The Battle of the Neighborhoods

A. Introduction

A.1 Description & Discussion of the Background

Buenos Aires City is the most important city in Argentina. It has a 203 km² of surface with 3 million of population approximately. As a resident of this city, I decide to use this city for the final project of the Applied Data Science specialization on Coursera platform. The city is divides into 15 communes with a 48 neighborhood [1]

As you can see, Buenos Aires is a big city with high population. Every day, people go outside to work, to eat or why not to see friends. For that a family want to live and move in a safest neighborhood. In consequence they must research where is the most convenient to live due the crimes that is happens in different commune and they want to buy things in venues that they like

Because when we saw the news on TV or the website, every time we see that a crime has happened. The project aims to select the safest commune in Buenos Aires City based on the total crimes considering all the three years and explore the 10 most common venues for each commune and finally cluster the communes using k-means

It will be interest for the Argentinian people because its important to have noticed on what city is the safest and the dangerous

A.2

To consider the problem we can list the data below:

- I found three datasets of the last three years on the government website. The dataset has the date of the crime, the crime, the commune, the neighborhood, the latitude and longitude.
  I cleaned all the samples that do not have communes and dropped all the columns that I considered unnecessary. Also I created a new column with the year's crime and dropped all the rows that have one registered quantity because they were few. Other things were to capitalize the columns, remove the parenthesis on the column 'Tipo_delito' and rename and convert into the right type of data [2]

  After, I concatenated all the three datasets in one to start with the data exploratory. I dropped the date because it was not being considered

- I imported a geojson file for the government website called **CABA_comunes.geojson** which has the neighborhood grouped in commune where I used a choropleth map to create a map of the city and show the safest and dangerous neighborhood [3]

- And I finally I imported a. geojson for the government website called **barrios.geojson** to obtain the average of latitude and longitude for each commune and if I saw that some point in the map are so far form the center, I got the latitude and longitude on Google Maps to correct manually with an arbitrary longitude and latitude that represents the center of it. I must modify the data for get the most venues in the FOURSQUARE API and obtain the clusters [4]

- I used **Forsquare API** to get the most common venues of given neighborhood of Buenos Aires City [5]

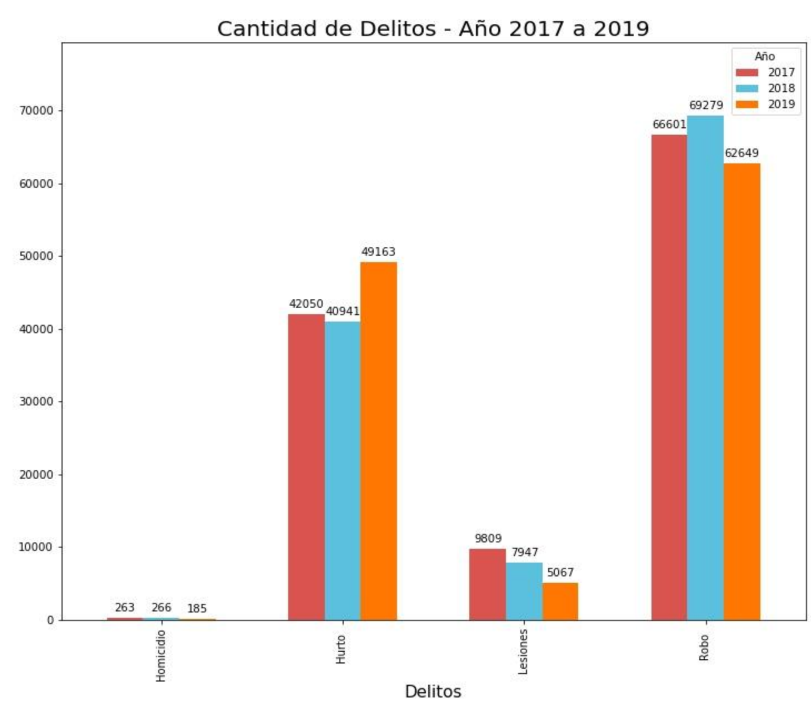
B. Methodology

After I cleaned all the three dataset and join them in one dataset and droppped the unnecessary columns. I obtained the final dataset to start the exploratory data analysis. This data frame has 354.220 samples and 6 features

