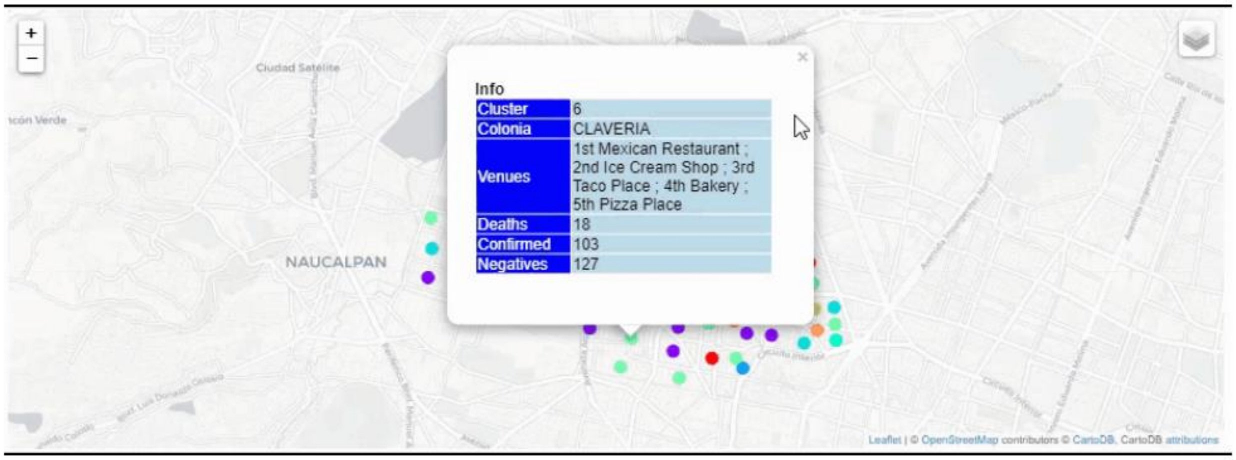
I used the **K-means** algorithm as part of this clustering study. When I tested the **Elbow** method, I set the optimum k value to 9. However, I already used 111 suburbs [COLONIA] coordinates. For more detailed and accurate guidance, the data set can be check and the details of the suburbs can also be drilled.

I also performed data analysis through this information by adding the coordinates of suburbs. In future studies, these data can also be accessed dynamically from specific platforms or packages.

5. Conclusion

I ended the study by visualizing the data and clustering information on the Azcapotzalco map. In future studies, web or telephone applications can be carried out to direct people.


+
−
Ciudad Satélite
NAUCALPAN
cartodbpositron
COLONIAS
Leaflet | © OpenStreetMap contributors © CartoDB, CartoDB attributions


+
−
Ciudad Satélite
NAUCALPAN
Leaflet | © OpenStreetMap contributors © CartoDB, CartoDB attributions

Info

Cluster	6
Colonia	CLAVERIA
Venues	1st Mexican Restaurant ; 2nd Ice Cream Shop ; 3rd Taco Place ; 4th Bakery ; 5th Pizza Place
Deaths	18
Confirmed	103
Negatives	127

As a result, people we can see that venues with high movement makes the difference, so we can evaluate all metrics surrounded of any suburb [COLONIA], as Restaurants or any others, COVID data impact to all of them, people can achieve better outcomes through their access to digital platforms where they can get access to many different services.

For this new point of view about New kind of Branch!

Not only for investors but also city mayors, they can manage the data or their suburb [COLONIA] more regularly by using similar data analysis types or platforms.

Alex Alvarado