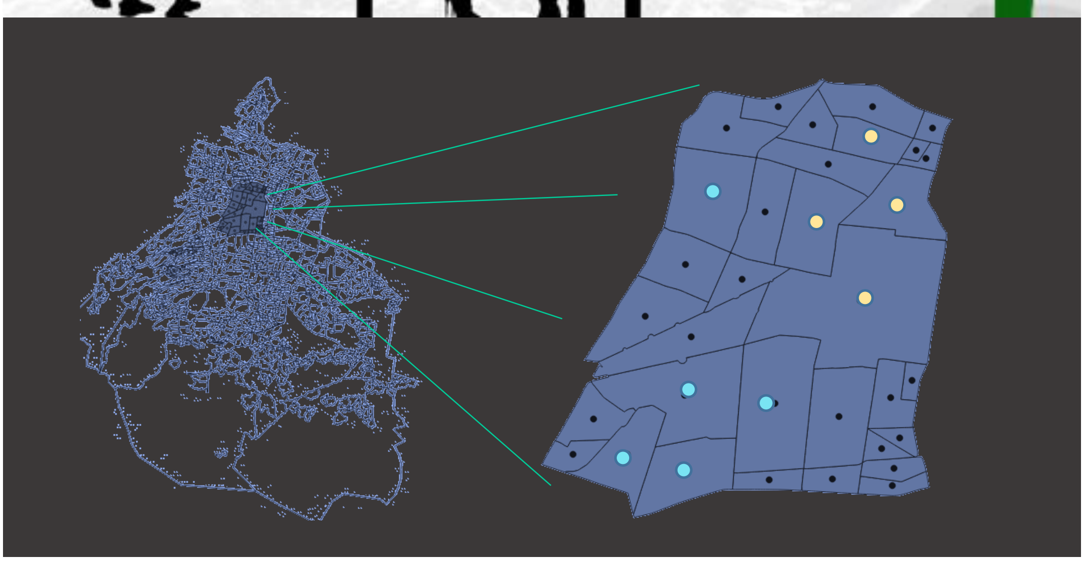
CAPSTONE PROJECT

Neighborhoods in central Mexico City
By parks availability and quality perception

Spatial analysis- clustering



Isaac Perez

INTRODUCTION AND THE PROBLEM

How are neighborhoods in central Mexico City similar or different to each other based on the availability and quality perception of their public spaces and particularity parks?

Can we grouped neighborhoods according to their parks availability and how people rate them?

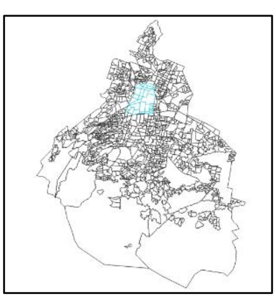
Are neighborhoods clustered with more and better parks associated with higher level of development and less poverty and the other way around?

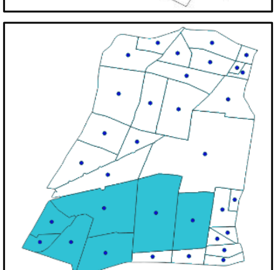
Public spaces such as parks are often unevenly distributed in cities. In some urban areas, usually those with high inequality, there are spatial patterns of concentrations and deficit of public spaces. Some areas might have not only more but also better public spaces than others. There are neighborhoods, on the other hand, that have comparatively few, less vibrant, even unsafe and poor quality public spaces or don't have any at all. This, in some cases, exacerbates, other forms of inequality and deprivation, prevents people to spend time outside and to meet with others. It might also mean they have less access to green spaces, impacting their health and the urban landscape.

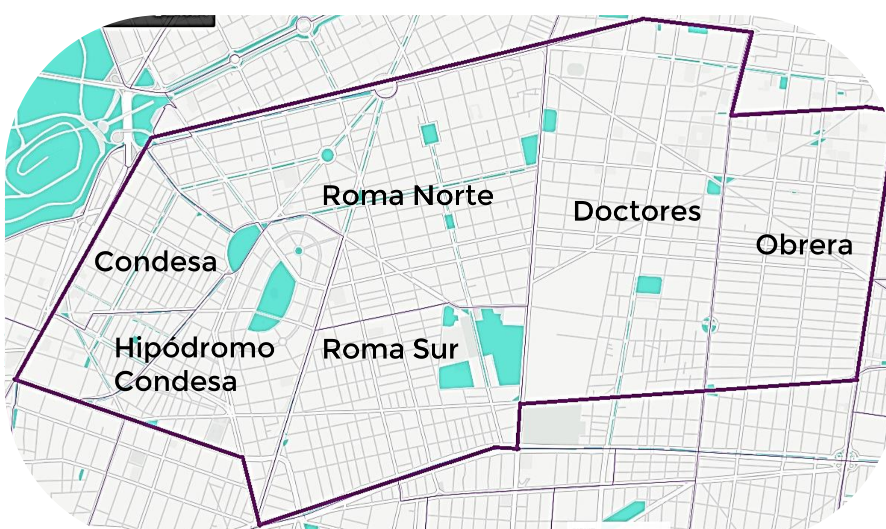
Mexico City stands as one of the largest metropolis in the world with about 22 million people and spanning over nearly 8 thousand square kilometers. The city is also full of contrast, with boroughs with a human development index comparable to Finland or France while also some comparable to countries in sub-Saharan Africa. About one third of the population lives in poverty and about 4 percent in extreme poverty, equivalent to 800 thousand people. The city also displays differences within and across boroughs and neighborhoods in terms of infrastructure, services and public spaces,

As shown in the map below, which shows neighborhoods in the central borough Cuauhtemoc and public spaces-parks in green, some neighborhoods in the west have more parks – and are more similar to each other in this regard, while the ones in the east are more similar to each other given their lack of parks. Cuauhtemoc houses about half a million people, areas with high, medium and low development.