The final features of the dataset are *Tipo_delito*, *Comuna, Barrio, Latitude, Longitude* and *Año.* Here is the head of the dataset:

	Tipo_delito	Comuna	Barrio	Latitude	Longitud	Año
0	Lesiones	4.0	Nueva Pompeya	-34.648387	-58.404748	2019
1	Robo	9.0	Liniers	-34.649827	-58.513859	2019
2	Lesiones	15.0	Chacarita	-34.588108	-58.439392	2019
3	Hurto	10.0	Floresta	-34.631877	-58.483975	2019
4	Robo	4.0	Parque Patricios	-34.633161	-58.397123	2019

To have a better understanding of the how much crimes had every year, I created a bar plot for each crime using **matplotlib** library:

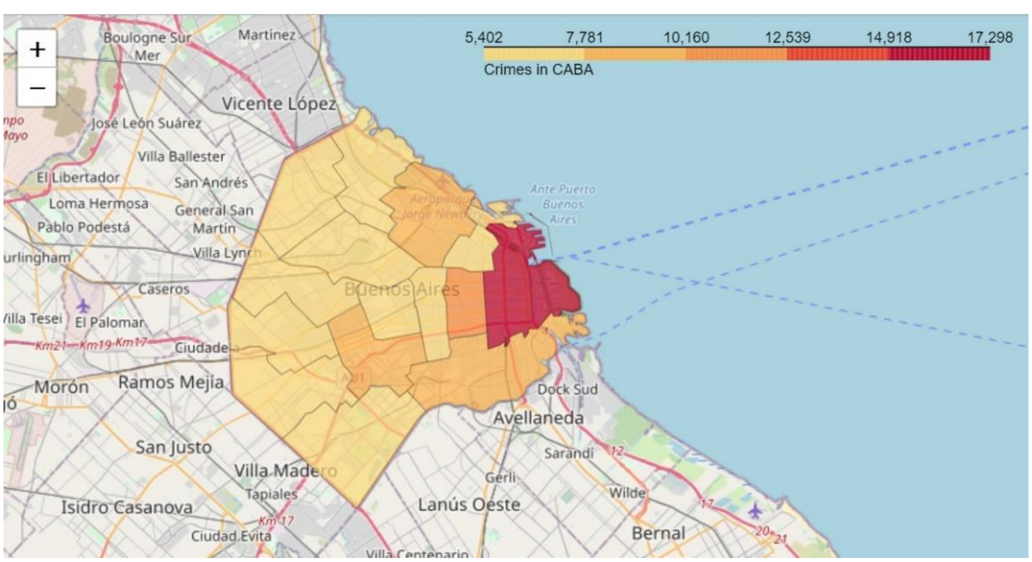


On the bar plot, we observe that ‘Robo’ (robbery) and ‘Hurto’ (pickpocketing) has the most data in comparison of the others crimes

I used python **folium** library to visualize communes of the city and therefore I created a table of the average amount of the last three years for each commune to display which commune has the highest and lowest of registered crimes. I used the **CABA_comuna.json** to merged and got which neighborhood are integrated on the communes

	Comuna	Barrios	Cantidad
0	1.0	CONSTITUCION - MONSERRAT - PUERTO MADERO - RE...	17297
1	3.0	BALVANERA - SAN CRISTOBAL	10268
2	4.0	BARRACAS - BOCA - NUEVA POMPEYA - PARQUE PATRI...	10030
3	14.0	PALERMO	9549
4	7.0	FLORES - PARQUE CHACABUCO	8396
5	13.0	BELGRANO - COLEGIALES - NUÑEZ	7293
6	15.0	AGRONOMIA - CHACARITA - PARQUE CHAS - PATERN...	7103
7	5.0	ALMAGRO - BOEDO	6940
8	9.0	LINIERS - MATADEROS - PARQUE AVELLANEDA	6729
9	12.0	COGHLAN - SAAVEDRA - VILLA PUEYRREDON - VILLA ...	6174
10	8.0	VILLA LUGANO - VILLA RIACHUELO - VILLA SOLDATI	5805
11	11.0	VILLA DEL PARQUE - VILLA DEVOTO - VILLA GRAL....	5798
12	10.0	FLORESTA - MONTE CASTRO - VELEZ SARSFIELD - VE...	5741
13	2.0	RECOLETA	5550
14	6.0	CABALLITO	5402

