

# Moving to a child friendly neighborhood in

Miguel Alegre Moreno

February 2020



---

## Introduction

### Moving from New York to Madrid

5 years ago, Irene and Carlos went to New York to develop their professional careers in marketing and finance respectively. Carlos, who is currently working for a FinTech startup, has been offered a position in Madrid as the company's Country Manager for Iberia (Spain and Portugal). As Irene's creative work is flexible and doesn't require her to physically attend her company's offices, they both have decided that Carlos should accept the offering and, therefore, move together to Madrid.

### Choosing a neighborhood

Irene and Carlos live in Chelsea, one of Manhattan's neighborhoods. They love living there because of the leisure venues that it offers them, and would like to find in Madrid a neighborhood with venues similar to the ones they enjoy in Chelsea.

On the other hand, Carlos and Irene are planning to have children in the near future, and they consider that a good neighborhood in which to raise a child should have a good amount of what they call child friendly spots: playgrounds, libraries, schools and hospitals.

Irene and Carlos are asking themselves which of Madrid's center neighborhoods will best fit their needs by offering similar venues to the ones they enjoy in Chelsea, and by presenting the greatest amount as possible of child friendly spots.

## Data

I gathered the following data to help Irene and Carlos find the neighborhoods that best fits their needs:

### Chelsea's data

I obtained New York's neighborhoods names and location data from the .json file saved in Cognitive Class's server that was used in week's 3 lab. After loading the relevant data into a new dataframe, I filtered it to make it contain only Chelsea's information.

	Borough	Neighborhood	Latitude	Longitude
0	MANHATTAN	CHELSEA	40.744035	-74.003116

Table1. Chelsea's geo dataframe

I then used Foursquare's API to obtain data regarding Chelseas venues; I limited the search to 100 results in a radius of 500 meters around Chelsea's center.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	CHELSEA	40.744035	-74.003116	Milk & Hops Chelsea	40.744751	-74.002595	Beer Bar
1	CHELSEA	40.744035	-74.003116	TAO Downtown	40.742545	-74.003837	Asian Restaurant
2	CHELSEA	40.744035	-74.003116	Bathtub Gin	40.743638	-74.003290	Speakeasy
3	CHELSEA	40.744035	-74.003116	Atlantic Theater Company (Linda Gross Theater)	40.743953	-74.001476	Theater

Table 2. Example of Chelsea's venues dataframe

### Madrid's center data

I obtained Madrid's center boroughs and neighborhoods names by merging and filtering data present in a pair of .csv files found in <https://datos.madrid.es>.

Once I had the names of the 31 neighborhoods in Madrid's center, I used GeoPy's library to obtain their location data.

	Borough	Neighborhood	Latitude	Longitude
0	CENTRO	PALACIO	40.415129	-3.715618
1	CENTRO	EMBAJADORES	40.409681	-3.701644
2	CENTRO	CORTES	40.414348	-3.698525
3	CENTRO	JUSTICIA	40.423957	-3.695747

Table 3. Example of Madrid's neighborhoods geo dataframe

In the same way I did with Chelsea, I obtained Madrid's venues data using Foursquare's API was used to retrieve data regarding Madrid's center neighborhoods venues.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	PALACIO	40.415129	-3.715618	Santa Iglesia Catedral de Santa María la Real ...	40.415767	-3.714516	Church
1	PALACIO	40.415129	-3.715618	Plaza de La Almudena	40.416320	-3.713777	Plaza
2	PALACIO	40.415129	-3.715618	Cervecería La Mayor	40.415218	-3.712194	Beer Bar
3	PALACIO	40.415129	-3.715618	Corral de la Morería	40.412619	-3.714249	Performing Arts Venue

Table 4. Example of Madrid's venues dataframe

I obtained the last piece of data, the data regarding Madrid's child friendly spots, in .csv files found in <https://datos.madrid.es>. After filtering the dataframes containing the information of each type of spot, I merged them all together into one single dataframe.

	Playgrounds	Schools	Libraries	Hospitals
Neighborhood				
GUINDALERA	22	15	2	4
LEGAZPI	17	6	0	1
ACACIAS	16	17	1	2
DELICIAS	13	13	1	1

Table 5. Example of Madrid's child friendly spots dataframe

## Methodology

I did the following analysis of the data obtained previously to find out which are the most suited neighborhoods for Irene and Carlos:

### Neighborhood visualization

First of all I used Folium to visualize the location of the different neighborhoods.