Although this is something one could observed by looking a map like the one below and with this overlay, **there is not a tool, analysis or map that groups neighborhoods based on the availability and quality of parks. This prevents understanding this key dimension in the city, guide policy or interventions based on this knowledge and notion. This also would make more difficult for someone to take a decision to move to another neighborhood taking this aspect into consideration.**







THE GOAL

This project seeks to provide insights into the similarities and differences in neighborhood in central Mexico City (Cuauhtémoc Borough) based on the availability and ratings of parks. It intends to group – cluster the 33 neighborhoods in this borough according to these variables. It addition, it also aims at shedding light into how the resulted grouping is associated with the neighborhood social development.

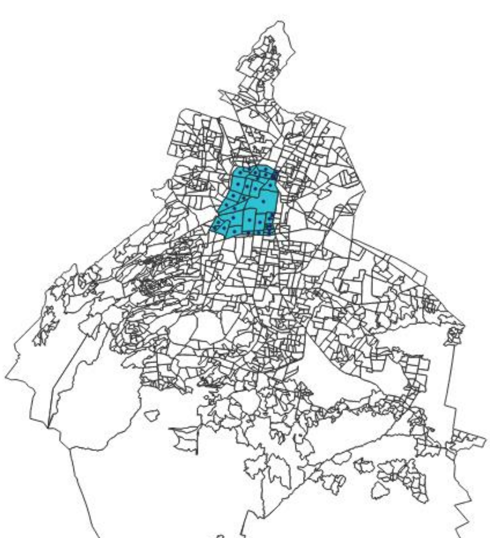
THE DATA

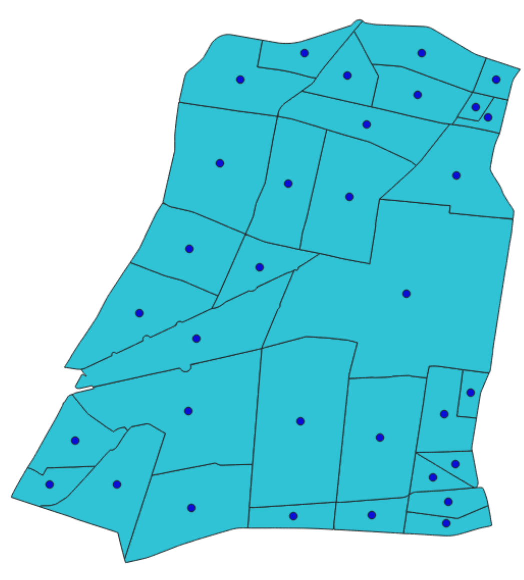
To develop this project data will be used from various sources including foursquare and leveraging the data available through the API.

1) The shape files of the Neighborhoods in Mexico City are available for from the city's Urban Development Department. The city has over a thousand neighborhoods distributed in 16 boroughs. For this particular project, only 33 that are part of the Cuauhtemoc borough will be filtered and used (highlighted in blue below). I have download the shape files, created new shape and json files with only Cuauhtemoc and also generated the centroids of each neighborhoods to have them both as points and polygons.
2) Data about the parks location and rating will be gathered using the foursquare API. There is data about 40 parks in Cuauhtemoc in a 4km kilometer radius from its centroid coordinates in their data set. Cuauhtemoc has about this radius, so this is an appropriate radius. Points that were to be outside the area will be excluded. There is also ratings for the venues in foursquare (see figure in the next page).
3) Data on the socio-economic development of neighborhoods is available from the city's social development department (See figure in the next page).

Using data on the parks locations and their ratings, two variables will be crated, number of parks and average parks ratings per neighborhood. Each neighborhood in a new data set-frame will have information on these two dimensions. K-means clustering then will be used to group neighborhoods based on these characteristics. The two variables will be normalized to facilitate the method execution. In addition, and once this step is completed, it will be explored what the most common development level of each group using bar plots and spatial overlays of these two dimensions.

1 Neighborhoods location data





2 Foursquare api data

Json file- request for parks

```
[35]: results = requests.get(url).json()
      results
```


```
[35]: {'meta': {'code': 200, 'requestId': '5ef946cc513bb60f5114b60b'},
       'response': {'venues': [{'id': '4cc8ae0b41e75481e20e5884',
         'name': 'Parque Alameda',
         'location': {'address': 'Av. Juárez 76',
          'crossStreet': 'Esq. con Azueta',
          'lat': 19.434515120135995,
          'lng': -99.14715262455131,
          'labeledLatLngs': [{'label': 'display',
            'lat': 19.434515120135995,
            'lng': -99.14715262455131}],
          'distance': 80,
          'postalCode': '06010',
```


