



Best Neighborhoods in Mexico City

Mejores Colonia en la Ciudad de México

A Coursera Capstone Final Project

August 15th, 2019

1. Summary / Resumen Ejecutivo

By analyze the neighborhoods in the Town Halls (boroughs) of Cuauhtémoc, Miguel Hidalgo and Benito Juarez in Mexico City, seeking to answer the question of which neighborhood or neighborhoods are the best to live, based in the nearest of particular interest sites, we found a group of neighborhoods which complies with those interests:

- 1) Supermarkets, 2) parks, 3) yoga centers, 4) coffee shops, 5) tacos, 6) pizza places, 7) music venues, 8) libraries, 9) spas and 10) pharmacies.

These group of neighborhoods are those on the Cluster 5, determined by applied the cluster methodology as the better neighborhoods to live with a score of 5 respect of the parameters defined in the methodology.

As a result, we can say that the neighborhood Roma Sur II on the Town Hall Cuauhtémoc, the neighborhood Angel Zimbron on the Town Hall Miguel Hidalgo, and the neighborhoods Narvarte VI and Residencial Emperadores on the Town Hall Benito Juarez are the best neighborhoods to live in Mexico City. Those neighborhoods are visualized in the Final Map.

Al analizar las colonias dentro de las Alcaldías Cuauhtémoc, Miguel Hidalgo y Benito Juárez en la Ciudad de México, buscando contestar la pregunta de cuál o cuáles son las mejores colonias para vivir, basado en la cercanía de sitios basado en intereses particulares, encontramos un grupo de colonias que cumplen con dichos intereses:

- 1) Supermercados, 2) parques, 3) yoga, 4) cafés, 5) tacos, 6) pizzerías, 7) lugares de música, 8) librerías, 9) spas y 10) farmacias.

Este conjunto de colonias son las correspondientes al Grupo 5 que se determinaron al aplicar la metodología de agrupación como las mejores colonias para vivir con un puntaje de 5 respecto a los parámetros establecidos en la metodología.

Como resultado podemos decir que la colonia Roma Sur II de la Alcaldía Cuauhtémoc, la colonia Ángel Zimbrón en la alcaldía Miguel Hidalgo y las colonias Narvarte VI y Residencial Emperadores de la alcaldía Benito Juárez son las mejores colonias para vivir en la Ciudad de México. Estas colonias se pueden visualizar en el Mapa Final.



Neighborhoods in Mexico City Map / Mapa de Colonias en Ciudad de México

2. Content / Contenido

1. Summary / Resumen Ejecutivo	2
2. Content / Contenido	3
3. Introduction / Introducción	3
4. Methodology / Metodología	5
5. Data Retrieval / Recuperación de Datos	7
6. Data Visualization / Visualización de Datos	9
7. Results / Resultados	14
8. Conclusion / Conclusión	18
Annex – K-Mean Algorithm / Anexo – Algoritmo K-Mean	19
References / Referencias	21
Appendix – Mexico City / Apéndice – Ciudad de México	21

3. Introduction / Introducción

Objective and Purpose

Objective: To qualify the neighborhoods in Mexico City based on the availability of services or venues

Purpose: The identification of the best neighborhoods to live in Mexico City.

Background

Mexico City is one of the largest metropolitan zones in the world. In 1900 the city size was less than 10% of what's today and the growth in the following four decades ended by swallowing small towns that became in important touristy centers (like «Coyoacan» and «Tlalpan») or business spots (like «La Condesa» o «Santa Fe»).

Between 1950 and 1970 many people begin to move to new neighborhoods in the city edges, reaching a point that exceeds the «Federal District» limits and into some of the «State of Mexico» municipalities, creating transport problems because of the transfer from residential zones to workplaces zones.

The city continues to grow as the same as services providers seek for the bests spots for their venues, creating different zones of concentration, as well as more residential zones were created each time more far away from these zones, magnifying the transfer times up to 2-3 hours, in the morning as well as at the afternoon.

Objetivo y Propósito

Objetivo: Calificar las colonias de la Ciudad de México en base a la disponibilidad de servicios establecidos.

Propósito: La identificación de las mejores colonias para vivir en la Ciudad de México.

Antecedentes

La Ciudad de México es una de las zonas metropolitanas más grandes del mundo. En 1900 el tamaño de la ciudad era menor al 10% de su tamaño actual y el crecimiento en las siguientes cuatro décadas terminó por abarcar pequeños pueblos que se han convertido en importantes centros turísticos (como Coyoacán o Tlalpan) o centros de negocios (como La Condesa o Santa Fe).

Entre 1950 y 1970 mucha gente comenzó a mudarse a nuevas colonias en las periferias de la ciudad, al grado de sobrepassar los límites del Distrito Federal abarcando algunos municipios del Estado de México y creando problemas de transporte por el traslado desde las zonas residenciales hasta las zonas de trabajo.

La ciudad siguió creciendo al tiempo que los proveedores de servicio buscaban los mejores puntos para sus locales, creando diferentes zonas de concentración al tiempo que se creaban zonas de residencia cada vez más alejados de estas zonas, magnificando los tiempos de traslado hasta 2-3 horas, tanto en las mañanas como en las tardes.

Problem

Mexico City, being a city with a lot of attractiveness for living in, generates the typical question of: Where is the best place or spot to live in Mexico City? To answer that question we must first establish how to qualify or grade the neighborhoods based on their services, and given those neighborhoods value based on the requirements from a person interested in identify the best option to live.

The parameters are not the same for everyone, because the interests can be variable depending on the living conditions that are required, such as socioeconomic level, age, marital status, etc. For example, an elder person can be looking for a neighborhood that has parks, theaters, and coffee shops, other person with children might be looking for a neighborhood that has schools, movie theaters, and malls, and another younger person might be looking for neighborhoods with venues with live music, bars and gyms. The combinations of interests for parameters can be as many as people living in the city.

The project aims to achieve the objective for a combination of specific interests (taking as example from a particular person) and can be adapted for any other people by filling their particular interests.

Stakeholders

The stakeholders of this project can be people from the outside that want to establish in Mexico City as well as people that want to move to other neighborhood for different reasons, especially for the time expend in the transfer from their homes to their work places.

It also can be used to those businesses that can identify the surplus value of their services in certain neighborhoods that lack them.

Problema

La Ciudad de México, al ser una urbe con gran atractivo para establecerse, genera la pregunta típica de: ¿Dónde es el mejor lugar para vivir en la Ciudad de México? Para responder la pregunta debemos establecer primero cómo calificar a las colonias en base a sus servicios y darle valor a estos en base a los requerimientos de las personas interesadas en identificar la mejores opción para vivir.

Los parámetros no son iguales para todos, ya que los intereses pueden ser muy variados dependiendo de las condiciones de vida que requieren, nivel socioeconómico, edad, etc. Por ejemplo, una persona de edad mayor puede estar buscando una colonia que contenga parques, teatros y cafeterías, otra persona con hijos puede estar buscando colonias con escuelas, cines y centro comerciales, y otra persona más joven puede estar buscando colonias que tengan establecimientos con música, bares y gimnasios. La combinación de intereses puede ser tan variada como personas que viven en la ciudad.

El proyecto busca lograr el objetivo planteado para una combinación de intereses específicos (tomando el ejemplo de una persona) y que pueda adaptarse a cualquier otra integrando sus intereses particulares.

Interesados

Los interesados en este proyecto pueden ser tanto personas que quieren establecerse en la Ciudad de México como aquellas que buscan mudarse a otras colonias por diferentes motivos, especialmente por el tiempo de traslado de sus respectivas residencias hasta sus zonas de trabajo.

También puede servir para aquellos comercios que pueden identificar la plusvalía de sus servicios en determinadas colonias que carecen de estos.

4. Methodology / Metodología

Information Requirement

For the project, it's required Mexico City neighborhoods information, their geospatial data, and information of places, venues and/or services and their qualifications.

Also is required, for the analysis, to define the interests of a particular person as an example to assure that the project reaches the objective and purpose mentioned earlier in this document.

Data Source

«Datos Abiertos Ciudad de México – ‘Mexico City.csv’»: A database from the government of CDMX will be used and the site has the list in CSV format for the neighborhoods and their geospatial data.

The image below shows an example of the database:

Requerimientos de Información

Para el proyecto se requiere información de las colonias de la Ciudad México, sus datos geográficos, información de tipos de lugares, establecimientos y/o servicios y la calificación de estos.

De igual manera se requiere, para el análisis, establecer los intereses de una persona en particular a manera de ejemplo para asegurar que el proyecto cumple con el objetivo y propósito antes planteado en este documento.

Fuentes de Información

Datos Abiertos Ciudad de México – ‘Mexico City.csv’: Se utilizará la base de datos del gobierno de la CDMX que cuenta con un listado en formato CSV de las colonias y sus datos geográficos.

