Finding a location to a gourmet restaurant in Buenos Aires, Argentina

Guadalupe Lamelas Fabbro

June 09, 2020

Capstone Project - The Battle of the Neighborhoods

Applied Data Science Capstone by IBM/Coursera

Introduction

The Ciudad Autónoma de Buenos Aires (CABA) is one of the twenty-four federal entities and the capital city of Argentina. Buenos Aires is a cosmopolitan city and a major tourist destination. Its complex infrastructure makes it one of the most important metropolises in Latin America and is a global city of alpha, category given its influences on commerce, finance, fashion, art, gastronomy, education, entertainment and mainly in its marked culture.

In this project I am going to try to find an optimal location for a company that wants to open a gourmet restaurant in this city. The decision will be taken based on the places where we can find more restaurants, galleries, because those neighborhoods are supposed to by haunted.

In order to do that, first we are going to collect the data of the Buenos Aires neighborhoods and their respective coordinates. Then we will work with our database to finally apply K means clustering and thus determine in which neighborhoods the investment would be convenient. When we consider all these problems, we can create a map and information chart where the real estate index is placed on Buenos Aires and each district is clustered according to the venue density.

Data description

To consider the problem we can list the datas as below:

- I found the data from “Datos Abiertos de Buenos Aires” that contains a list of the different neighborhoods from CABA and their coordinates[1]. I cleaned the data and convert it to a dataframe.
- I used Forsquare API to get the most common venues of given the neighborhoods from Buenos Aires.

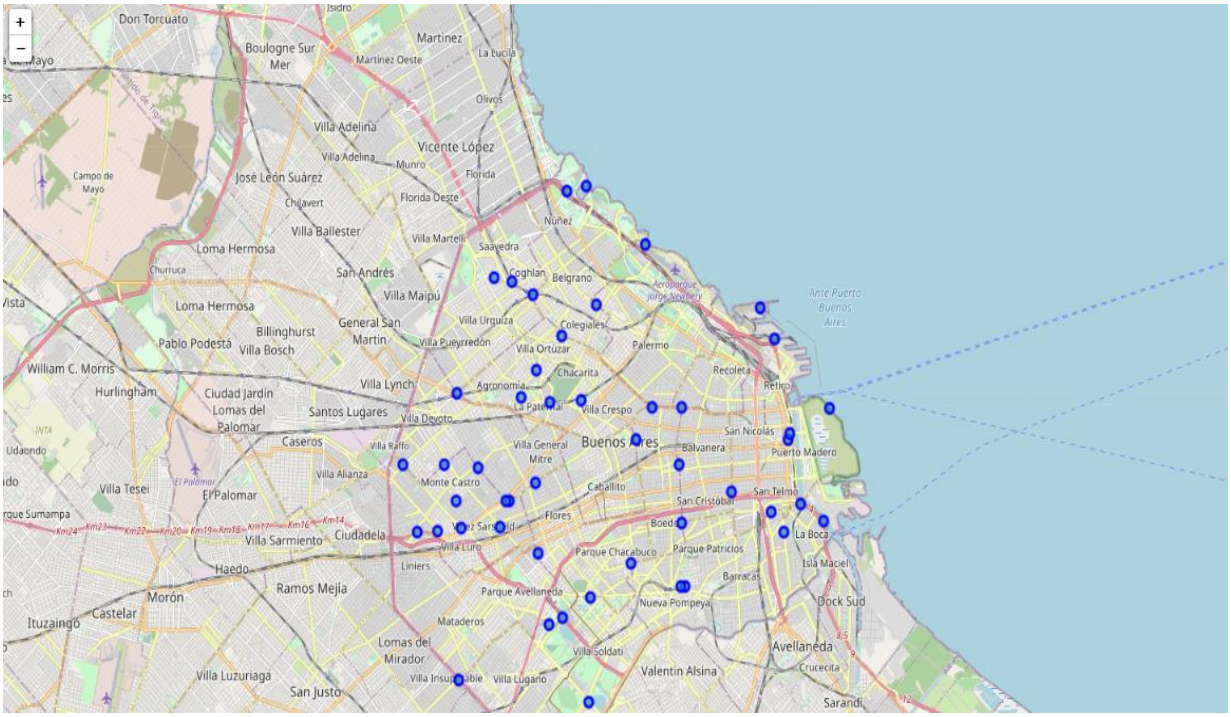
[1] http://cdn.buenosaires.gob.ar/datosabiertos/datasets/barrios/barrios.geojson

Methodology

As a database, I used GitHub repository in my study. My master data which has the main components Comuna, Neighborhoods, Latitude and Longitude information of the city.