Json to data frame

```
[39]: venues = results['response']['venues']

      # tranform venues into a dataframe
      dataframe = json_normalize(venues)
      dataframe.head(5)
```

	id	name	categories	referralId	hasPerk	location.address	location.crossStreet	location.la
0	4cc8ae0b41e75481e20e5884	Parque Alameda	[{'id': '4bf58dd8d48988d1fd941735', 'name': 'S...	v-1593395281	False	Av. Juárez 76	Esq. con Azueta	19.43451
1	4ec7df32f79041351d2d6868	Parque Ciudadela	[{'id': '4bf58dd8d48988d182941735', 'name': 'T...	v-1593395281	False	NaN	NaN	19.42732
2	53e8320c498ee437284d3f15	Parque Olof Palme	[{'id': '4bf58dd8d48988d163941735', 'name': 'P...	v-1593395281	False	Rio Grijalva	Rio Danubio	19.43332
3	4da21ca29aa4721e3473091a	Parque José Martí	[{'id': '4bf58dd8d48988d164941735', 'name': 'P...	v-1593395281	False	Reforma	Hidalgo	19.43710

Csv file with parks in Cuauhtemoc

	id	name	categories	referralId	hasPerk	location.add	location.cros	location.lat	location.lng	location.labe	location.dist	lo
0	4cc8ae0b41e	Parque Alameda	[{'id': '4bf58c	v-159339528	False	Av. JuÃ¡rez 7	Esq. con Azu	19.4345151	-99.14715262	[{'label': 'dis	80	
1	4ec7df32f79	Parque Ciudadela	[{'id': '4bf58c	v-159339528	False			19.4273216	-99.14975289	[{'label': 'dis	918	
2	53e8320c498	Parque Olof Palme	[{'id': '4bf58c	v-159339528	False	Rio Grijalva	Rio Danubio	19.4333273	-99.17075567	[{'label': 'dis	2475	
3	4da21ca29aa	Parque JosÃ© MartÃ-	[{'id': '4bf58c	v-159339528	False	Reforma	Hidalgo	19.4371054	-99.14597898	[{'label': 'dis	248	
5	5e694bde14l	Dorothy Gaynor Parque Alm	[{'id': '4bf58c	v-159339528	False			19.435038	-99.146905	[{'label': 'dis	43	
6	5488464e498	parque para perros Pushkin	[{'id': '4bf58c	v-159339528	False	Morelia	Colima	19.4209741	-99.15446091	[{'label': 'dis	1757	
7	55fd8f89498	Parque Galeria	[{'id': '4bf58c	v-159339528	False	Puebla 170	Roma Norta	19.4223491	-99.16258604	[{'label': 'dis	2154	
8	4d037f229d3	Parque San Fernando	[{'id': '4bf58c	v-159339528	False	Puente de A	San Fernand	19.4389003	-99.14609471	[{'label': 'dis	426	
9	4fdfe1bbe4b	Parque	[{'id': '4bf58c	v-159339528	False			19.4274756	-99.1428993	[{'label': 'dis	977	
10	5277f9b111d	Parque Via 198	[{'id': '50327(	v-159339528	False			19.4353684	-99.16816103	[{'label': 'dis	2193	
11	4ebf01607ee	Parque de Los Cuatro Vient	[{'id': '4bf58c	v-159339528	False			19.4626279	-99.13855394	[{'label': 'dis	3183	
12	50282730e4b	Parque Recreativo	[{'id': '4bf58c	v-159339528	False			19.4301949	-99.14349365	[{'label': 'dis	686	
13	5249ecb2498	Parque De Los CaÃ±ones	[{'id': '4bf58c	v-159339528	False			19.4301762	-99.15020693	[{'label': 'dis	641	

2 Socio-economic development level data

Cuauhtemoc Borough

33 Neighborhoods

Development Level (High, medium and low

Nombre de la Delegación	Nombre de la Colonia o Barrio	Habitantes	Indice de Desarrollo Social		
			Valor	Estrato	Grado
Cuauhtémoc	Algarin	5,556	0.90156	4	Alto
Cuauhtémoc	Ampliación Asturias	5,708	0.86853	3	Medio
Cuauhtémoc	Asturias	4,364	0.86687	3	Medio
Cuauhtémoc	Atlampa	14,433	0.86445	3	Medio
Cuauhtémoc	Buenavista	15,605	0.86896	3	Medio
Cuauhtémoc	Buenos Aires	5,772	0.79160	2	Bajo
Cuauhtémoc	Centro	61,229	0.75839	2	Bajo
Cuauhtémoc	Condesa	8,453	0.96018	4	Alto
Cuauhtémoc	Cuauhtémoc	11,399	0.95240	4	Alto
Cuauhtémoc	Doctores	44,703	0.85298	3	Medio
Cuauhtémoc	Esperanza	4,072	0.84225	3	Medio
Cuauhtémoc	Ex-Hipódromo de Peralvillo	11,711	0.84403	3	Medio
Cuauhtémoc	Felipe Pescador	1,988	0.88338	3	Medio
Cuauhtémoc	Guerrero	42,339	0.83420	3	Medio
Cuauhtémoc	Hipódromo	13,572	0.96310	4	Alto
Cuauhtémoc	Hipódromo Condesa	3,204	0.96037	4	Alto
Cuauhtémoc	Juárez	10,184	0.91452	4	Alto
Cuauhtémoc	Maza	2,503	0.82626	3	Medio
Cuauhtémoc	Morelos	36,590	0.74957	2	Bajo
Cuauhtémoc	Nonoalco Tlatelolco	27,843	0.96114	4	Alto
Cuauhtémoc	Obrera	35,224	0.83980	3	Medio
Cuauhtémoc	Paulino Navarro	5,307	0.82021	3	Medio
Cuauhtémoc	Peralvillo	20,213	0.84848	3	Medio
Cuauhtémoc	Roma Norte	27,770	0.91764	4	Alto
Cuauhtémoc	Roma Sur	17,435	0.93359	4	Alto
Cuauhtémoc	San Rafael	19,684	0.92592	4	Alto
Cuauhtémoc	San Simón Tolnáhuac	9,885	0.86051	3	Medio
Cuauhtémoc	Santa María Insurgentes	1,480	0.89881	3	Medio
Cuauhtémoc	Santa María La Ribera	40,960	0.90479	4	Alto
Cuauhtémoc	Tabacalera	3,267	0.89747	3	Medio
Cuauhtémoc	Tránsito	9,720	0.81415	3	Medio
Cuauhtémoc	Valle Gómez	6,281	0.78673	2	Bajo
Cuauhtémoc	Vista Alegre	3,377	0.88345	3	Medio