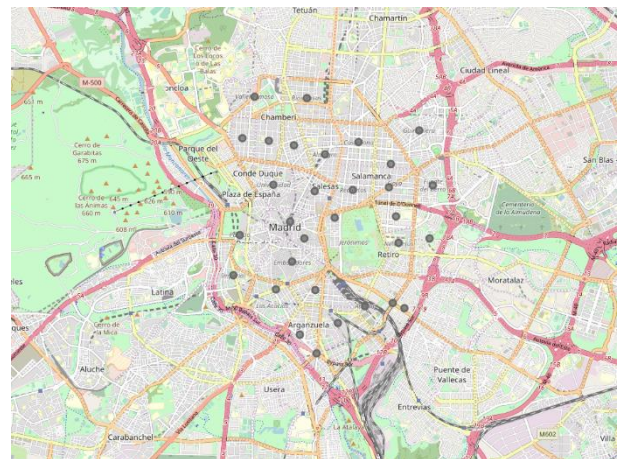
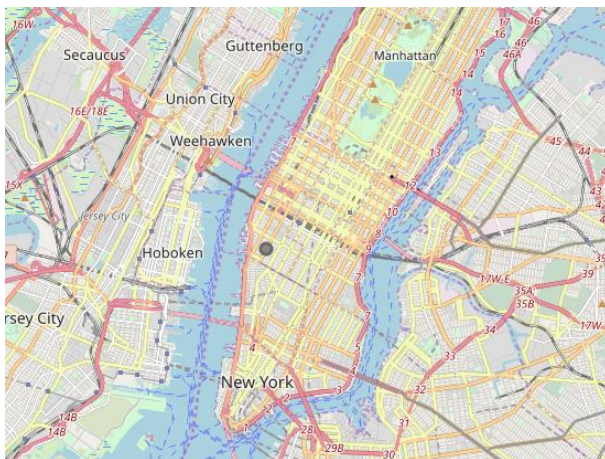


Figure1. Chelsea & Madrid's center neighborhoods

### Neighborhood clustering by venues

I one hot encoded Chelsea's and Madrid's neighborhoods venues category data. Once the data was one hot encoded, I calculated the frequency of occurrence of each category in each neighborhood. The resulting dataframes were then merged together and fed to the k-means algorithm that I used to cluster the neighborhoods.

	Neighborhood	American Restaurant	Art Gallery	Asian Restaurant	Bakery	Bar	Beer Bar	Bookstore	Breakfast Spot	Burger Joint	Butcher	Café	Cheese Shop	Chinese Restaurant	Clothing Store	Cocktail Bar	Coffee Shop
0	CHELSEA	0.030000	0.020000	0.010	0.050000	0.010000	0.010000	0.020000	0.02	0.020000	0.01	0.010000	0.010000	0.01	0.01	0.020000	0.060000
1	ACACIAS	0.000000	0.068182	0.000	0.043455	0.113636	0.000000	0.022727	0.00	0.022727	0.00	0.022727	0.000000	0.00	0.00	0.000000	0.000000
2	ADEFAS	0.000000	0.000000	0.025	0.050000	0.075000	0.000000	0.025000	0.00	0.000000	0.00	0.050000	0.000000	0.00	0.00	0.000000	0.025000
3	ALMAGRO	0.010000	0.020000	0.020	0.010000	0.040000	0.010000	0.020000	0.00	0.000000	0.00	0.010000	0.000000	0.00	0.00	0.020000	0.010000

Table 6. Example of frequency of occurrence dataframe

## Neighborhood clusters visualization

After applying k-means, I used Folium to visualize the resulting neighborhood clusters.