	comuna	Neighborhood	Latitude	Longitude
0	15	CHACARITA	-34.595989	-58.452820
1	15	PATERNAL	-34.596558	-58.465577
2	15	VILLA CRESPO	-34.597827	-58.423753
3	11	VILLA DEL PARQUE	-34.614865	-58.494610
4	5	ALMAGRO	-34.614116	-58.412870

I used python folium library to visualize geographic details of Buenos Aires and its neighborhoods and I created a map of Buenos Aires with neighborhoods superimposed on top. I used latitude and longitude values to get the visual as below:



I utilized the Foursquare API to explore the neighborhoods and segment them. I designed the limit as 100 venue and the radius 500 meter for each neighborhood from their given latitude and longitude information. Here is a head of the list Venues name, category, latitude, and longitude information from Forsquare API.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	CHACARITA	-34.595989	-58.45282	Yeite	-34.596012	-58.449280	Deli / Bodega
1	CHACARITA	-34.595989	-58.45282	Movistar Arena	-34.594348	-58.448033	Stadium
2	CHACARITA	-34.595989	-58.45282	Margen del Mundo	-34.596987	-58.456835	Museum
3	CHACARITA	-34.595989	-58.45282	Tiro Loco	-34.598935	-58.452126	Café
4	CHACARITA	-34.595989	-58.45282	Alumni Fútbol 5	-34.597889	-58.451936	Soccer Field

In summary, 146 unique categories were returned by Foursquare, then I created a table which shows list of top 10 venue category for each borough in below table.

	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
0	AGRONOMIA	Hardware Store	Argentinian Restaurant	Train Station	Big Box Store	Bus Stop	Jewish Restaurant	Fast Food Restaurant	Furniture / Home Store	Food Truck
1	ALMAGRO	Argentinian Restaurant	Pizza Place	Indie Theater	Coffee Shop	Camera Store	Grocery Store	Restaurant	Pharmacy	Sushi Restaurant
2	BALVANERA	Theater	Gym / Fitness Center	Hotel	Café	Italian Restaurant	Argentinian Restaurant	Gym	Pizza Place	Bakery
3	BARRACAS	Supermarket	Gym	Restaurant	Café	Argentinian Restaurant	Pet Store	Plaza	Fish Market	Deli / Bodega
4	BELGRANO	Harbor / Marina	Argentinian Restaurant	Paintball Field	Sports Club	Gun Range	Women's Store	Fast Food Restaurant	Food Truck	Food Court

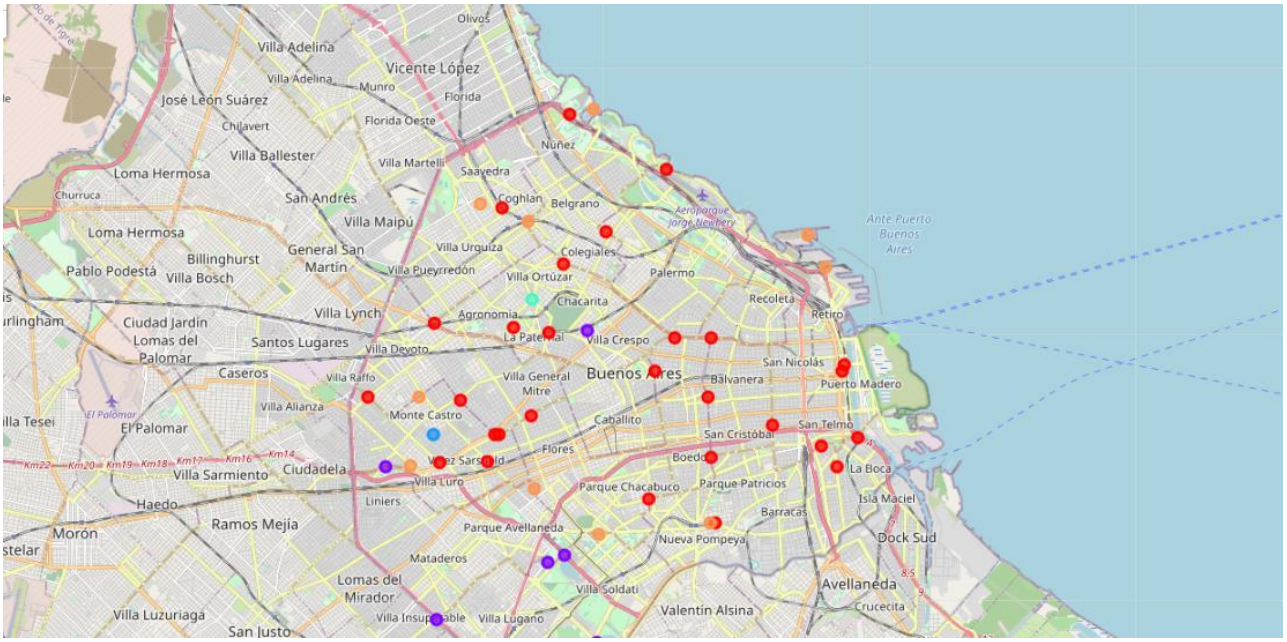
We have some common venue categories in neighborhoods. In this reason I used unsupervised learning K-means algorithm to cluster the neighborhoods. K-Means algorithm is one of the most common cluster methods of unsupervised learning.