METHODOLOGY

To begin with, data was gathered as explained before using the foursquare API. This was turn into a data frame an into a data set- with the venues of interest and their coordinates.

```
[35]: results = requests.get(url).json()
      results
```


```
[35]: {'meta': {'code': 200, 'requestId': '5ef946cc513bb60f5114b60b'},
       'response': {'venues': [{'id': '4cc8ae0b41e75481e20e5884',
          'name': 'Parque Alameda',
          'location': {'address': 'Av. Juárez 76',
           'crossStreet': 'Esq. con Azueta',
           'lat': 19.434515120135995,
           'lng': -99.14715262455131,
           'labeledLatLngs': [{'label': 'display',
             'lat': 19.434515120135995,
             'lng': -99.14715262455131}],
           'distance': 80,
           'postalCode': '06010',
```


```
[39]: venues = results['response']['venues']

      # tranform venues into a dataframe
      dataframe = json_normalize(venues)
      dataframe.head(5)
```

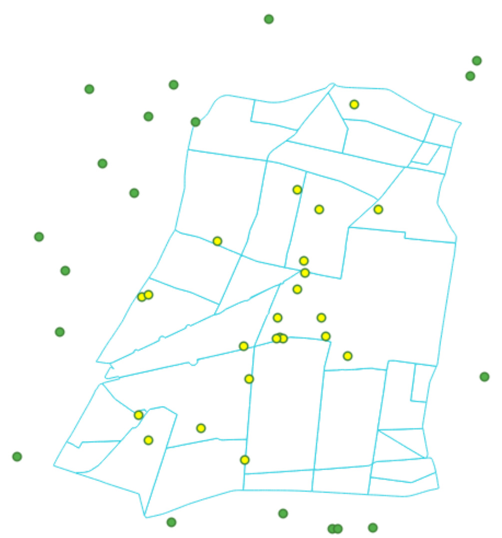
[39]:

	id	name	categories	referralId	hasPerk	location.address	location.crossStreet	location.la
0	4cc8ae0b41e75481e20e5884	Parque Alameda	[{'id': '4bf58dd8d48988d1fd941735', 'name': 'S...	v-1593395281	False	Av. Juárez 76	Esq. con Azueta	19.43451
1	4ec7df32f79041351d2d6868	Parque Ciudadela	[{'id': '4bf58dd8d48988d182941735', 'name': 'T...	v-1593395281	False	NaN	NaN	19.42732
2	53e8320c498ee437284d3f15	Parque Olof Palme	[{'id': '4bf58dd8d48988d163941735', 'name': 'P...	v-1593395281	False	Rio Grijalva	Rio Danubio	19.43332
3	4da21ca29aa4721e3473091a	Parque José Martí	[{'id': '4bf58dd8d48988d164941735', 'name': 'P...	v-1593395281	False	Reforma	Hidalgo	19.43710

After getting the json file from foursquare with the venues of interest, using QGIS, I started by identifying points, venues-parks, that were outside the Cuauhtemoc Area. This is done using the select by location tool The adjacent figure shows in yellow areas that fall within the polygon of interest while the green ones not.

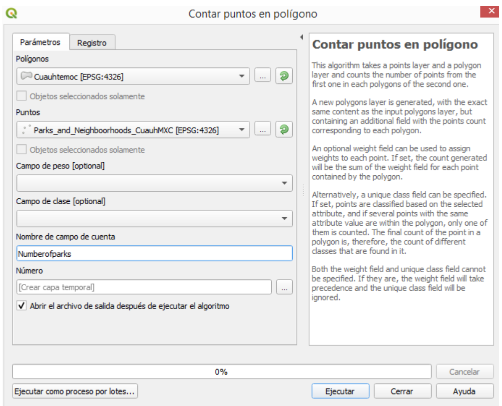
I also created a new shape file with only those points that intersect the Cuauhtemoc area or the yellow points. That is exporting the selected features into a new shape file.

Then, as shown below I also merged the parks layer with the neighborhood layers. This way, each park now has information on the neighborhood in which it is located.