Taking this table, I created a choropleth map



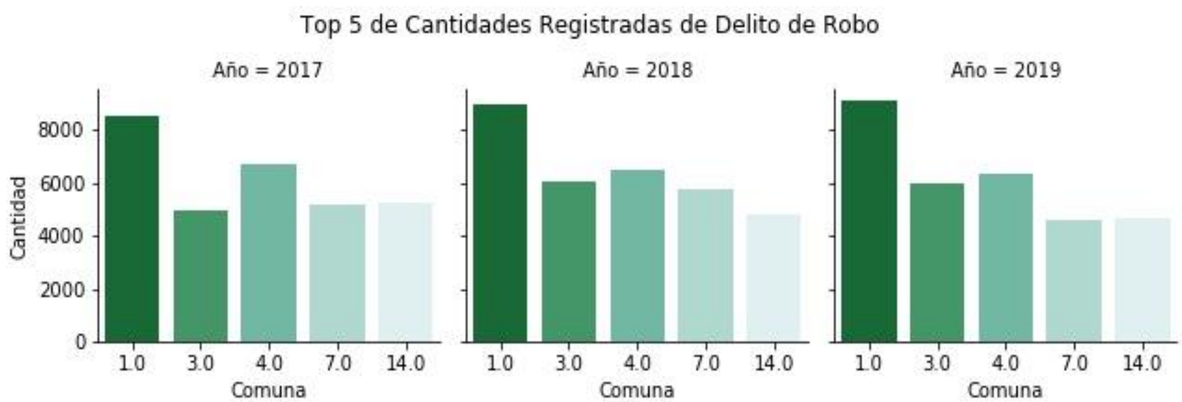
We can observe that the three highest with registered crimes. The communes are *1, 3* and *4'*. The lowest observed are *10, 2* and *6*

We get a first view of which commune are the safest and the dangerous. But this information in not enough to make a decision because is not the same if someone is murdered or is got robbery. Therefore, I took the top 5 of communes that has the highest amount recorded for each crime

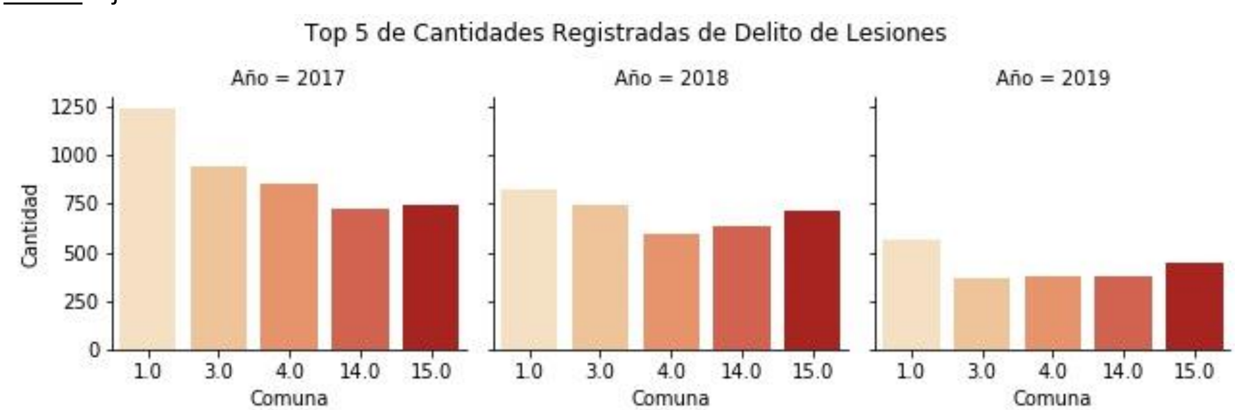
In consequence, I created four bar plots for each crime in every year using **seaborn** library