I am going to run k-means to cluster the neighborhood into 6 clusters, and then, analyze the clusters looking for those with the greatest presence of soccer and sport venues

This is the new dataframe that includes the cluster as well as the top venues for each neighborhood.

comuna	Neighborhood	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Mos Commo Venue
15	CHACARITA	-34.595989	-58.452820	1.0	Museum	Bus Stop	Soccer Field	Stadium	Café	Deli / Bodega
15	PATERNAL	-34.596558	-58.465577	0.0	Art Gallery	BBQ Joint	Diner	Arts & Entertainment	Fish Market	Gastrop
15	VILLA CRESPO	-34.597827	-58.423753	0.0	Pizza Place	Middle Eastern Restaurant	Bar	BBQ Joint	Indie Theater	Café
11	VILLA DEL PARQUE	-34.614865	-58.494610	0.0	Breakfast Spot	BBQ Joint	Plaza	Argentinian Restaurant	Burger Joint	Restaur
5	ALMAGRO	-34.614116	-58.412870	0.0	Argentinian Restaurant	Pizza Place	Indie Theater	Coffee Shop	Camera Store	Grocery Store

Finally, let's visualize the resulting clusters in a map of Buenos Aires.



Results

The findings suggest that neighborhoods in cluster number one, will be possible places to open the restaurant. That is because the results shown as that neighborhoods of that cluster have many Argentinian restaurants, which means that residents and turists visit those places and choose them to eating out.

	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
1	PATERNAL	Art Gallery	BBQ Joint	Diner	Arts & Entertainment	Fish Market	Gastropub	Gas Station	Furniture / Home Store
2	VILLA CRESPO	Pizza Place	Middle Eastern Restaurant	Bar	BBQ Joint	Indie Theater	Café	Cajun / Creole Restaurant	Restaurar
3	VILLA DEL PARQUE	Breakfast Spot	BBQ Joint	Plaza	Argentinian Restaurant	Burger Joint	Restaurant	Café	Fish Market
4	ALMAGRO	Argentinian Restaurant	Pizza Place	Indie Theater	Coffee Shop	Camera Store	Grocery Store	Restaurant	Pharmacy
5	CABALLITO	Café	Argentinian Restaurant	Bar	Indie Theater	Bakery	Hotel	Park	Restaurar
6	VILLA SANTA RITA	Restaurant	Middle Eastern Restaurant	Ice Cream Shop	Korean Restaurant	Deli / Bodega	Café	Furniture / Home Store	Comfort Food Restaurar
8	VILLA REAL	Pharmacy	Bus Stop	Café	Pizza Place	Argentinian Restaurant	Fast Food Restaurant	Breakfast Spot	Bakery
10	FLORESTA	Ice Cream Shop	Korean Restaurant	Deli / Bodega	Restaurant	Furniture / Home Store	Middle Eastern Restaurant	Comfort Food Restaurant	Toy / Game Store

Neighborhoods in cluster number two seems to be places with little people movement, the most common venues are bus stop and soccer field, so I think that there is no place to a gourmet restaurant.

Clusters three, four and five although they are very tiny and have some restaurants, are discarded because comparing to cluster number one, they are not representative of restaurants. Same happens with cluster six, in which we can see some restaurants, but based on the results of our method, those neighborhoods seem to be more centrical, I mean, there are more pharmacies, gyms, etc.

Discussion

Although we have already given information to the stackholders about the optimal neighborhoods to establish the business, now it would be their job to find the location in one of these neighborhoods. With the use of data science tools, we could find the best place to rent according to cost and other variables of interest, but that is not part of this work.

Conclusion

The purpose of this project was to identify Buenos Aires neighborhoods that are optimal to invest in a gourmet restaurant. In order to do that first we collect the data of the Buenos Aires neighborhoods and their respective coordinates, we work the data and finally cluster the neighborhoods into 6 clusters. After a brief analysis we determined that cluster 1 was the most appropriate for our objective since there are lots of restaurants, and also some turistics attractions.