La siguiente imagen muestra un ejemplo de esta base de datos por colonia:

ENTIDAD	COLONIA	Geo Point	Geo Shape	ALCALDIA	
37	9	7 DE NOVIEMBRE	19.46048775211773, -99.12046074465486	{"type": "Polygon", "coordinates": [[[-99.1...	GUSTAVO A. MADERO
38	9	TRIUNFO DE LA REPUBLICA	19.491661026741767, -99.1097111580985	{"type": "Polygon", "coordinates": [[[-99.1...	GUSTAVO A. MADERO
39	9	INDUSTRIAL I	19.473773253527792, -99.1240192960837	{"type": "Polygon", "coordinates": [[[-99.1...	GUSTAVO A. MADERO
40	9	CASAS ALEMAN (AMPL) II	19.47880386453557, -99.08737442769691	{"type": "Polygon", "coordinates": [[[-99.0...	GUSTAVO A. MADERO
41	9	TEPETATL	19.562424437601376, -99.1380859685765	{"type": "Polygon", "coordinates": [[[-99.1...	GUSTAVO A. MADERO
42	9	VISTA HERMOSA	19.548099967962443, -99.1224855548496	{"type": "Polygon", "coordinates": [[[-99.1...	GUSTAVO A. MADERO
43	9	VILLA DE ARAGON (FRACC)	19.468435220290466, -99.0605841043596	{"type": "Polygon", "coordinates": [[[-99.0...	GUSTAVO A. MADERO
44	9	VASCO DE QUIROGA	19.494654258573487, -99.0939298050141	{"type": "Polygon", "coordinates": [[[-99.0...	GUSTAVO A. MADERO
45	9	SCT (U HAB)	19.50769610693348, -99.1567486453785	{"type": "Polygon", "coordinates": [[[-99.1...	GUSTAVO A. MADERO
46	9	TORRES DE SAN JUAN (U HAB)	19.47882791228001, -99.0930244644736	{"type": "Polygon", "coordinates": [[[-99.0...	GUSTAVO A. MADERO
47	9	NUEVA VALLEJO	19.49102264556759, -99.14986258207156	{"type": "Polygon", "coordinates": [[[-99.1...	GUSTAVO A. MADERO
48	9	NARCISO BASSOLS (U HAB)	19.45614878143942, -99.05688712605318	{"type": "Polygon", "coordinates": [[[-99.0...	GUSTAVO A. MADERO
49	9	SAN ANTON DE ARAGON 3A SECCION (II)	19.45129182286436, -99.07388914486776	{"type": "Polygon", "coordinates": [[[-99.0...	GUSTAVO A. MADERO

Data base from the source / Base de datos de la fuente

«Foursquare – ‘Places API’»: We will use Foursquare’s API Places to acquire data related to ‘venues’ (as defined by Foursquare) as the ‘interests’ of the particular persona of example. It is important to note that Foursquare defines ‘venues’ as a place that one can go to, check-in, and that ‘venue’ is not necessarily a business, but can be any place such as a park or bus station. Each Foursquare ‘venue’ is assigned a ‘category’ and each ‘category’ is associated with a particular ‘categoryId’.

The image below shows the ‘categoryId’ values provided by Foursquare for music related to venues as an example:

«Foursquare – ‘Places API’»: Se usará Foursquare API Places para adquirir la información relacionada con ‘lugares’ (definidos como ‘venues’) como los intereses de la persona en particular del ejemplo. Es importante hacer notar que Foursquare define como ‘lugar’ sitios donde uno puede ir o registrarse y que ese lugar no es necesariamente un comercio, sino cualquier tipo de lugar como parques o paradas de autobús. Cada ‘lugar’ tiene asignado una ‘categoría’ que está asociada una ‘categoryId’ particular.

La siguiente imagen muestra los valores de ‘categoryId’ de sitios relacionados con música como ejemplo:

Foursquare Music-Related Venue CategoryIDs

```

Music Venue = '4bf58dd8d48988d1e5931735'
Jazz Club = '4bf58dd8d48988d1e7931735'
Piano Bar = '4bf58dd8d48988d1e8931735'
Rock Club = '4bf58dd8d48988d1e9931735'
Concert Hall = '5032792091d4c4b30a586d5c'
Amphitheater = '56aa371be4b08b9a8d5734db'
Music Festival = '5267e4d9e4b0ec79466e48d1'
Nightlife Spot = '4d4b7105d754a06376d81259'
Music Store = '4bf58dd8d48988d1fe941735'
Record Shop = '4bf58dd8d48988d10d951735'
Recording Studio = '52f2ab2ebcbc57f1066b8b37'

```

Category ID from Foursquare / ID de Categorías de Foursquare

Other Considerations

The average cost of house or apartments rental will be desirable as it is an important interest of a person's decision to move to a determinate neighborhood, but no database or information related to that has been found.

Also in further analysis or as a complementary project, a Google maps database will be used to define the average travel times from determinate neighborhood to a particular spot (workplace), but it will not be in the scope of the present project.

Otras Consideraciones

El promedio de costo de rentas de casas o departamentos sería deseable, ya que es un interés muy importante en la decisión de una persona para moverse a una determinada colonia, pero nos encontraron bases de datos relacionadas al respecto.

Igualmente, en posteriores análisis o proyectos complementarios se usará la base de datos de Google para definir el tiempo promedio de traslado de una determinada colonia a un punto en particular, pero éste no está dentro de los alcances del proyecto.

5. Data Retrieval / Recuperación de Datos

Neighborhood Dataset

The dataset was hosted by «Agencia Digital de Innovación Pública» of the Mexico City Government and was easy to download as a CSV file and imported in Jupyter Notebook. However, we found problems with the data file when trying to read them, because the information has rows that exceed the length of a cell that can be read in python. We can download the data in Excel and JSON format, but those files also have errors at reading it.

We opted to download the Excel file, and converted it manually into a CSV from where we can import and read the information to process it.

Data Cleaning

Once we have read the database in Jupyter Notebook and put it in a data frame, we eliminate the unnecessary columns with only three left: neighborhood name (COLONIA), geospatial data (Geo Point), and the town hall name or borough (ALCALDIA).

Once with the data frame with those three columns we separate the column «Geo Point» whose string has the latitudes and longitudes separated by a comma and two new columns «Latitude» and «Longitude» were made, at the same time as the «Geo Point» column was drop.

With the new columns of latitudes and longitudes as 'strings' we need to change them as 'float' type to be able to reference them in a map. Finally we drop those rows with null data in the 'Longitude' and 'Latitude' columns to map the results.

The image below shows an example of the data frame:

Base de Colonias

La base de datos está hospedada por la «Agencia Digital de Innovación Pública» del Gobierno de la Ciudad de México y fue fácil bajar información en formato CSV para importarla en Jupyter Notebook. Sin embargo se tuvo problemas con el archivo al tratar de leerlo debido a que la información presentaba filas con demasiados datos para leerse en python. Se pudo bajar igualmente los datos en Excel y JSON, pero también presentaban errores al leerla.

Se optó por bajar la información en Excel y convertirla manualmente como CSV para desde ahí poderla importar y leer la información y procesarla.

Limpieza de Datos

Una vez leída la base de datos en la Jupyter Notebook y colocada en una tabla, se eliminaron las columnas innecesarias quedando solo tres: nombre de la colonia (COLONIA), referencias geográficas (Geo Point) y nombre de la alcaldía (ALCALDIA).

Una vez con la tabla con las tres columnas se procedió a separar la columna «Geo Point» cuya cadena contenía las latitudes y longitudes separadas por una coma y se crearon las columnas «Latitude» y «Longitude» al tiempo que se eliminaba la columna «Geo Point».

Con las columnas de latitud y longitud como 'cadena' se necesitaba cambiar a 'número fraccionario' para referenciar las colonias en un mapa. Finalmente había que eliminar aquellas filas que no contenían datos para ver los resultados en el mapa.

La siguiente imagen muestra un ejemplo de la tabla:

[72]:	COLONIA	ALCALDIA	Latitude	Longitude
0	LOMAS DE CHAPULTEPEC	MIGUEL HIDALGO	19.4228411174	-99.2157935754
1	LOMAS DE REFORMA	MIGUEL HIDALGO	19.4016815485	-99.2354719599
2	LOMAS DE REFORMA (LOMAS DE CHAPULTEPEC)	MIGUEL HIDALGO	19.4106158914	-99.2262487268
3	IRRIGACION	MIGUEL HIDALGO	19.4429549298	-99.2099357048
4	DEL BOSQUE (POLANCO)	MIGUEL HIDALGO	19.4342189235	-99.2094037513

First data frame example for neighborhoods / Ejemplo de la primera tabla para colonias

The neighborhoods data frame has a shape of 1,808 x 3 and the original project needs to explore the venues from Foursquare for each neighborhood, but the free developer account from Foursquare has the limitation of 950 API calls per day and for that we need to narrow the scope of the objective by limit the results to three or four boroughs.

To do that we chose three boroughs: CUAUHTEMOC, MIGUEL HIDALGO and BENITO JUAREZ, as those are neighbors between each other and includes the most city attractions.

La tabla de colonias tiene una forma de 1,808 x 3 y el proyecto original necesitaba explorar los lugares de Foursquare para cada colonia, pero la cuenta gratis de desarrollador de Foursquare tiene un límite de 950 llamadas API por día y por eso necesitamos adelgazar el alcance del objetivo limitando los resultados de solo tres alcaldías.

Para hacer eso se escogieron tres alcaldías: CUAUHTEMOC, MIGUEL HIDALGO y BENITO JUAREZ, ya que son vecinos entre ellos e incluyen muchas de las atracciones de la ciudad.

So we filter the data frame to obtain a 215×3 shape and we can't exceed the limits of API calls per day from Foursquare.

Así que filtramos la tabla para obtener 215×3 de forma y así no excedemos el número de llamadas API por día de Foursquare.

	COLONIA	ALCALDIA	Latitude	Longitude
0	LOMAS DE CHAPULTEPEC	MIGUEL HIDALGO	19.422841	-99.215794
1	LOMAS DE REFORMA	MIGUEL HIDALGO	19.401682	-99.235472
2	LOMAS DE REFORMA (LOMAS DE CHAPULTEPEC)	MIGUEL HIDALGO	19.410616	-99.226249
3	IRRIGACION	MIGUEL HIDALGO	19.442955	-99.209936
4	DEL BOSQUE (POLANCO)	MIGUEL HIDALGO	19.434219	-99.209404

Final data frame example for neighborhoods / Ejemplo de la tabla final para colonias

Nearby Dataset

To obtain a table with the nearby venues for the 215 neighborhoods we need to make a API call for each one using an endpoint for explore the venues limited to 250 in a 1 km radius from the neighborhood center with a loop function.

Also the free developer account from Foursquare has a limitation to acquire the qualification of the venues, so we need to refine our scope.

The function creates a new data frame with more than 20,000 records related to a specific neighborhood name. We are aware that many of those records can be the same venue but related to a different neighborhood and with a different distance from the neighborhoods centers.