<u>Crime:</u> Homicide

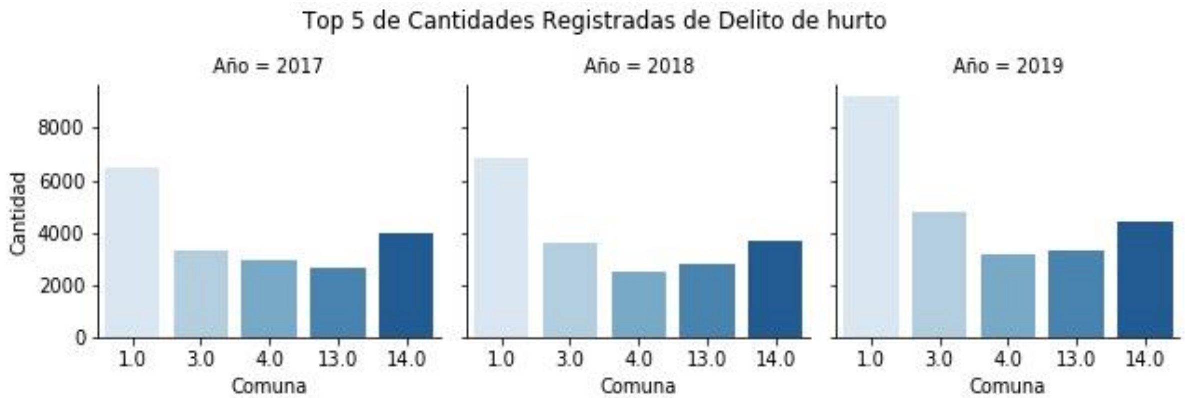
<u>Crime:</u> Robbery



<u>Crime:</u> Injuries

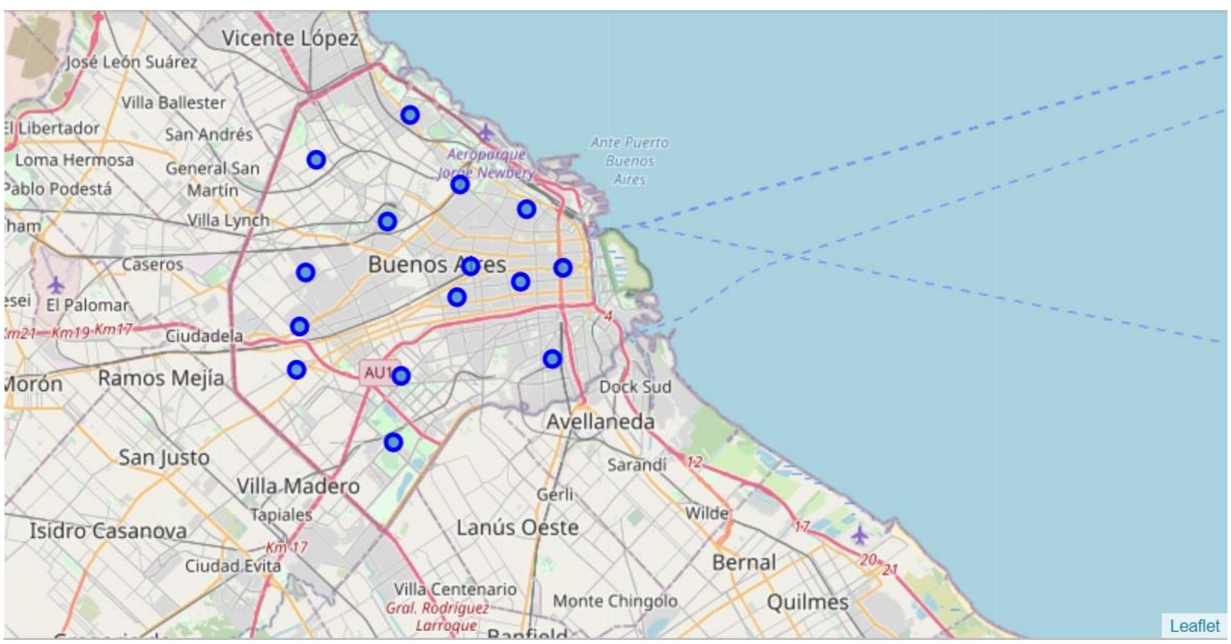


<u>Crime:</u> Pickpocketing



We can observe that taking all the three communes that we got previously, we get the same communes as the dangerous places to live