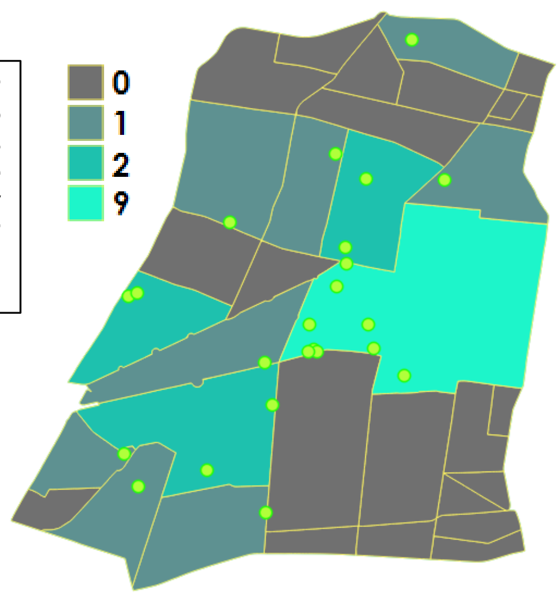
field_1	id	Parks Names	tegori	referralId	hasPerk	ocation.a	catior	location.l	location_1	cation	catior	catior	cation	cation	catior	cation	ocation.	uePa	catior	Neighborhood
40	51ccd113498e6...	Parque De Bols...	[]	v-1594178470	False			19.42709667999...	-99.1493420900...	[{'l...	931		MX			Mé...	['Méx...			CENTRO
8	4b54d09ff964a5...	Parque Espana	[{'i...	v-1594178470	False	Av. S...	Oa...	19.41541926000...	-99.1713066199...		3351		MX	Cu...	Dis...	Mé...	['Av. ...			CONDESA
6	53e8320c498ee...	Parque Olof Pa...	[{'i...	v-1594178470	False	Rio G...	Ri...	19.43332734999...	-99.1707556700...	[{'l...	2475		MX			Mé...	['Rio ...			CUAUHTEMOC
15	5488464e498ed...	parque para pe...	[{'i...	v-1594178470	False	Morelia	Co...	19.42097413000...	-99.1544609099...	[{'l...	1757		MX	Ro...	Dis...	Mé...	['Mor...		Ro...	ROMA NORTE
11	4da21ca29aa47...	Parque Jose Ma...	[{'i...	v-1594178470	False	Refor...	Hi...	19.43710537999...	-99.1459789799...	[{'l...	248		MX	Ci...	Dis...	Mé...	['Refo...			CENTRO
0	4cc8ae0b41e75...	Parque Alameda	[{'i...	v-1594178470	False	Av. J...	Es...	19.43451512000...	-99.1471526200...	[{'l...	80	6010	MX	Cu...	Dis...	Mé...	['Av. J...			CENTRO
2	4ec7df32f79041...	Parque Ciudadela	[{'i...	v-1594178470	False			19.42732162000...	-99.1497528900...	[{'l...	918		MX			Mé...	['Méx...			CENTRO
1	4b5626fef964a5...	Parque Mexico	[{'i...	v-1594178470	False	Av. ...	En...	19.41168309000...	-99.1697224600...		3525	6100	MX	Cu...	Dis...	Mé...	['Av. ...	93...		HIPODROMO
32	50282730e4b0fe...	Parque Recreat...	[{'i...	v-1594178470	False			19.43019484999...	-99.1434936499...	[{'l...	686		MX			Mé...	['Méx...			CENTRO
31	5675d79d498eb...	Parque Giordano	[{'i...	v-1594178470	False			19.42585272000...	-99.1552554200...	[{'l...	1339		MX			Mé...	['Méx...			JUAREZ
35	4ebf01607ee54e...	Parque de Los ...	[{'i...	v-1594178470	False			19.46262791999...	-99.1385539399...	[{'l...	3183		MX		CD...	Mé...	['CD...			PERALVILLO
33	5249ecb2498e6...	Parque De Los ...	[{'i...	v-1594178470	False			19.43017615999...	-99.1502069299...	[{'l...	641		MX			Mé...	['Méx...			CENTRO
19	4d037f229d33a...	Parque San Fer...	[{'i...	v-1594178470	False	Puen...	Sa...	19.43890030000...	-99.1460947100...	[{'l...	426		MX	Ci...	Dis...	Mé...	['Pue...			GUERRERO
18	5110018de4b00...	Parque Edith G...	[{'i...	v-1594178470	False	Av. Y...	Ca...	19.41349981000...	-99.1618252300...	[{'l...	2861	6700	MX	M...	Dis...	Mé...	['Av. ...			ROMA NORTE

I also count points in polygon tool to identify the number of points in each neighborhood. This adds a new column with the count for each neighborhood



	Name	FolderPath	Shape_Leng	Shape_Area	Delegacion	Numberofparks
1	CENTRO	Colonias/Colon...	0.09944792955	0.00043105979	Cuauhtemoc	9
2	ROMA NORTE	Colonias/Colon...	0.07086460649	0.00020287239	Cuauhtemoc	2
3	GUERRERO	Colonias/Colon...	0.04948405604	0.00013117851	Cuauhtemoc	2
4	CUAUHTEMOC	Colonias/Colon...	0.04889174521	0.00010557328	Cuauhtemoc	2
5	SANTA MARIA ...	Colonias/Colon...	0.05015670381	0.00015545499	Cuauhtemoc	1
6	PERALVILLO	Colonias/Colon...	0.03960491647	0.00007483128	Cuauhtemoc	1
7	ROMA SUR	Colonias/Colon...	0.04688972490	0.00012134003	Cuauhtemoc	1
8	HIPODROMO	Colonias/Colon...	0.04931849603	0.00009590027	Cuauhtemoc	1
9	CONDESA	Colonias/Colon...	0.03450069073	0.00005723402	Cuauhtemoc	1
10	BUENAVISTA	Colonias/Colon...	0.04291676886	0.00009335234	Cuauhtemoc	1
11	MORELOS	Colonias/Colon...	0.04739208614	0.00010296778	Cuauhtemoc	1
12	JUAREZ	Colonias/Colon...	0.06814846132	0.00012806204	Cuauhtemoc	1

By generating the new data set with the number of parks per neighborhood it is already possible to see visualize this variable. In the adjacent map it is possible to see the parks (the green dots) and the polygons according the number of parks they have. Centro has 9 parks, while some other neighborhoods have 2 or 1, while the majority has none.