The image below shows an example of the nearby venues final data frame:

Base de Sitios Cercanos

Para obtener una tabla de sitios cercanos para las 215 colonias, necesitamos hacer una llamada API por cada una usando el punto-final de explorar sitios limitados a 250 en un radio de un 1 km desde el centro de la colonia en una función circulante.

También la cuenta gratis de desarrollador de Foursquare tiene la limitación de adquirir las calificaciones de los sitios, así que se tuvo que redefinir el alcance del proyecto.

La función crea una nueva tabla con más de 20,000 registros relacionados a los nombres específicos de las colonias. Estamos conscientes que muchos de esos registros tienen el mismo nombre de sitio pero relacionado a una colonia y distancia al centro del mismo.

La imagen siguiente muestra un ejemplo de la tabla final de sitios cercanos:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Distance	Venue Category
0	LOMAS DE CHAPULTEPEC	19.422841	-99.215794	Loma Linda	19.420156	-99.218127	386	Steakhouse
1	LOMAS DE CHAPULTEPEC	19.422841	-99.215794	Studio Gourmet	19.420263	-99.215545	288	Gastropub
2	LOMAS DE CHAPULTEPEC	19.422841	-99.215794	City Market	19.418977	-99.214751	443	Gourmet Shop
3	LOMAS DE CHAPULTEPEC	19.422841	-99.215794	Los Tulipanes	19.426797	-99.213360	509	Bakery
4	LOMAS DE CHAPULTEPEC	19.422841	-99.215794	punto i coma	19.418724	-99.213941	497	Accessories Store
5	LOMAS DE CHAPULTEPEC	19.422841	-99.215794	Siete 30 Salon Boutique	19.420477	-99.211307	539	Salon / Barbershop
6	LOMAS DE CHAPULTEPEC	19.422841	-99.215794	Silence Track	19.420381	-99.217477	325	Recording Studio

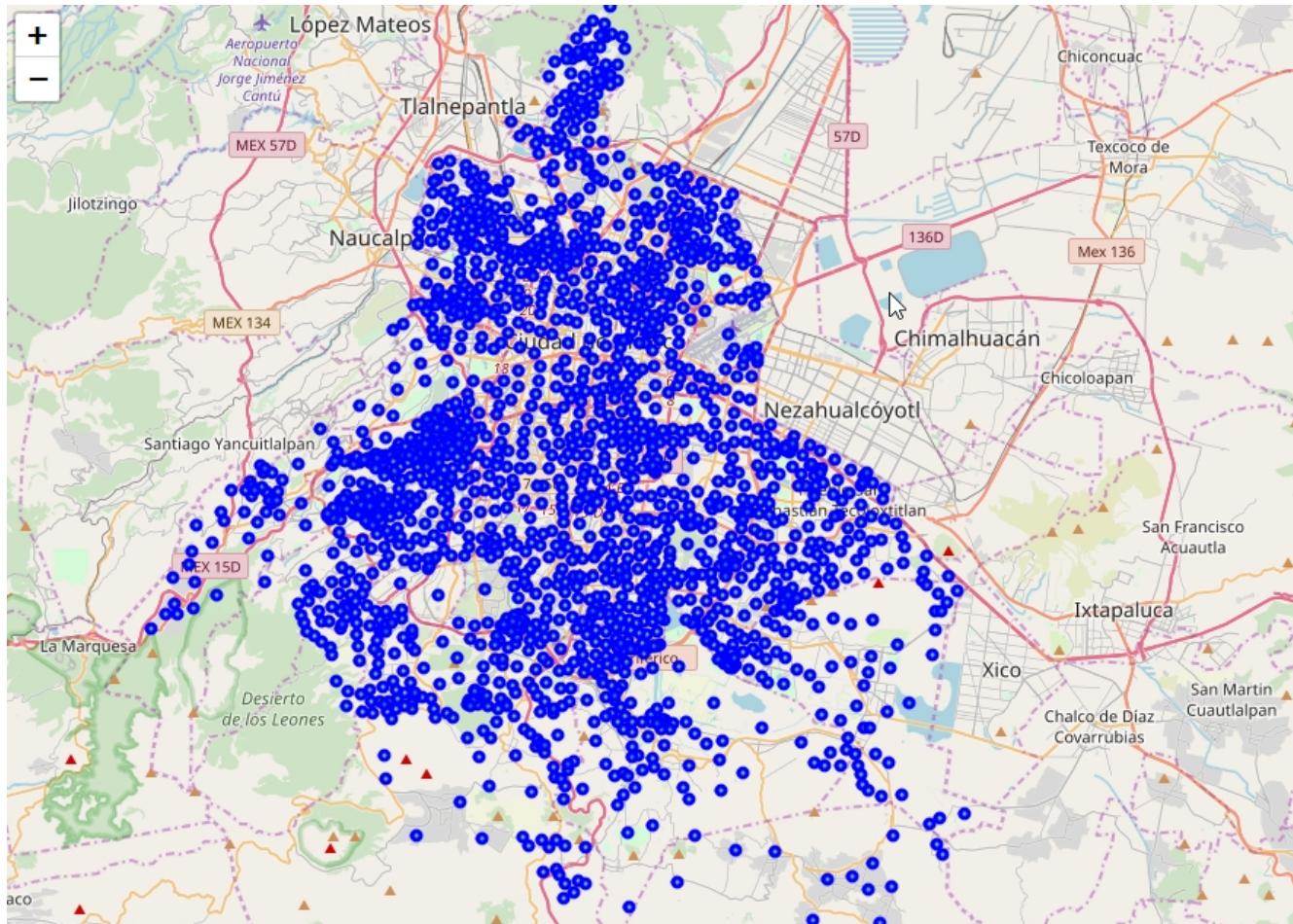
Final data frame example for nearby venues / Ejemplo de la tabla final para sitios cercanos

6. Data Visualization / Visualización de Datos

Folium Map

We use the first data frame from the 1,808 neighborhoods in Mexico City and with the help «folium» we can visualize them in a map.

The image below shows the neighborhoods geospatial centers in Mexico City:



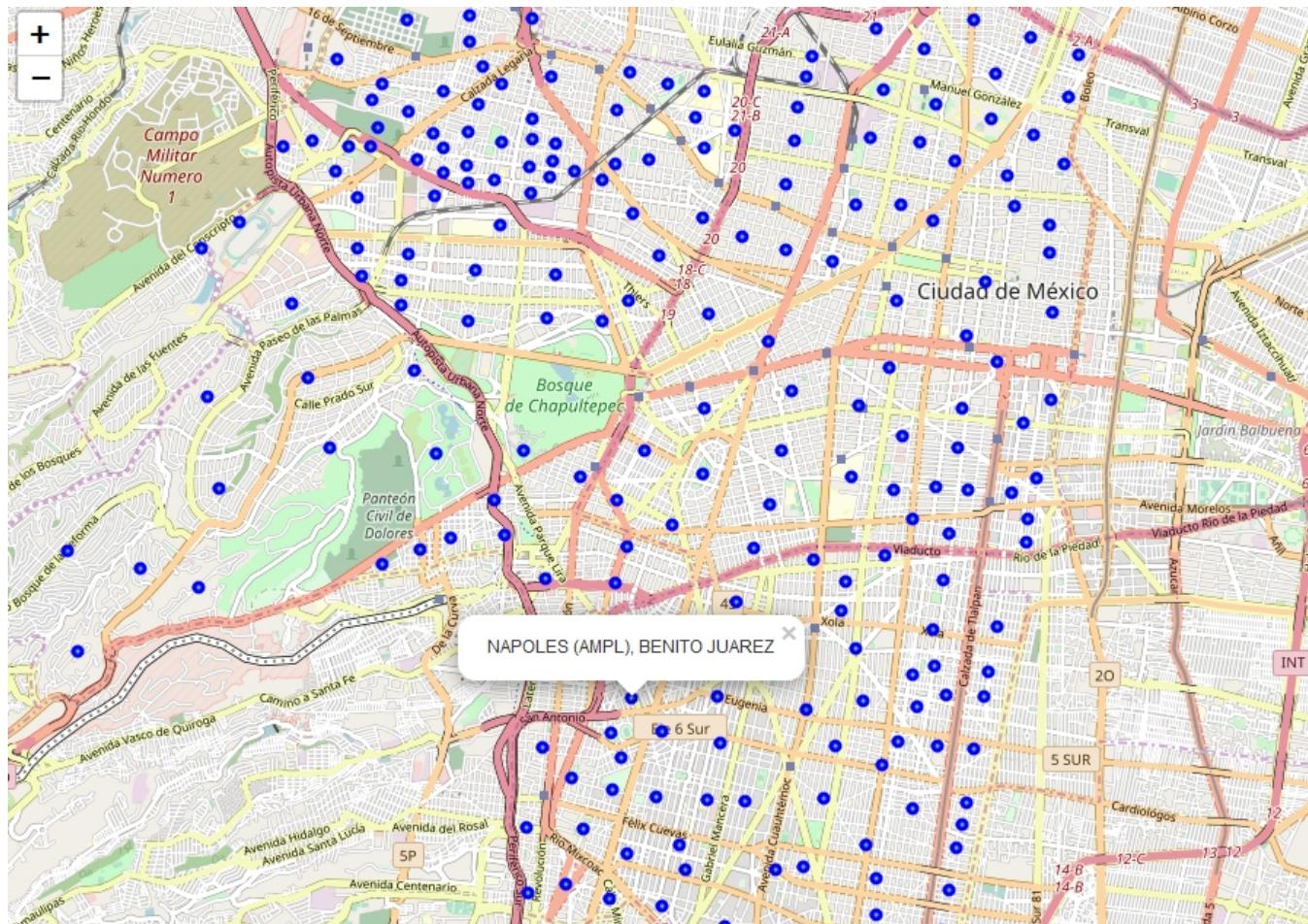
Neighborhoods in Mexico City Map / Mapa de Colonias en Ciudad de México

Using the final table of neighborhoods, specifically with the boroughs of Cuahutemoc, Miguel Hidalgo and Benito Juarez, we use again the tool to create a second map:

Mapa Folium

Usamos la primera tabla de las 1,808 colonias en la Ciudad de México y con la ayuda de «folium» podemos visualizarlas en un mapa.