Now that we have a view of the security of the communes, I calculated the average from the latitude and longitude for each neighborhood using **barrios.geojson** to obtain the centers of the communes. In some commune to get the center, I searched on **Google Maps** and got arbitrarily the geolocation. Here is the map with their center:

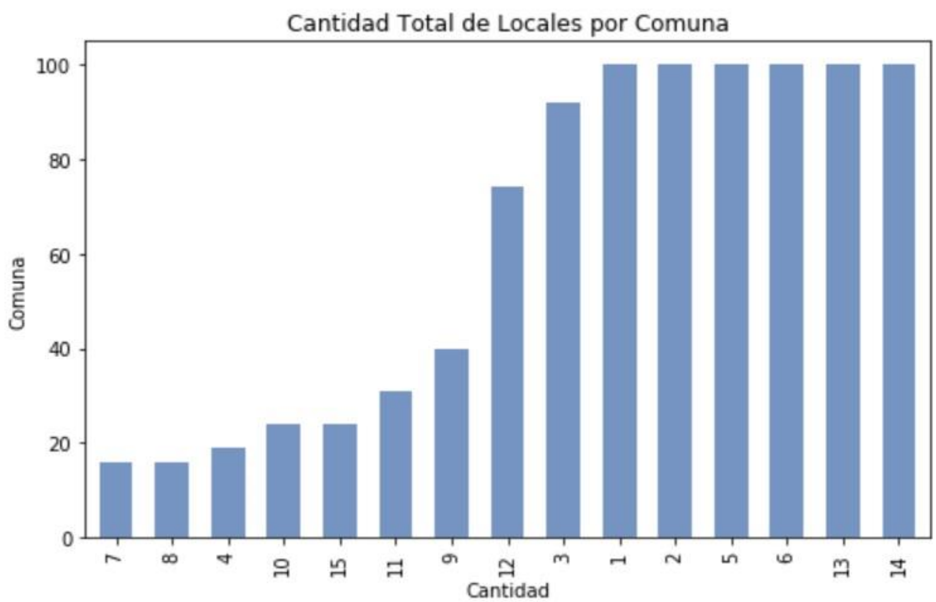


I utilized the FOURSQUARE API to explore and segment them. I designed the limit of 100 venue and the radius 1000 meter for each commune from the latitude and longitude obtained. Here is the head of the venues and geolocation from FOURSQUARE API with his commune respectively:

	Comuna	Latitud	Longitude	Lugar	Latitud del Lugar	Longitud del Lugar	Categoria del Lugar
0	1	-34.6087	-58.3802	de Dios Editores	-34.609999	-58.379109	Bookstore
1	1	-34.6087	-58.3802	Foto Club Buenos Aires	-34.609475	-58.378273	Camera Store
2	1	-34.6087	-58.3802	Starbucks	-34.609013	-58.380196	Coffee Shop
3	1	-34.6087	-58.3802	Hostel Portal del Sur BA	-34.609428	-58.378450	Bed & Breakfast
4	1	-34.6087	-58.3802	Latino Sándwich	-34.610413	-58.379110	Sandwich Place

We can see that the following communes: 1, 2, 5, 6, 13 and 14 has reached the 100 limits of venues. On the other hand: the 4,7,8 communes are below 20 venues on in our given coordinates with Latitude and Longitude, in below graph.

The result does not mean that inquiry run all the possible results in commune. It depends on the pair geolocation that we used in this project.