Then, for each of the 23 parks I went back to the foursquare Api to find their rating.

```
[80]: venue_id = '4b5626fef964a520bd0228e3' # ID of Parque Mexico
      url = 'https://api.foursquare.com/v2/venues/{}?client_id={}&client_secret={}&v={}'.format(venue_id, CI
      url
```

```
[80]: 'https://api.foursquare.com/v2/venues/4b5626fef964a520bd0228e3?client_id=PN15LV3TD50YLNOKXA2Y1WF0HU1P
      4GJF4UKWDGJETN1J4AAE&amp;client_secret[redacted]v=2020071
      1'
```

```
[81]: results = requests.get(url).json()
      results
```


```
[81]: {'meta': {'code': 200, 'requestId': '5f0a47a33395b26d601e0aa2'},
       'response': {'venue': {'id': '4b5626fef964a520bd0228e3',
         'name': 'Parque México',
         'contact': {},
         'location': {'address': 'Av. México s/n',
          'crossStreet': 'Entre Michoacán y Amsterdam',
          'lat': 19.411683086011223,
          'lng': -99.16972246246875,
          'postalCode': '06100',
          'cc': 'MX',
          'city': 'Cuauhtemoc',
```

I created a new data frame CNPD which stand for Cuauhtemoc Neighborhood parks data, with the normalized number of parks and the normalized average rating of parks for each neighborhood.

And then I ran k-means clustering to identify neighborhoods clusters based on these two variables.

[126]:
```
CNPD.head(33)
```

[126]:

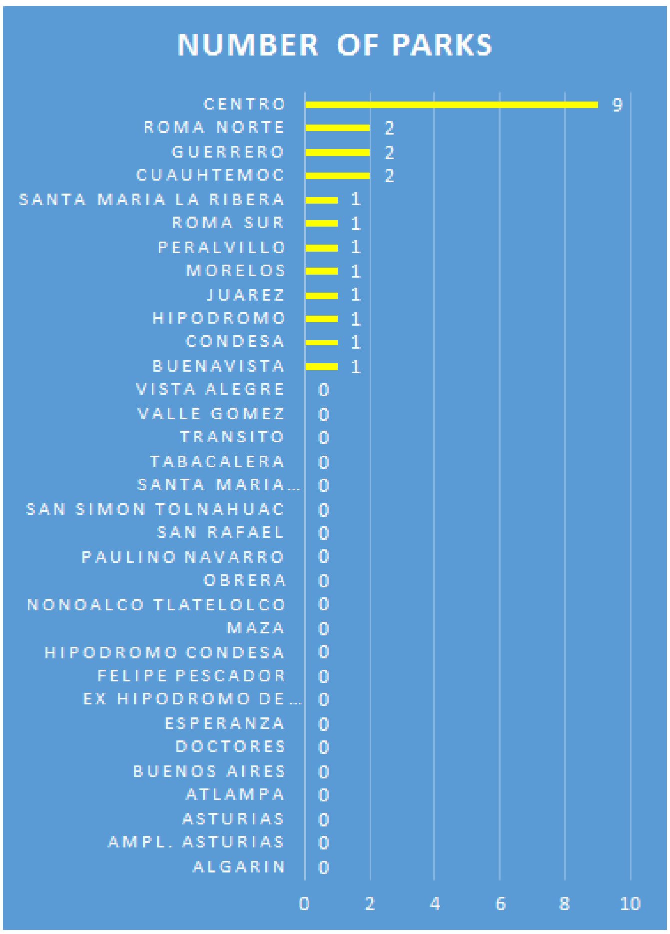
Neighborhood	Nparksnormalized	Arnormalized
CENTRO	1.00	0.86
ROMA NORTE	0.22	0.79
CUAUHTEMOC	0.22	0.72
GUERRERO	0.22	0.87
ROMA SUR	0.11	0.00
HIPODROMO	0.11	0.96
CONDESA	0.11	1.00
JUAREZ	0.11	0.00
MORELOS	0.11	0.56
BUENAVISTA	0.11	0.00
SANTA MARIA LA RIBERA	0.11	0.00
PERALVILLO	0.11	0.61
FELIPE PESCADOR	0.00	0.00
MAZA	0.00	0.00
NONOALCO TLATELOLCO	0.00	0.00
EX HIPODROMO DE PERALVILLO	0.00	0.00
VALLE GOMEZ	0.00	0.00
ATLAMPA	0.00	0.00
SAN SIMON TOLNAHUAC	0.00	0.00
SANTA MARIA INSURGENTES	0.00	0.00