La siguiente imagen muestra los centros geográficos de la colonia en la Ciudad de México:



Neighborhoods of Selected Town Halls Map / Mapa de Colonia de Alcaldías Seleccionadas

Nearby Places Categories

With the nearby venues data frame we need to see how many categories entries we have and if there are any null values.

Then we need to create a new data frame with the categories of interest from the particular person in example. Those categories are: **supermarket, park, yoga, coffee, tacos, music, pizza, library, spa and pharmacy**. This data frame will be called 'Interests'.

Now we have first verify if all of the 'Interests' categories data frame are in the 'categoriesID' of the nearby data frame and count them:

Categorías de Lugares Cercanos

Con la tabla de sitios cercanos necesitamos ver cuántos categorías tenemos y si hay datos con categorías nulas.

Luego necesitamos crear una tabla con las categorías de los intereses de la persona en particular del ejemplo. Estas categorías son: **supermercado, parque, yoga, café, tacos, música, pizza, librería, spa y farmacia**. Esta tabla se llamará 'Interests'.

Ahora tenemos primero que verificar si todas las categorías en la tabla de 'Interests' se encuentran en las 'categoriesID' de la tabla de sitios cercanos y contabilizarlos:

15708 entries were removed based on "Venue Category" not being related to interests
 ['Coffee Shop' 'Spa' 'Bookstore' 'Café' 'Park' 'Music Venue' 'Taco Place'
 'Pharmacy' 'Pizza Place' 'Supermarket' 'Grocery Store' 'Yoga Studio'
 'Farmers Market' 'Market' 'Garden' 'Cafeteria' 'Library']

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Distance	Venue Category
9	LOMAS DE CHAPULTEPEC	19.422841	-99.215794	Starbucks	19.419363	-99.215274	391	Coffee Shop
13	LOMAS DE CHAPULTEPEC	19.422841	-99.215794	La Llorena	19.418286	-99.214768	518	Coffee Shop
14	LOMAS DE CHAPULTEPEC	19.422841	-99.215794	Cafe Vida	19.426790	-99.213686	492	Coffee Shop
16	LOMAS DE CHAPULTEPEC	19.422841	-99.215794	Hand Spa	19.427244	-99.211779	646	Spa
17	LOMAS DE CHAPULTEPEC	19.422841	-99.215794	Gandhi	19.428012	-99.217575	605	Bookstore

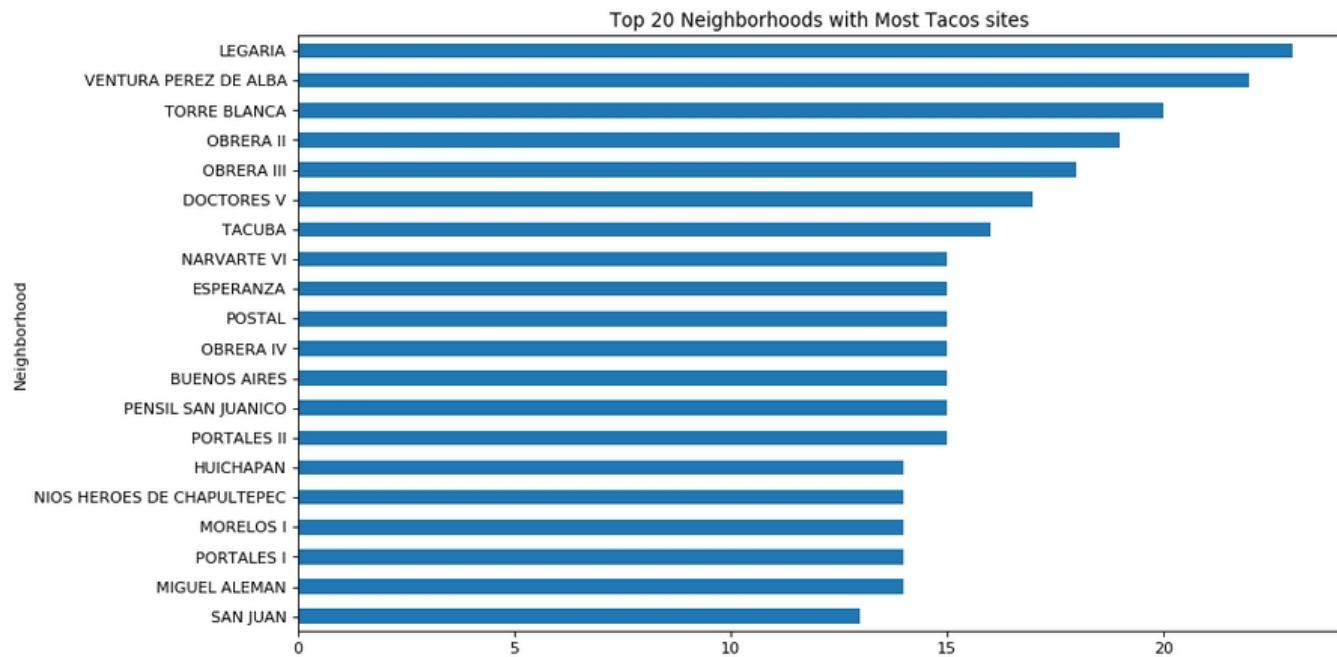
Verification of interest categories / Verificación de categorías de interés

Then we visualize for the best five categories, which neighborhoods has the more venues of those categories. In this case, we found that the categories that have more sites in the selected town halls were: Tacos Places, Pizza Places, Coffee Shops, Supermarkets and Parks.

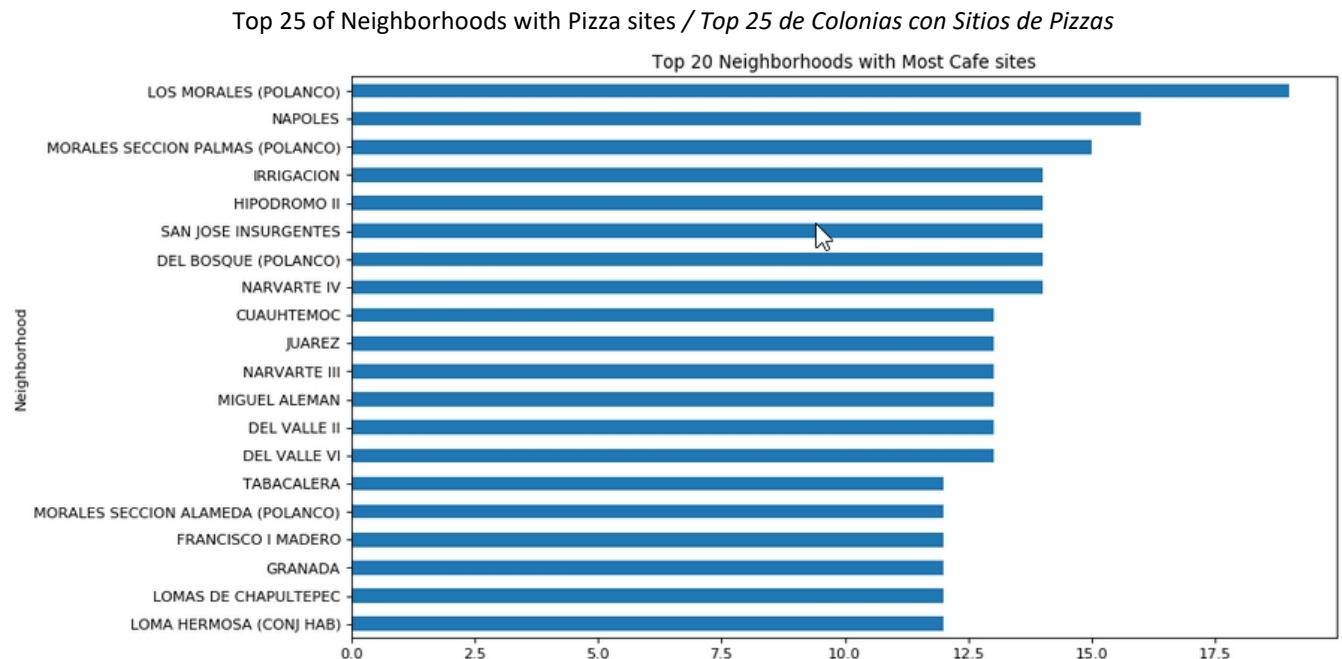
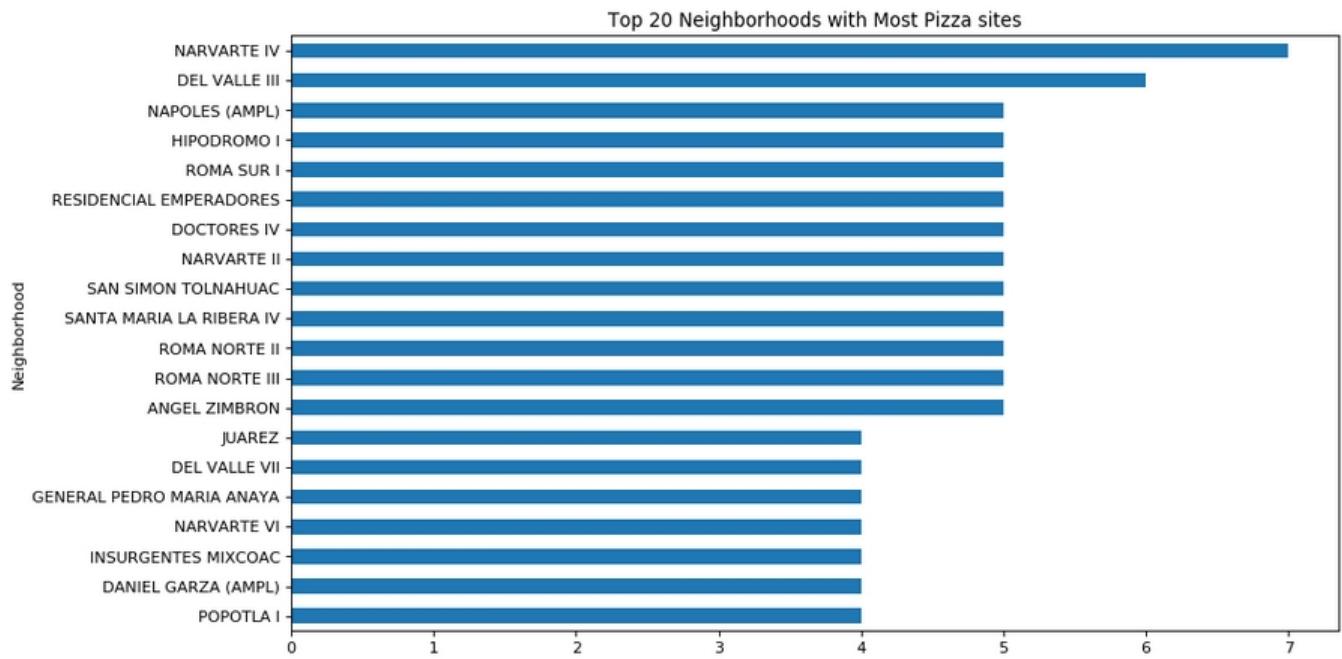
The following images shows the top 20 neighborhoods with more venues per categories:

Luego visualizamos para las cinco mejores categorías, cuáles colonias cuentan con más locales de dicha categoría. En este caso, encontramos que las categorías que tenían más locales en las alcaldías seleccionadas eran: Tacos, Pizza, Café, Supermercado y Parques.

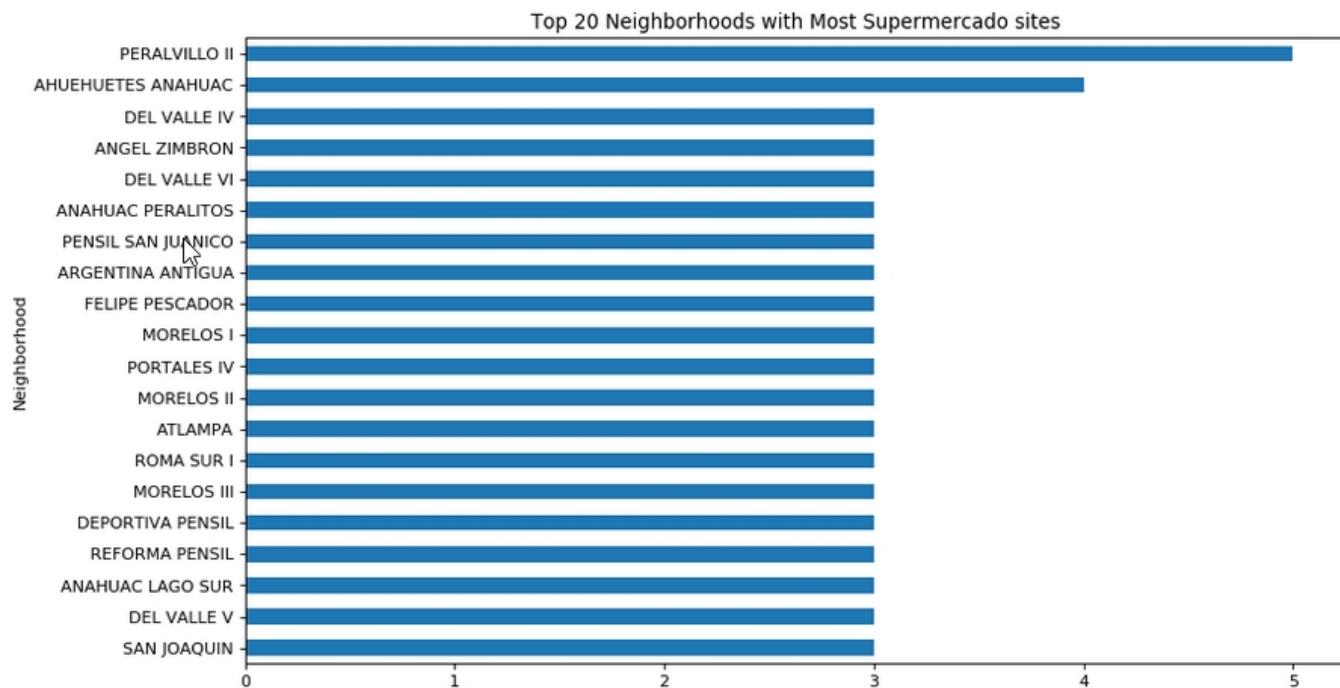
Las siguientes imágenes muestran las 20 colonias con más locales por categoría:



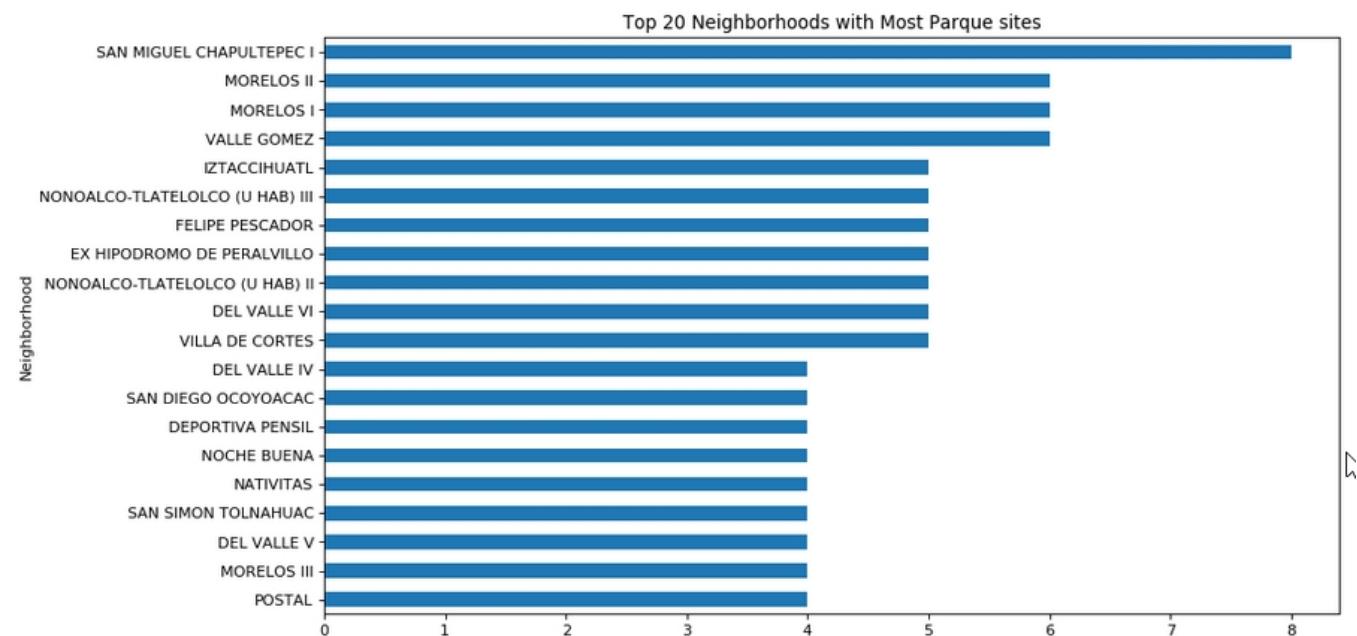
Top 25 of Neighborhoods with Tacos sites / Top 25 de Colonia con Sitos de Tacos



Top 25 of Neighborhoods with Coffee sites / Top 25 de Colonia con Sitos de Cafés



Top 25 of Neighborhoods with Supermarkets sites / Top 25 de Colonia con Sitos de Supermercados



Top 25 of Neighborhoods with Parks sites / Top 25 de Colonia con Sitos de Parques

With the information of venues of interest related to neighborhoods in selected town halls, we can proceed to show the results in the next chapter.

Con la información de los sitios de interés relacionada con las colonias de las alcaldías seleccionadas, podemos proceder a mostrar los resultados en el siguiente capítulo.

7. Results / Resultados

Results Modeling

The data frame with the results of the neighborhoods with the count of venues in each category is ready, so we need to add a value and apply a factor to define the clusters.

By applying the factor to the sum of count values, we can get a fractional number that we need to convert to an integer to know how many clusters we have and found that we have 6 clusters.

The image below shows the final table with the cluster value:

[64] :	COLONIA	ALCALDIA	Latitude	Longitude	Cafe	Farmacia	Libreria	Musica	Parque	Pizza	Spa	Supermercado	Tacos	Yoga	Value	
0	LOMAS DE CHAPULTEPEC	MIGUEL HIDALGO	19.422841	-99.215794	12	1	1	1	1	1	3		0	6	0	4
1	LOMAS DE REFORMA	MIGUEL HIDALGO	19.401682	-99.235472	7	1	1	1	2	0	0		2	2	1	4
2	LOMAS DE REFORMA (LOMAS DE CHAPULTEPEC)	MIGUEL HIDALGO	19.410616	-99.226249	2	1	1	0	3	0	0		1	0	0	2
3	IRRIGACION	MIGUEL HIDALGO	19.442955	-99.209936	14	1	0	0	3	0	1		2	3	0	3
4	DEL BOSQUE (POLANCO)	MIGUEL HIDALGO	19.434219	-99.209404	14	1	0	0	2	0	1		2	6	0	3

Neighborhoods with Cluster Value and Venues Count / Colonias con Valor de Grupo y Conteo de Sitos

Clusters Visualization

We plot the neighborhoods in a map with a color that identifies the clusters. The greater the cluster, the best qualification of the neighborhood.