In summary of this graph **171** unique categories were returned by Foursquare, then I created a table which shows list of top 10 venues category for each commune in below table

	Comuna	1st Lugares Más Comunes	2nd Lugares Más Comunes	3rd Lugares Más Comunes	4th Lugares Más Comunes	5th Lugares Más Comunes	6th Lugares Más Comunes	7th Lugares Más Comunes	8th Lugares Más Comunes	9th Lugares Más Comunes	10th Lugares Más Comunes
0	1	Hotel	Coffee Shop	Theater	Café	Spanish Restaurant	Restaurant	Hostel	Nightclub	Argentinian Restaurant	Gym / Fitness Center
1	2	Hotel	Argentinian Restaurant	Ice Cream Shop	Italian Restaurant	Plaza	Bakery	Coffee Shop	Tea Room	Deli / Bodega	Café
2	3	Café	Pizza Place	Japanese Restaurant	Argentinian Restaurant	Spanish Restaurant	Bakery	Coffee Shop	Gym	Hotel	Bar
3	4	Argentinian Restaurant	Café	Pedestrian Plaza	Train Station	Flower Shop	Clothing Store	General Entertainment	Soccer Field	Grocery Store	Restaurant
4	5	Argentinian Restaurant	Bar	Café	Ice Cream Shop	Gym	Restaurant	Coffee Shop	Pizza Place	Hotel	Indie Theater

We have some common venue categories in commune. In this reason I used unsupervised learning **K-means algorithm** to cluster the boroughs. K-Means algorithm is one of the most common cluster method of unsupervised learning

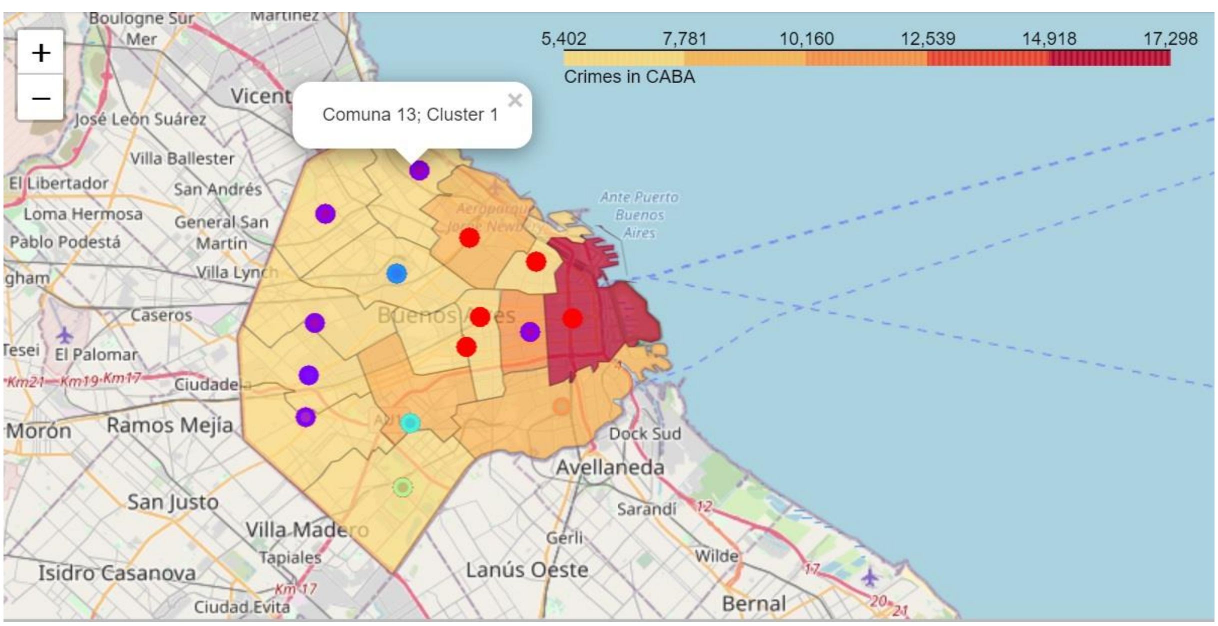
I clustered the communes into 6 clusters. In further project, I will investigate which k (number of clusters) is the optimum. Because it was not part of the course, I took it arbitrary