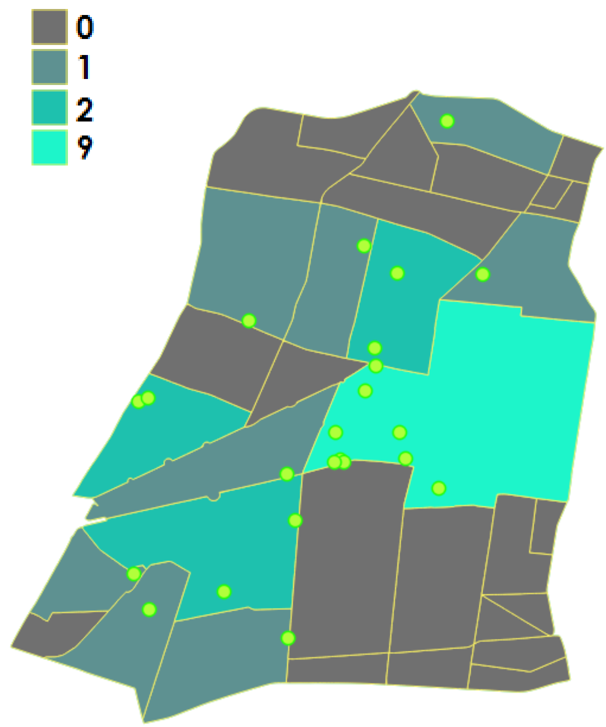
[125]:
```
# set number of clusters
kclusters = 4

kmeans = KMeans(n_clusters=kclusters, random_state=0).fit(CNPD)

# check cluster labels generated for each row in the dataframe
kmeans.labels_[0:33]
```

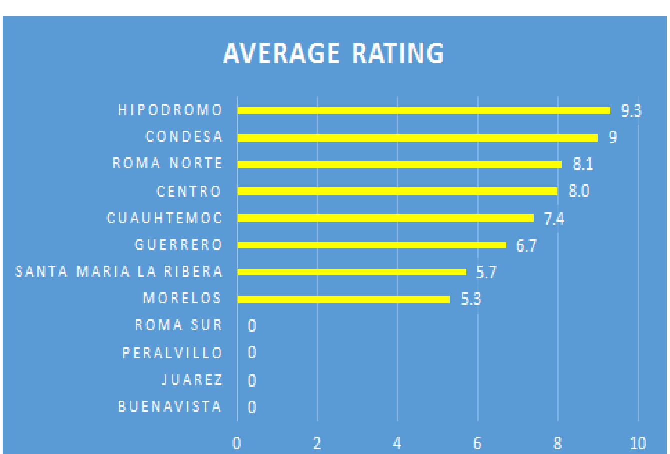
[125]:
```
array([0, 2, 3, 2, 1, 2, 2, 1, 3, 1, 1, 3, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
       1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1], dtype=int32)
```

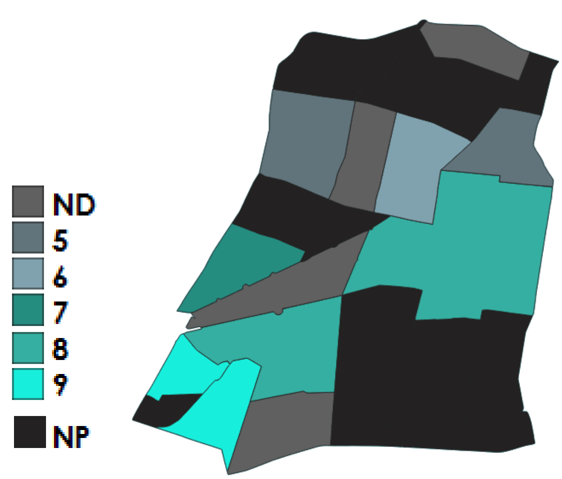
RESULTS

The project first allow to identify the number of parks in each neighborhood and how they were rated. It was possible to visualize neighborhoods according to the availability and perceived quality –rating of parks. Centro was the neighborhood with the largest number of parks: 9 . Then there were also three neighborhoods with two parks, and eight neighborhoods with one park. Also, it was found that there were 21 neighborhoods with no parks at all, most of which, are located in the south east and northern parts of the city. In addition, while centro has the highest number of parks, the highest average rating of parks are in Condesa and Hipodromo. These are among the richest neighborhoods in Cuauhtemoc and the city.