- Best neighborhoods – Value of 5 (orange dots)
- Second best neighborhoods – Value of 4 (green dots)
- Third best neighborhoods – Value of 3 (light blue dots)
- Fourth best neighborhoods – Value of 2 (blue dots)
- Second worst neighborhoods – Value of 1 (purple dots)
- Worst neighborhoods – Value of 0 (red dots)

You can see the clusters in the map image below:

Modelaje de Resultados

La tabla con los resultados de las colonias con la cuenta de sitios de cada categoría está lista, y ahora necesitamos añadir un valor y aplicarle un factor para definir los grupos.

Al aplicar el factor a la suma de la cuenta como valor, obtenemos un número fraccionario que tenemos que convertir en integral para saber cuántos grupos tenemos y encontramos que son 6.

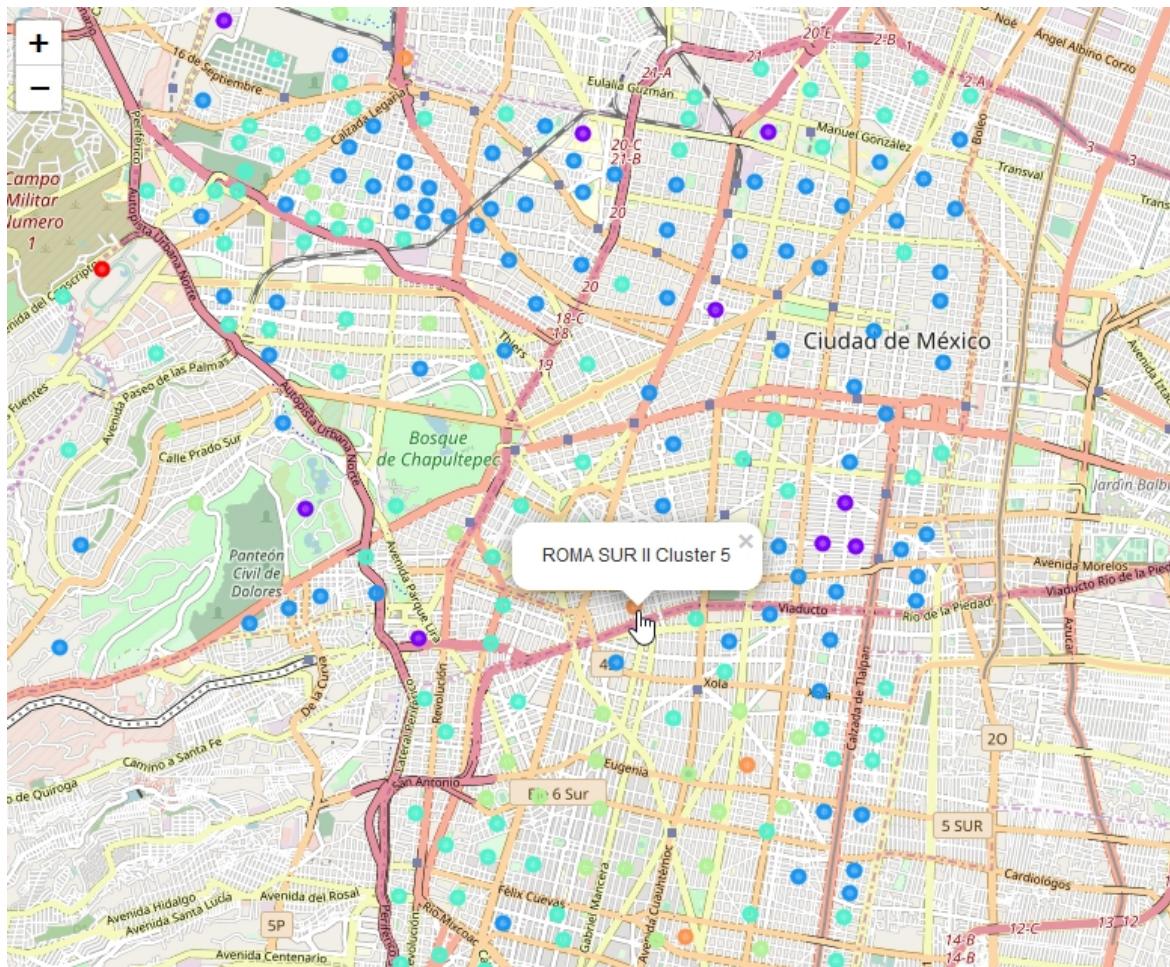
La siguiente imagen muestra la tabla final con el valor del grupo:

Visualización de Grupos

Delineamos las colonias en un mapa con colores para identificar los grupos. Entre más alto el valor del grupo, mejor calificación tiene la colonia.

- Mejores colonias – Valor de 5 (puntos naranjas)
- Segundas mejores colonias – Valor de 4 (puntos verdes)
- Terceras mejores colonias – Valor de 3 (puntos azul claro)
- Cuartas mejores colonias – Valor de 2 (puntos azules)
- Segundas peores colonias – Valor de 1 (puntos morados)
- Peores colonias – Valor de 0 (puntos rojos)

Pueden observar los grupos en la imagen del mapa siguiente:



Clusters Map of Neighborhoods / *Mapa de Grupos de Colonia*

Clusters Lists

The clusters analysis shows the following results:

- Cluster 5 Neighborhoods: 4
 - Cluster 4 Neighborhoods: 32
 - Cluster 3 Neighborhoods: 88
 - Cluster 2 Neighborhoods: 81
 - Cluster 1 Neighborhoods: 9
 - Cluster 0 Neighborhoods: 1

The following images show the lists of neighborhoods in each cluster:

Listas de Grupos

El análisis de los grupos muestra los siguientes resultados:

- *Grupo 5 de Colonias: 4*
 - *Grupo 4 de Colonias: 32*
 - *Grupo 3 de Colonias: 88*
 - *Grupo 2 de Colonias: 82*
 - *Grupo 1 de Colonias: 9*
 - *Grupo 0 de Colonias: 1*

Las siguientes imágenes muestran las listas de colonias en cada grupo:

This cluster have 4 neighborhoods												
	COLONIA	Cafe	Farmacia	Libreria	Musica	Parque	Pizza	Spa	Supermercado	Tacos	Yoga	Value
1221	ROMA SUR II	7	2	1	1	1	2	1		1	8	1
1276	NARVARTE VI	11	3	1	1	1	4	2		1	15	1
1559	RESIDENCIAL EMPERADORES	6	2	1	1	1	5	1		2	9	1
1621	ANGEL ZIMBRON	3	1	1	1	2	5	1		3	13	2

Neighborhoods in Cluster 5 / *Colonias en Grupo 5*

This cluster have 32 neighborhoods

	COLONIA	Cafe	Farmacia	Libreria	Musica	Parque	Pizza	Spa	Supermercado	Tacos	Yoga	Value
0	LOMAS DE CHAPULTEPEC	12	1	1	1	1	1	3	0	6	0	4
1	LOMAS DE REFORMA	7	1	1	1	2	0	0	2	2	1	4
31	5 DE MAYO	8	1	1	0	2	1	1	2	3	0	4
38	PENSIL SUR	7	1	1	0	2	1	1	1	4	0	4
379	VERTIZ NARVARTÉ	10	2	1	1	1	2	3	2	11	0	4
811	DEL VALLE V	11	1	2	0	4	3	2	3	5	1	4
813	DEL VALLE II	13	1	1	0	1	3	3	1	4	0	4
1218	HIPODROMO I	6	0	1	1	3	5	2	0	3	2	4
1223	ROMA SUR I	5	1	1	1	1	5	1	3	4	0	4
1400	HIPODROMO II	14	1	3	0	2	1	1	0	3	2	4
1459	NARVARTÉ IV	14	1	0	1	0	7	1	1	11	1	4
1461	PORTALES IV	9	1	1	1	2	2	3	3	7	0	4
1519	NARVARTÉ V	10	3	1	1	1	3	4	0	11	0	4
1520	PORTALES III	8	2	0	1	1	3	2	3	11	0	4
1541	NAPOLES (AMPL)	11	1	0	0	2	5	2	1	7	1	4

Neighborhoods in Cluster 4 / *Colonias en Grupo 4*

This cluster have 88 neighborhoods

	COLONIA	Cafe	Farmacia	Libreria	Musica	Parque	Pizza	Spa	Supermercado	Tacos	Yoga	Value
3	IRRIGACION	14	1	0	0	3	0	1	2	3	0	3
4	DEL BOSQUE (POLANCO)	14	1	0	0	2	0	1	2	6	0	3
5	DANIEL GARZA (AMPL)	4	0	0	1	3	4	0	0	12	2	3
8	LEGARIA (U HAB)	4	2	0	0	3	0	1	1	9	0	3
32	MORALES SECCION ALAMEDA (POLANCO)	12	1	0	0	0	1	2	1	1	0	3
35	TORRE BLANCA	2	1	0	1	1	1	0	3	20	0	3
350	INDEPENDENCIA	7	1	0	0	1	2	2	2	13	0	3
351	ALBERT	5	1	0	0	1	2	0	1	11	0	3
374	SAN PEDRO DE LOS PINOS	8	2	0	0	3	2	0	1	4	0	3
375	NAPOLES	16	0	0	1	1	4	2	0	3	0	3
376	NOCHE BUENA	7	0	0	1	4	3	1	1	11	0	3
380	MODERNA	5	0	0	1	4	1	0	2	8	0	3
382	INSURGENTES MIXCOAC	11	1	0	1	2	4	0	0	4	0	3
383	XOCO	8	0	1	1	3	2	0	1	4	0	3
386	RINCON DEL BOSQUE	7	0	1	0	2	1	2	0	4	0	3