Here is my merged table with cluster labels for each borough

	Comuna	Latitud	Longitud	Cluster Labels	1st Lugares Más Comunes	2nd Lugares Más Comunes	3rd Lugares Más Comunes	4th Lugares Más Comunes	5th Lugares Más Comunes	6th Lugares Más Comunes	7th Lugares Más Comunes	8 Lugar M Comun
0	1	-34.608700	-58.380200	0	Hotel	Coffee Shop	Theater	Café	Spanish Restaurant	Restaurant	Hostel	Nightcl
1	2	-34.587700	-58.396700	0	Hotel	Argentinian Restaurant	Ice Cream Shop	Italian Restaurant	Plaza	Bakery	Coffee Shop	Tea Roo
2	3	-34.613767	-58.399165	0	Café	Pizza Place	Japanese Restaurant	Argentinian Restaurant	Spanish Restaurant	Bakery	Coffee Shop	Gy
3	4	-34.642343	-58.384727	4	Argentinian Restaurant	Café	Pedestrian Plaza	Train Station	Flower Shop	Clothing Store	General Entertainment	Socc Fie
4	5	-34.608500	-58.421900	0	Argentinian Restaurant	Bar	Café	Ice Cream Shop	Gym	Restaurant	Coffee Shop	Piz Pla

C. Results

I created a choropleth map to show how to integrate the final table and the map with most crimes registered. Here is the result:



We can observe the commune that has the highest and lowest crimes and which communes have common venues

Cluster “0”:

	Comuna	1st Lugares Más Comunes	2nd Lugares Más Comunes	3rd Lugares Más Comunes	4th Lugares Más Comunes	5th Lugares Más Comunes	6th Lugares Más Comunes	7th Lugares Más Comunes	8th Lugares Más Comunes	9th Lugares Más Comunes	10th Lugares Más Comunes
0	1	Hotel	Coffee Shop	Theater	Café	Spanish Restaurant	Restaurant	Hostel	Nightclub	Argentinian Restaurant	Gym / Fitness Center
1	2	Hotel	Argentinian Restaurant	Ice Cream Shop	Italian Restaurant	Plaza	Bakery	Coffee Shop	Tea Room	Deli / Bodega	Café
4	5	Argentinian Restaurant	Bar	Café	Ice Cream Shop	Gym	Restaurant	Coffee Shop	Pizza Place	Hotel	Indie Theater
5	6	Ice Cream Shop	Bakery	Coffee Shop	Café	Pizza Place	Burger Joint	Argentinian Restaurant	Bar	Gym	Pharmacy
13	14	Argentinian Restaurant	Hotel	Ice Cream Shop	Pizza Place	Gym / Fitness Center	Coffee Shop	Deli / Bodega	Bakery	Sushi Restaurant	Wine Shop

Cluster “1”:

	Comuna	1st Lugares Más Comunes	2nd Lugares Más Comunes	3rd Lugares Más Comunes	4th Lugares Más Comunes	5th Lugares Más Comunes	6th Lugares Más Comunes	7th Lugares Más Comunes	8th Lugares Más Comunes	9th Lugares Más Comunes	10th Lugares Más Comunes
2	3	Café	Pizza Place	Japanese Restaurant	Argentinian Restaurant	Spanish Restaurant	Bakery	Coffee Shop	Gym	Hotel	Bar
8	9	Pizza Place	Argentinian Restaurant	Café	Dessert Shop	Coffee Shop	Bus Stop	Bakery	Sports Club	Ice Cream Shop	Deli / Bodega
9	10	Argentinian Restaurant	Pizza Place	BBQ Joint	Bakery	Gas Station	Bar	Event Space	Sports Club	Steakhouse	Supermarket
10	11	Plaza	Café	Ice Cream Shop	Breakfast Spot	Pharmacy	Bakery	Pizza Place	Coffee Shop	Supermarket	Food
11	12	Pizza Place	Ice Cream Shop	Bakery	Café	Argentinian Restaurant	Italian Restaurant	Coffee Shop	BBQ Joint	Bus Stop	Grocery Store
12	13	BBQ Joint	Café	Coffee Shop	Pizza Place	Italian Restaurant	Sushi Restaurant	Restaurant	Ice Cream Shop	Sandwich Place	Grocery Store

