

IBM Data Science Capstone

Opening a new McDonald's franchise in CABA, Argentina

1. Introduction

McDonald's is a fast food restaurant all over the world that works with a franchise business model. When someone wants to open a new franchise it's often difficult to define where's the best place to locate it. The reason of this is due to the market saturation with not only McDonald's but also with other fast food restaurants like Burger King, Wendy's, KFC, Friday's and others.

For my final Data Science Capstone, I am going to analyze in which neighborhood in "Ciudad de Buenos Aires (CABA), Argentina" it's the best place to locate a new McDonald's using a clustering algorithm. This project could be useful to anyone interested in opening a McDonald's or any fast food restaurant in CABA.

2. Data

In order to solve this problem, I have used data from Wikipedia containing all the CABA neighborhood's details and added to a Pandas data frame in order to clean the data and make an exhaustive analysis.

Nombre del barrio	Superficie ¹	Habitantes (año 2007)	Densidad de habitantes (año 2007)	Comuna de pertenencia de la Ciudad de Buenos Aires
Agronomía	2,1 km²	34.580	8.645	Comuna 15
Almagro	4,1 km²	139.262	33.960	Comuna 5
Balvanera	4,4 km²	152.198	34.950	Comuna 3
Barracas	7,6 km²	77.474	10.194	Comuna 4
Belgrano	6,8 km²	138.942	20.433	Comuna 13
Boedo	2,6 km²	48.520	18.662	Comuna 5
Caballito	6,8 km²	183.396	25.830	Comuna 6
Chacarita	3,1 km²	27.440	9.800	Comuna 15
Coghlan	1,3 km²	19.177	14.752	Comuna 12
Colegiales	2,3 km²	56.998	21.922	Comuna 13
Constitución	2,1 km²	45.860	21.838	Comuna 1
Flores	7,8 km²	150.484	18.578	Comuna 7
Floresta	2,3 km²	39.473	16.447	Comuna 10
La Boca	3,1 km²	46.494	14.089	Comuna 4

3. Methodology

After creating the data frame, I cleaned the data by dropping the unnecessary columns and changing their names.

	Neighborhood	Population Density
0	Agronomía	8.645
1	Almagro	33.960
2	Balvanera	34.950
3	Barracas	10.194
4	Belgrano	20.433

Further, I used the Geocoder library to get the coordinates (latitude and longitude) for each Neighborhood in CABA.

	Neighborhood	Population Density	Latitude	Longitude
0	Agronomía	8.645	-34.59243	-58.49659
1	Almagro	33.960	-34.60324	-58.42043
2	Balvanera	34.950	-34.61011	-58.40602
3	Barracas	10.194	-34.64990	-58.38910
4	Belgrano	20.433	-34.56153	-58.45702

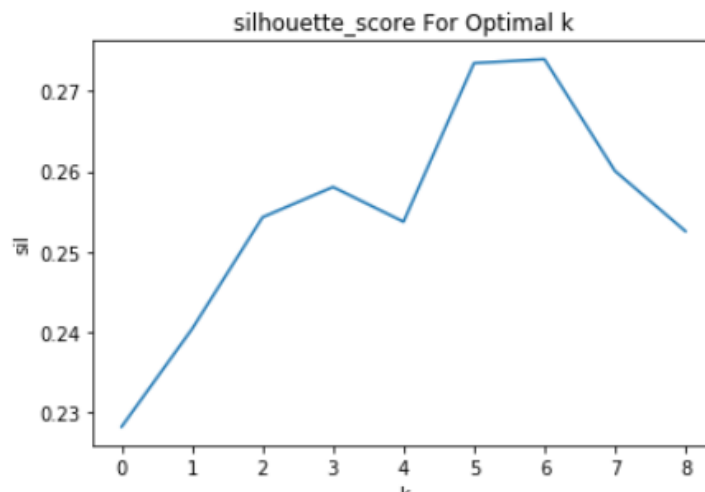
Then, I plotted the neighborhood's to visualize geographic details by using the Folium library.