Neighborhoods in Cluster 3 / *Colonias en Grupo 3*

This cluster have 81 neighborhoods

		COLONIA	Cafe	Farmacia	Libreria	Musica	Parque	Pizza	Spa	Supermercado	Tacos	Yoga	Value
2	LOMAS DE REFORMA (LOMAS DE CHAPULTEPEC)	2	1	1	0	3	0	0		1	0	0	2
6	IGNACIO MANUEL ALTAMIRANO	0	0	1	1	3	1	0		0	11	0	2
7	LEGARIA	4	0	0	0	1	1	0		3	23	0	2
9	MARINA NACIONAL (U HAB)	8	0	0	0	0	0	3	0	3	11	0	2
36	UN HOGAR PARA NOSOTROS	3	0	0	0	1	2	0		0	10	0	2
353	ATENOR SALAS	5	0	0	0	1	3	0		1	9	0	2
377	NATIVITAS	9	0	0	1	4	3	0		0	11	0	2
378	SAN SIMON TICUMAC	7	0	0	0	1	2	0		3	12	0	2
381	ALAMOS II	8	0	0	0	2	0	0		2	11	0	2
384	PENSIL NORTE	4	1	0	0	3	0	0		2	12	0	2
385	PENSIL SAN JUANICO	6	0	0	0	2	2	0		3	15	0	2
387	MORALES SECCION PALMAS (POLANCO)	15	1	0	0	0	0	2		1	0	0	2
643	ASTURIAS	4	0	0	1	0	1	0		2	6	0	2
657	CENTRO VI	7	3	0	1	0	0	1		0	6	0	2
658	CENTRO VIII	6	0	4	0	0	2	0		2	5	0	2

Neighborhoods in Cluster 2 / *Colonias en Grupo 2*

This cluster have 9 neighborhoods

		COLONIA	Cafe	Farmacia	Libreria	Musica	Parque	Pizza	Spa	Supermercado	Tacos	Yoga	Value
37	PLUTARCO ELIAS CALLES	4	0	0	0	1	0	0		0	5	0	1
1004	TABACALERA	12	0	0	0	0	2	0		0	4	0	1
1405	OBRERA III	3	0	0	0	0	0	0		3	18	0	1
1406	OBRERA II	4	0	0	0	0	0	0		2	19	0	1
1493	OBRERA IV	4	0	0	0	0	0	0		1	15	0	1
1566	NONOALCO-TLATELOLCO (U HAB) I	0	0	0	2	3	0	0		0	10	0	1
1604	TACUBAYA	5	0	0	0	0	3	0		0	11	0	1
1656	BOSQUE DE CHAPULTEPEC I, II Y III SECCIONES	7	0	0	0	3	0	0		0	2	0	1
1753	SAN LORENZO TLALTENANGO	0	0	0	0	2	0	0		0	3	0	1

Neighborhoods in Cluster 1 / *Colonias en Grupo 1*

This cluster have 1 neighborhoods

		COLONIA	Cafe	Farmacia	Libreria	Musica	Parque	Pizza	Spa	Supermercado	Tacos	Yoga	Value
1748	MILITAR 1 K LOMAS DE SOTELO (U HAB)	0	0	0	0	0	0	0	0	0	1	0	0

Neighborhoods in Cluster 0 / *Colonias en Grupo 0*

8. Conclusion / Conclusión

As a result, we can say that the neighborhood Roma Sur II on the Town Hall Cuauhtemoc, the neighborhood Angel Zimbron on the Town Hall Miguel Hidalgo, and the neighborhoods Narvarte VI and Residencial Emperadores on the Town Hall Benito Juarez are the best neighborhoods to live in Mexico City.

Machine learning and clustering algorithms can be applied to multi-dimensional datasets to find similarities and patterns in the data. Clusters of neighborhoods of specific interests can be generated using venue exploring data, but analysis models are only as good as the input into them. Unfortunate Foursquare ‘Places API’ service has limitations on the number of calls in the free developer account that delivers not so perfect results. Although the exercise scope defined in the objective and purpose are fulfill, a different approach to answer the original question could delivers better results.

This project is not finished and could be expanded on with a number of different approaches. Foursquare’s API could be used to search the specific venues instead of explore nearby venues and filter them. Also other methodologies, like DBSCAN clustering algorithm, could be better to define clusters ignoring outliers. The clustering model could become the basis for a recommendation system aimed to provide neighborhoods of similar interest to users.

As a surplus of this project we include an Annex with a K-Mean algorithm to define clusters of neighborhoods with similar venues of interest in the three town halls selected.

Como resultado podemos decir que la colonia Roma Sur II de la Alcaldía Cuauhtémoc, la colonia Ángel Zimbrón en la alcaldía Miguel Hidalgo y las colonias Narvarte VI y Residencial Emperadores de la alcaldía Benito Juárez son las mejores colonias para vivir en la Ciudad de México.

Machine Learning y algoritmos de grupos se pueden aplicar a bases de datos multidimensionales para encontrar similitudes y patrones en la información. Grupos de colonias de intereses específicos se pueden generar usando la exploración de sitios, pero los modelos de análisis son tan buenos como el registro de estos. Desafortunadamente Foursquare tiene limitaciones en el número de llamadas en la cuenta gratis de desarrollador que entrega resultados no tan buenos. A pesar de que el alcance del objetivo y propósito se cumplió, un enfoque diferente para contestar la pregunta original puede entregar mejores resultados.

Este proyecto no está terminado y podrá ser explorado con un número de diferentes enfoques. Foursquare puede ser usado con la función de localización de sitios específicos en lugar de explorar sitios cercanos y filtrarlos. También otras metodologías, como el algoritmo DBSCAN, pueden ser mejores al definir grupos ignorando extremos. Los modelos de grupos se pueden convertir en la base recomendada para mostrar colonias con intereses similares.

Como valor agregado de este proyecto incluimos un anexo con el algoritmo K-Mean para definir grupos de colonias con sitios de intereses similares en las tres alcaldías seleccionadas.

Annex – K-Mean Algorithm / Anexo – Algoritmo K-Mean

Cluster Modeling

We use the Scikit-learn's K-Means tool to cluster the results of the neighborhoods lists of interests based on the interest percentage.

The images below shows the percentages of venues data, the data being scaled and the K-mean model being created, and the data frame created with top venues categories:

	Neighborhood	Cafe	Farmacia	Libreria	Musica	Parque	Pizza	Spa	Supermercado	Tacos	Yoga
0	10 DE ABRIL	0.006887	0.005988	0.000000	0.0000	0.005602	0.000000	0.007874	0.006579	0.003639	0.0
1	16 DE SEPTIEMBRE	0.002066	0.011976	0.000000	0.0000	0.000000	0.005155	0.000000	0.003289	0.001819	0.0
2	5 DE MAYO	0.005510	0.005988	0.011236	0.0000	0.005602	0.002577	0.007874	0.006579	0.001819	0.0
3	ACACIAS	0.007576	0.005988	0.000000	0.0125	0.005602	0.005155	0.007874	0.000000	0.003032	0.0
4	ACTIPAN	0.006887	0.000000	0.000000	0.0000	0.002801	0.002577	0.000000	0.000000	0.002426	0.0

Percentage Neighborhood Venue Data / Datos de Porcentaje de Sitios por Colonia

```
# Define Clusters
kclusters = 5

venue_grouped_clustering = venue_mean.drop('Neighborhood', 1)
kmeans = KMeans(n_clusters=kclusters, random_state=0).fit(venue_grouped_clustering)

kmeans.labels_

array([1, 1, 4, 2, 1, 1, 1, 1, 1, 1, 4, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0,
       3, 1, 2, 2, 1, 2, 1, 4, 4, 2, 1, 1, 4, 4, 4, 4, 2, 4, 2, 2, 4, 0,
       0, 1, 2, 1, 0, 4, 1, 0, 1, 1, 3, 3, 0, 3, 3, 4, 1, 4, 4, 2, 2, 1,
       3, 4, 3, 1, 1, 2, 1, 1, 2, 3, 4, 2, 1, 4, 4, 0, 0, 0, 2, 2, 3, 2,
       3, 1, 2, 2, 4, 1, 1, 3, 1, 1, 4, 2, 3, 2, 4, 1, 3, 1, 1, 1, 1, 4,
       1, 3, 1, 3, 1, 1, 2, 2, 3, 3, 1, 1, 1, 3, 3, 1, 1, 4, 3, 3, 3, 2,
       1, 3, 2, 1, 2, 2, 1, 1, 1, 1, 0, 1, 1, 2, 1, 1, 4, 2, 1, 1,
       3, 2, 1, 0, 4, 4, 1, 1, 3, 3, 3, 1, 2, 1, 3, 3, 3, 4, 1, 2,
       3, 2, 1, 1, 2, 1, 0, 0, 1, 1, 1, 2, 2, 2, 1, 1, 1, 2, 1, 2,
       1, 1, 2, 1, 2, 1, 1, 4, 1, 1, 1, 3, 3, 2, 2, 2, 1, 1], dtype=int32)
```

Clustering Neighborhood Venue Data / Agrupación de Datos de Sitios por Colonia

	COLONIA	Latitude	Longitude	Cluster Labels	1st Top Venue Category	2nd Top Venue Category	3rd Top Venue Category	4th Top Venue Category	5th Top Venue Category
0	LOMAS DE CHAPULTEPEC	19.422841	-99.215794	3	Spa	Musica	Libreria	Cafe	Farmacia
1	LOMAS DE REFORMA	19.401682	-99.235472	2	Yoga	Musica	Libreria	Supermercado	Farmacia
2	LOMAS DE REFORMA (LOMAS DE CHARULTEPEC)	19.410616	-99.226249	4	Libreria	Parque	Farmacia	Supermercado	Cafe
3	IRRIGACION	19.442955	-99.209936	1	Cafe	Parque	Spa	Supermercado	Farmacia
4	DEL BOSQUE (POLANCO)	19.434219	-99.209404	1	Cafe	Spa	Supermercado	Farmacia	Parque

Top Venues Categories per Neighborhood /Principales Categoría de Sitios por Colonia

Cluster Visualization

Using the code of folium again, we visualize the neighborhoods with similar interest venues, coloring each neighborhood point based on cluster label.