Cluster “2”:

	Comuna	1st Lugares Más Comunes	2nd Lugares Más Comunes	3rd Lugares Más Comunes	4th Lugares Más Comunes	5th Lugares Más Comunes	6th Lugares Más Comunes	7th Lugares Más Comunes	8th Lugares Más Comunes	9th Lugares Más Comunes	10th Lugares Más Comunes
14	15	Pizza Place	Coffee Shop	Bar	Theater	Museum	Sandwich Place	Electronics Store	Spanish Restaurant	Sporting Goods Shop	Office

Cluster "3":

	Comuna	1st Lugares Más Comunes	2nd Lugares Más Comunes	3rd Lugares Más Comunes	4th Lugares Más Comunes	5th Lugares Más Comunes	6th Lugares Más Comunes	7th Lugares Más Comunes	8th Lugares Más Comunes	9th Lugares Más Comunes	10th Lugares Más Comunes
6	7	Plaza	Athletics & Sports	Pizza Place	Light Rail Station	Hockey Field	Soccer Field	Convenience Store	Moving Target	Restaurant	Korean Restaurant

Cluster "4":

	Comuna	1st Lugares Más Comunes	2nd Lugares Más Comunes	3rd Lugares Más Comunes	4th Lugares Más Comunes	5th Lugares Más Comunes	6th Lugares Más Comunes	7th Lugares Más Comunes	8th Lugares Más Comunes	9th Lugares Más Comunes	10th Lugares Más Comunes
7	8	Fast Food Restaurant	Electronics Store	Ice Cream Shop	Café	Hardware Store	Sporting Goods Shop	Coffee Shop	Mobile Phone Shop	Shopping Mall	Light Rail Station

Cluster "5":

	Comuna	1st Lugares Más Comunes	2nd Lugares Más Comunes	3rd Lugares Más Comunes	4th Lugares Más Comunes	5th Lugares Más Comunes	6th Lugares Más Comunes	7th Lugares Más Comunes	8th Lugares Más Comunes	9th Lugares Más Comunes	10th Lugares Más Comunes
3	4	Argentinian Restaurant	Café	Pedestrian Plaza	Train Station	Flower Shop	Clothing Store	General Entertainment	Soccer Field	Grocery Store	Restaurant

D. Discussion

As I mentioned before, Buenos Aires city is a big city with high population. It is divided into 15 communes and 48 neighborhoods. It can be very varied where to live because of the different options of venues that has. In consequence, I used an algorithm like k-means to obtain which venues has in common and united with the information of the security of the communes, we can get a conclusion of which commune is the best to live

I also performed data analysis thorough the three last datasets and show with different plots how the crimes are distributed by the communes

So taking the final map, we can observe that living in the commune 1 is not the best but some others communes has the same venues and is more safe to live such as commune 5 or 6.
In this cluster we have in common this venue: Hotel, Coffee shop or Argentinian Restaurant

On the purple cluster, only one is dangerous than others. So, we can have a quiet life. They have in common this venue: Pizza Place, Bakery and Cafe

Some communes do not have any common venue, so we can live although commune 4 and 7 are not the best to live.

E. Conclusion

As a result, people want to get information of which commune they should be. Searching is the best way to have an idea and finally make the best decision

For further investigation, we have other features than only one like security because maybe some crimes are not reported or there another crime that is not considering or how many cops has each commune. On other hand, it is important to consider the price of the houses.

F. Bibliography

[1]. https://es.wikipedia.org/wiki/Buenos_Aires

[2]. https://data.buenosaires.gob.ar/dataset/mapa-del-delito

[3]. https://data.buenosaires.gob.ar/dataset/comunas/archivo/b0b627ac-5b47-4574-89ac-6999b63598ee

[4]. http://cdn.buenosaires.gob.ar/datosabiertos/datasets/barrios/barrios.geojson

[5]. https://developer.foursquare.com/