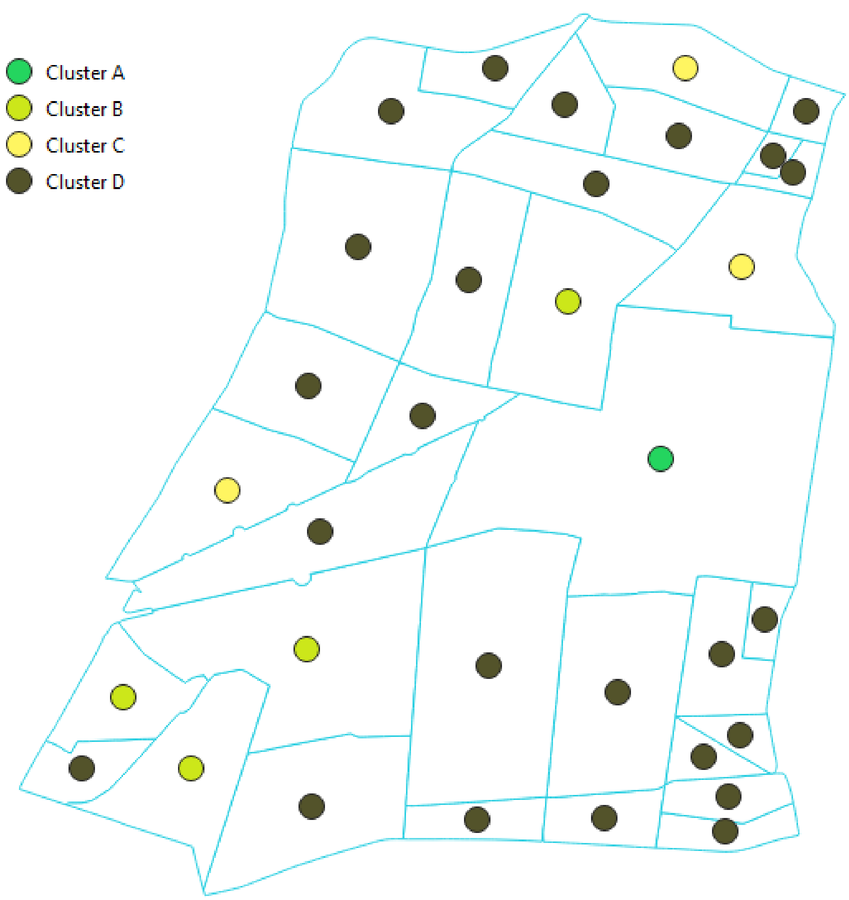


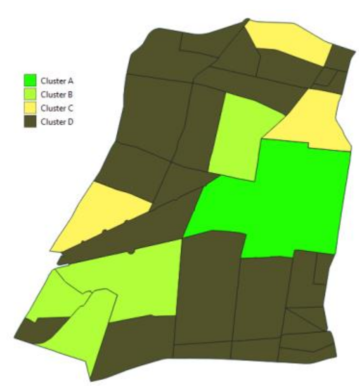


RESULTS

The project allow to cluster neighborhoods based on both availability and rating of parks. Centro stands alone in cluster A, as it houses nearly half of all parks in Cuauhtemoc, but not necessarily the best rated ones. In cluster B are found neighborhoods characterized by few parks but highly rated parks. Most are found in the western part of the city, in Condesa, Hipodromo and Roma Norte. Cluster C is characterized by few and badly rated parks while cluster D is comprised by neighborhoods by no parks at all.

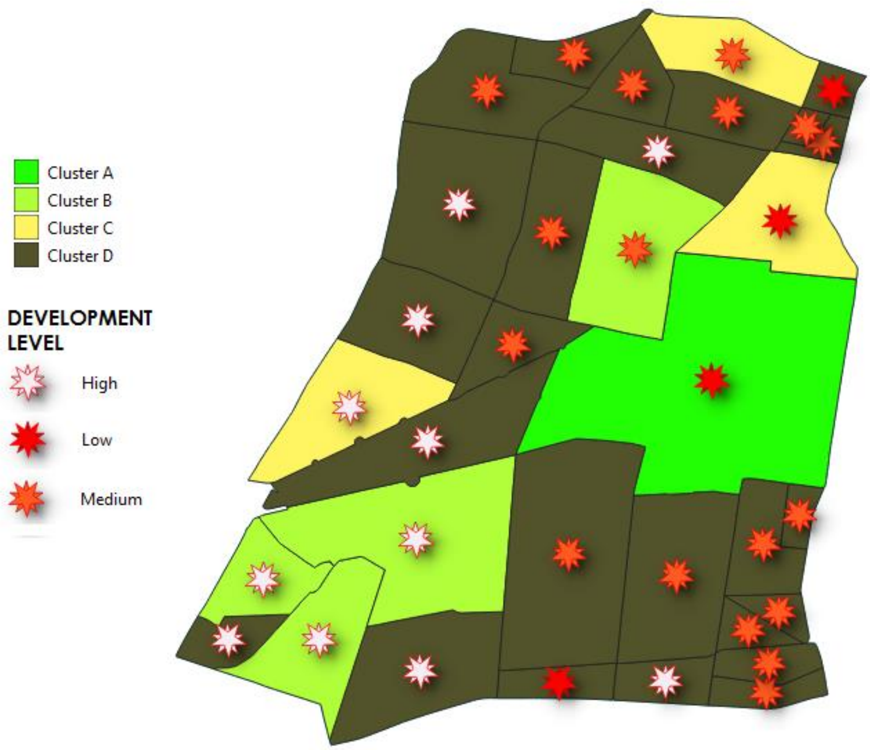
The map below shows each cluster by different color dots. It also shows in the right and corner with the same color ramp the whole neighborhood polygons according the cluster they belong to.





RESULTS, conclusions final remarks

The map below shows in an overlay neighborhood polygons according their cluster and the development level with dot stars. Although parks are found in areas with low development as Centro in cluster 1, and in areas with medium development as well, high development areas coincide with cluster B, which is characterized by few but good parks. Nonetheless, some high development areas are also found in cluster D which is characterized by having no parks. All we can way is that neighborhoods in cluster B, with few but highly rated parks, are mainly areas with high development, but not the other way around.



This research has contributed to the understanding on how neighborhoods in central Mexico City, the Cuauhtemoc borough, are similar to each other based on the availability and quality perception- rating of parks. It was that there is uneven provision of parks, with nearly two thirds of neighborhoods having no parks. The latter was one of the characteristics of one of the 4 clusters generated, cluster D. Cluster A, was mainly Centro, which was characterized by a large number of parks but not necessarily the highest rated ones. Cluster B on the other hand, was characterized by few but good – high rated parks. Cluster C was characterized by few and not so good parks according to how people rated them.

The are and there are not parks in low, medium and high development neighborhoods. However, cluster B, which again, was characterized by the best parks according to people, are mainly high development areas. Condesa, Roma Norte and Hipodromo, which are among the richest neighborhoods not only in Cuauhtemoc, but the city and even the country, are similar in terms of their availability and quality of parks, the have few but the ones people consider the best.

It is recommended that parks not only by be available for neighborhoods in cluster D, in particular in areas such as Buenos Aires, which has no parks and has low development but also to improve those in cluster A and B so that they might be as good as in cluster B.