The map below show the neighborhoods clusters in the town hall selected:

Modelo de Grupos

Usamos la herramienta Scikit-learn K-Means para agrupar los resultados de las colonias listadas basadas en el porcentaje de intereses.

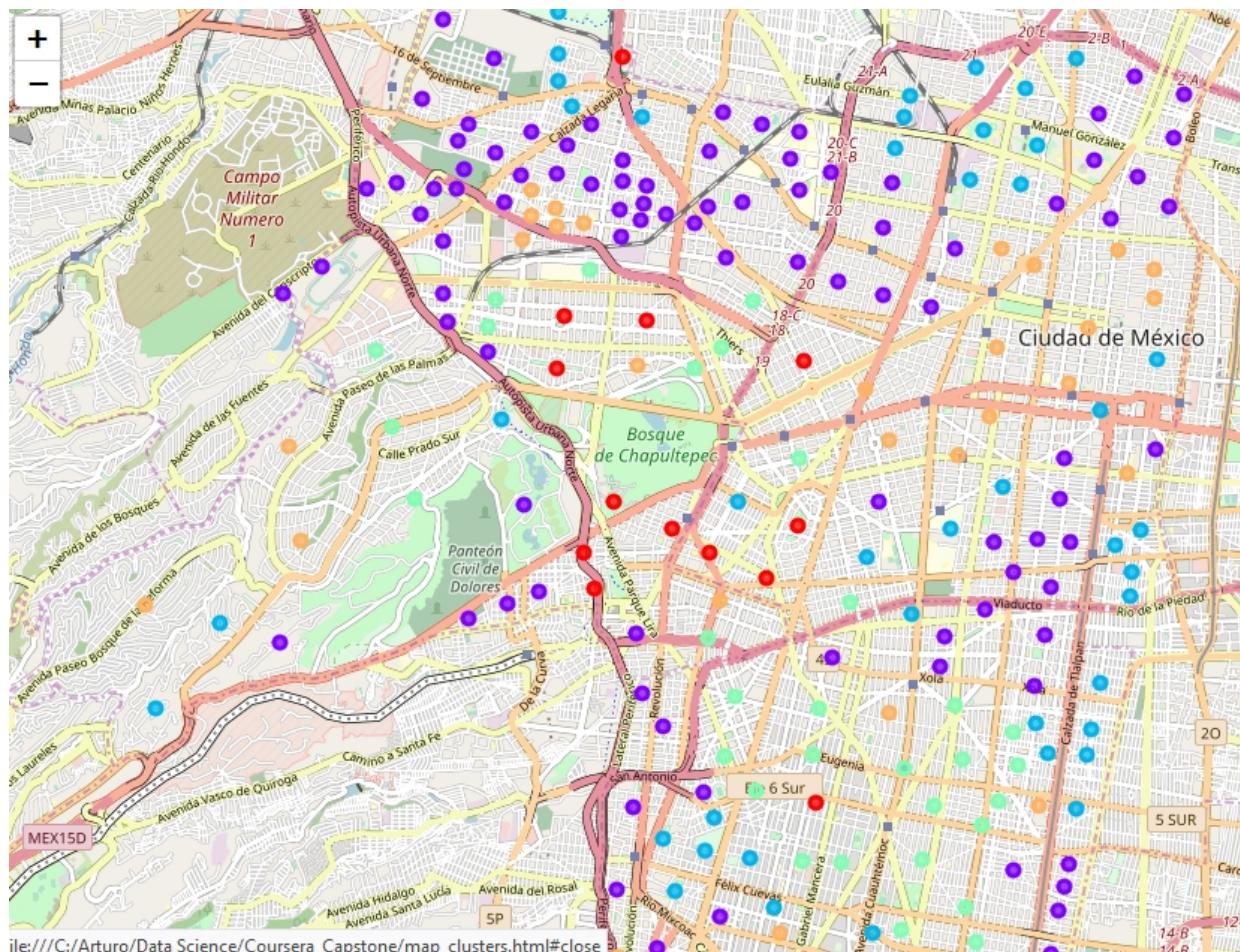
Las imágenes siguientes muestra los datos de porcentajes de los sitios, los datos escalados y el modelo K-mean siendo cread y la tabla creada con las principales categorías de sitios:

	Neighborhood	Cafe	Farmacia	Libreria	Musica	Parque	Pizza	Spa	Supermercado	Tacos	Yoga
0	10 DE ABRIL	0.006887	0.005988	0.000000	0.0000	0.005602	0.000000	0.007874	0.006579	0.003639	0.0
1	16 DE SEPTIEMBRE	0.002066	0.011976	0.000000	0.0000	0.000000	0.005155	0.000000	0.003289	0.001819	0.0
2	5 DE MAYO	0.005510	0.005988	0.011236	0.0000	0.005602	0.002577	0.007874	0.006579	0.001819	0.0
3	ACACIAS	0.007576	0.005988	0.000000	0.0125	0.005602	0.005155	0.007874	0.000000	0.003032	0.0
4	ACTIPAN	0.006887	0.000000	0.000000	0.0000	0.002801	0.002577	0.000000	0.000000	0.002426	0.0

Visualización de Grupos

Usando el código de folium nuevamente, podemos visualizar las colonias con sitios de interés similares, coloreando cada punto de colonia en base a la etiqueta de cada grupo.

El mapa siguiente muestra los grupos de colonias en las alcaldías seleccionadas:



Map of Similar Neighborhoods Clusters / Mapa de Grupos de Colonias Similares

Results

We define only 6 clusters as similar as the previous analysis and the results of the Cluster 5 are:

Resultados

Se definieron solo 6 grupos de manera similar al análisis previo y los resultados del Grupo 5 son:

	COLONIA	1st Top Venue Category	2nd Top Venue Category	3rd Top Venue Category	4th Top Venue Category	5th Top Venue Category
2	LOMAS DE REFORMA (LOMAS DE CHAPULTEPEC)	Libreria	Parque	Farmacia	Supermercado	Cafe
31	5 DE MAYO	Libreria	Spa	Supermercado	Farmacia	Parque
38	PENSIL SUR	Libreria	Spa	Farmacia	Parque	Cafe
658	CENTRO VIII	Libreria	Supermercado	Pizza	Cafe	Tacos
661	ROMA NORTE II	Yoga	Pizza	Libreria	Cafe	Tacos
748	JUAREZ	Libreria	Yoga	Pizza	Cafe	Tacos
749	TRANSITO	Libreria	Pizza	Tacos	Cafe	Supermercado
808	BOSQUES DE LAS LOMAS	Libreria	Supermercado	Farmacia	Parque	Cafe
809	LOMAS DE BARRILACO (LOMAS DE CHAPULTEPEC)	Farmacia	Libreria	Spa	Supermercado	Parque
831	ESCANDON I	Libreria	Spa	Pizza	Cafe	Farmacia
1067	CENTRO III	Libreria	Musica	Pizza	Cafe	Tacos
1214	CENTRO I	Libreria	Supermercado	Tacos	Parque	Pizza
1215	CENTRO II	Libreria	Tacos	Pizza	Supermercado	Cafe
1217	CENTRO IV	Libreria	Cafe	Supermercado	Pizza	Yoga

Cluster 5 Results / Resultado del Grupo 5

References / Referencias

- Información de Colonias de la Agencia Digital de Innovación Pública del Gobierno de la Ciudad de México: (<https://datos.cdmx.gob.mx/explore/dataset/coloniascdmx/information/>)
- Foursquare API Places Documentation (<https://developer.foursquare.com/docs>)

Appedix – Mexico City / Apéndice – Ciudad de México

Denominations

When we talk of «Mexico City» it's easy to get confused with «CDMX», «Federal District», «Metropolitan Zone of Mexico Valley», «State of Mexico» and «Mexico» (the country). The city is the country capital and by mid XIX century an area from the «State of Mexico» was delimited and called «Federal District» (DF) and was divided in 16 delegations (the equivalent to municipalities or boroughs).

When the urban stain exceeded the «Federal District» limits, the city was denominated as «Metropolitan Zone of Mexico City» (ZMCM) or «Metropolitan Zone of Mexico Valley» (ZMVM). In 2017 their district status change and became the 32th state of the country and was called as the state of «Mexico City» (CDMX) as well as the change of name of delegations to «town halls» (boroughs).

Clarifications

In this work we do not made an analysis of the objective based on the ZMVM, which includes the 16 town halls from «CDMX», 23 municipalities form «State of Mexico» and 1 municipality form the state of «Hidalgo». For this work we will only use the information form the 16 town halls and we referrer the city as «Mexico City» in order to not create confusion with «ZMVM».

Denominaciones

Cuando se habla de la Ciudad de México es fácil confundirse con CDMX, Distrito Federal, Zona Metropolitana del Valle de México, Estado de México y México (el país). La ciudad es la capital del país y a mediados del siglo XIX fue delimitada una zona del Estado de México, al que pertenecía, llamándolo Distrito Federal (DF) y se dividió en 16 delegaciones (equivalentes a municipios en el resto del país).

Al rebasar la mancha urbana los límites del Distrito Federal se denominó a la ciudad como Zona Metropolitana de la Ciudad de México (ZMCM) o Zona Metropolitana del Valle de México (ZMVM). En 2017 se cambió el estatus de distrito para convertirse en estado 32 del país y se le denominó «Ciudad de México» (CDMX) al estado, al tiempo de cambiar las delegaciones por «alcaldías».

Clarificaciones

En este trabajo no haremos el análisis del objetivo en base a la ZMVM que incluiría las 16 alcaldías de la CDMX, 23 municipios del Estado de México y 1 municipio del estado de Hidalgo. Para este trabajo solo utilizaremos la información de las 16 alcaldías y nos referiremos a la urbe como «Ciudad de México» para no crear confusión con la ZMVM.