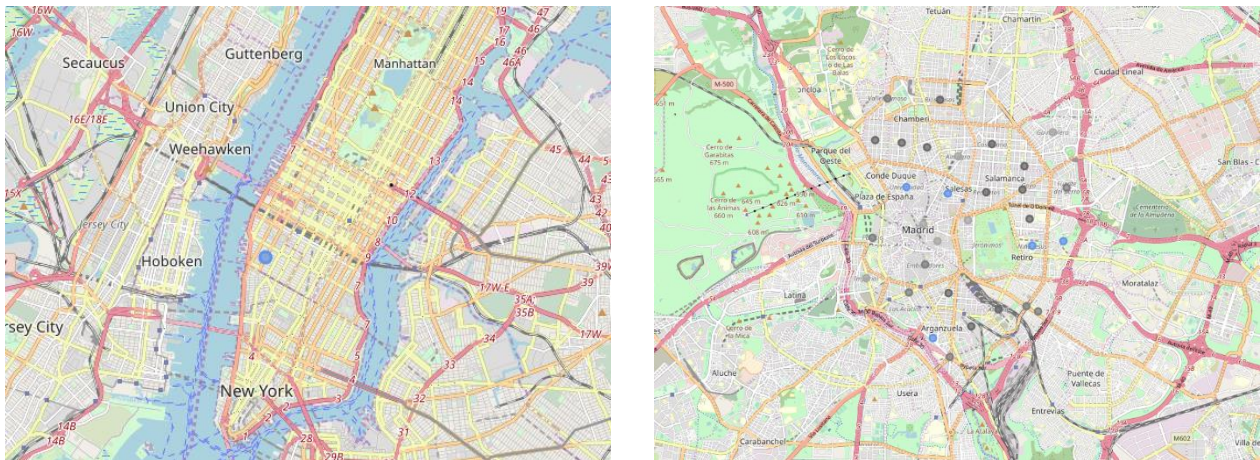


Figure 2. Chelsea & Madrid's center neighborhoods clusters

## Chelsea's cluster analysis

In order to understand the venue categories that Chelsea had in common with the rest of neighborhoods in its cluster, I created a dataframe in which I displayed the 10 most common venues for each neighborhood.

Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0 CHELSEA	Coffee Shop	Bakery	Ice Cream Shop	Italian Restaurant	Hotel	Nightclub	Wine Shop	American Restaurant	Theater	Breakfast Spot
4 JUSTICIA	Restaurant	Bakery	Spanish Restaurant	Italian Restaurant	Vegetarian / Vegan Restaurant	Hotel	Gay Bar	American Restaurant	Flower Shop	Deli / Bodega
5 UNIVERSIDAD	Bar	Bookstore	Cocktail Bar	Tapas Restaurant	Theater	Café	Restaurant	Argentinian Restaurant	Chinese Restaurant	Spanish Restaurant
9 CHOPERA	Italian Restaurant	Park	Art Gallery	Burger Joint	Tapas Restaurant	Coffee Shop	Spanish Restaurant	Bar	Restaurant	Mobile Phone Shop
16 ESTRELLA	Coffee Shop	Park	Cocktail Bar	Italian Restaurant	Sports Club	Spanish Restaurant	Café	Chinese Restaurant	Restaurant	Pool
19 NIÑO JESUS	Spanish Restaurant	Italian Restaurant	Park	Bar	Supermarket	Gym	Café	Cocktail Bar	Coffee Shop	Plaza

Table 7. Chealsea's cluster 10 most common venues dataframe

## Child friendly spots distribution analysis

I first used a box plot to determine if child friendly spots are evenly distributed within Madrid's neighborhoods.

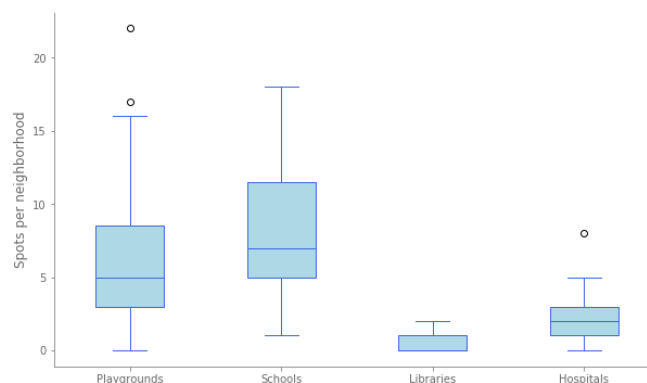


Figure 3. Box plot of Madrid's child friendly spots distribution

After observing that child friendly spots are not evenly distributed (some neighborhoods have a big number of playgrounds and / or schools while other have very little), I used a bar graph to visualize the number of child friendly spots in each neighborhood.

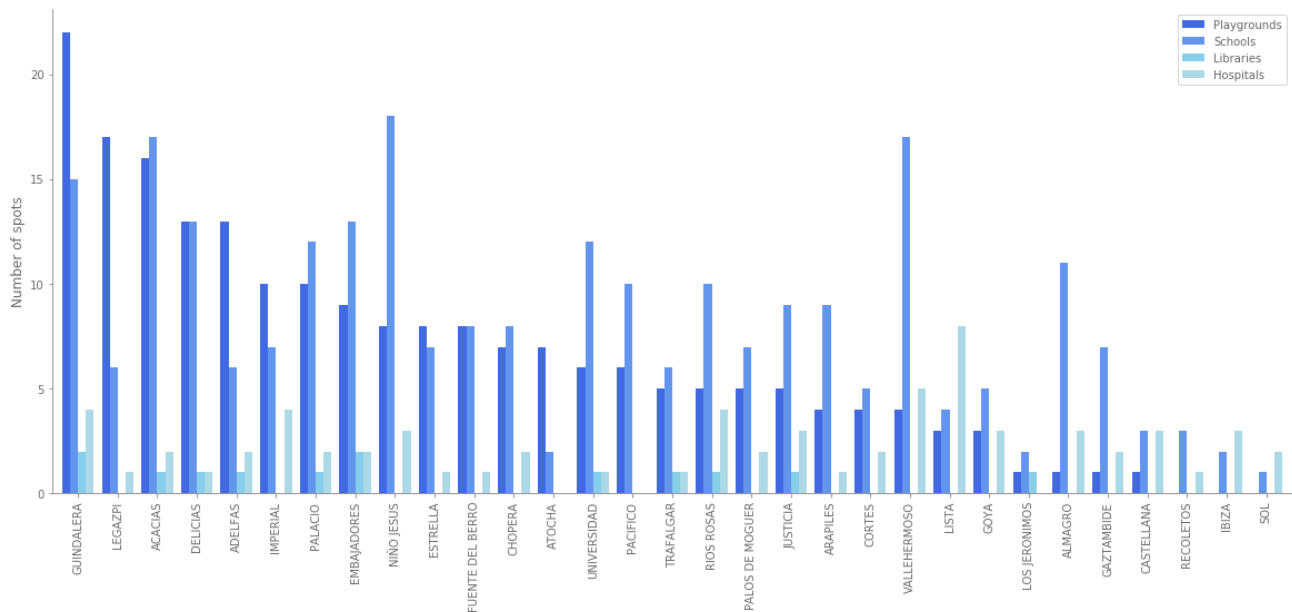


Figure 4. Example of Madrid's child friendly spots data

## Results

Chelsea is a neighborhood with a considerable amount of coffee shops, bakeries, Italian and American restaurants, parks and hotels. Within Madrid's center, the following neighborhoods offer similar type of venues: **Justicia**, **Universidad**, **Chopera**, **Estrella** and **Niño Jesús**.

Besides Niño Jesús in relation to its number of schools, this 5 neighborhoods all have a similar amount of child friendly spots.

	Playgrounds	Schools	Libraries	Hospitals
Neighborhood				
NIÑO JESUS	8	18	0	3
ESTRELLA	8	7	0	1
CHOPERA	7	8	0	2
UNIVERSIDAD	6	12	1	1
JUSTICIA	5	9	1	3

Table 8. Child friendly spots in Madrid's neighborhoods belonging to Chelsea's cluster



## Discussion

Something I noticed during the analysis of the data collected is that while for some neighborhoods I was able to retrieve 100 results from Foursquare, for other neighborhoods (a 40% of the total) I was able to retrieve less than 50 results.

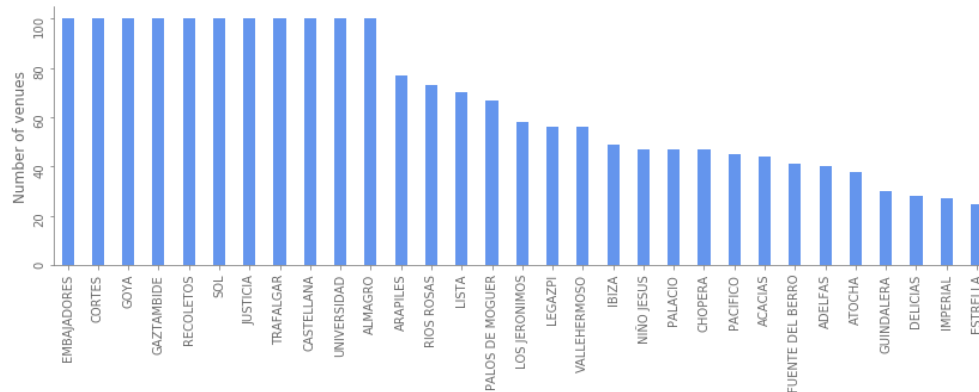


Figure 5. Example of Madrid's child friendly spots data

Another issue I identified regarding Foursquare's data is that there are some venue categories that strongly resemble one another, like Café and Coffee Shop, or like Gym and Gym / Fitness Center. As similar venues may have been divided into different categories, this issue may have had an impact on the results of the analysis by distorting the frequency of occurrence of each category with which the k-means algorithm is fed to cluster the neighborhoods.

In a future study, apart from trying to retrieve more venues from the neighborhoods with less Foursquare results, either by using different locations within the neighborhood to perform the search queries or by using a bigger radius, similar venue categories could be grouped together to avoid the mentioned frequency of occurrence distortion.

## Conclusion

Besides the limitation discussed in the previous section, and taking into account the result of the data analysis performed, any of the neighborhoods belonging to Chelsea's cluster (**Justicia**, **Universidad**, **Chopera**, **Estrella** and **Niño Jesús**) would be an excellent choice for Irene and Carlos as they all have a similar number of child friendly spots.

By the way, my personal favorite within the selected neighborhoods is **Justicia**.