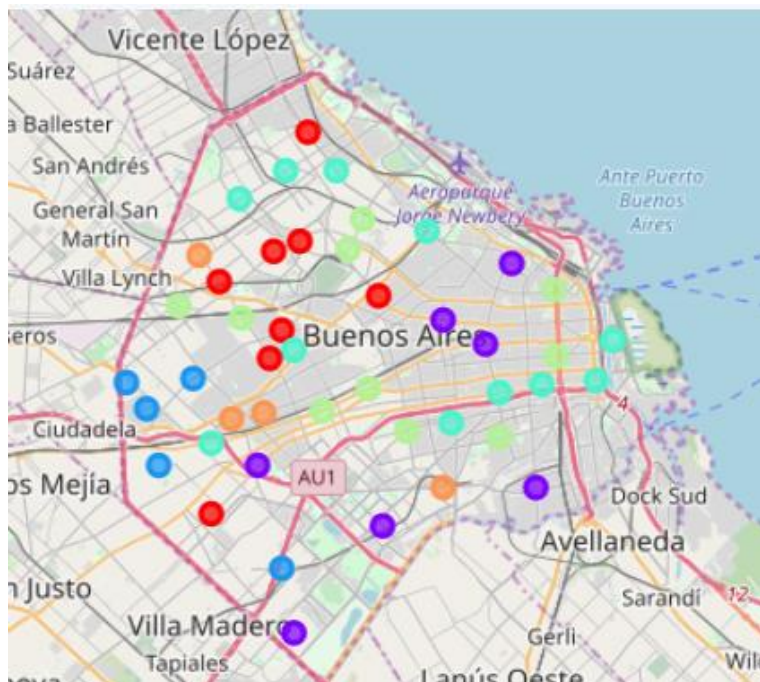
Next, I used the Foursquare API in order to get the venues for each neighborhood. Here I defined a 1.500 (m) radius and a limit of 100 venues per neighborhood. After converting the results into discrete values, I used the mean to determine the frequency in which the venues repeat for every neighborhood. Finally, I only kept the 8 more relevant venues that I considered to be competitors of McDonald's and I ordered them in a new data frame containing the top 3 for each neighborhood.

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
0	Agronomía	BBQ Joint	Restaurant	Burger Joint
1	Almagro	BBQ Joint	Restaurant	Hot Dog Joint
2	Balvanera	Sandwich Place	Hot Dog Joint	BBQ Joint
3	Barracas	BBQ Joint	Restaurant	Sandwich Place
4	Belgrano	BBQ Joint	Sandwich Place	Beer Bar

After I obtained the above data frame, I clustered the data using K-Means Clustering Machine Learning algorithm. I chose the value of K=6 using the Silhouette Score as shown below.



Finally, I plotted the clusters in a map using the Folium library.



4. Results

As result, I got 6 clusters of CABA based on the McDonald's competitors. Those clusters were named:

- **Cluster 1:** The most common venues are BBQ Joint and Restaurants.
- **Cluster 2:** The most common venues are BBQ Joint, Restaurants and Sandwich Places.
- **Cluster 3:** The most common venues are Fast Food Restaurants and Sandwich Places.
- **Cluster 4:** The most common venues are BBQ Joints and Sandwich Places.
- **Cluster 5:** The most common venues are Restaurants and Burger/Beer Bars.
- **Cluster 6:** The most common venues are BBQ Joints

5. Discussion

Based on the results we can identify which clusters could be a good option to locate a McDonald's.

Without any doubt, cluster 3 would be the worst option given that the top venues here are Fast Food Restaurants.

Then, clusters 2, 4 and 5 would be a bad option given that Sandwich Places and Burger Bars are direct competitors.

And finally, the best clusters would be Cluster 1 or 6 because these clusters top venues are not direct competitors of McDonald's.

6. Conclusion

I was able to cluster CABA in 6 clusters identifying the best/worst clusters to locate the McDonald's.

If I would have to choose the best neighborhood to locate the McDonald's it would be Parque Chas in Cluster 1 given that it's the place with the higher population density that's in either cluster 1 or 6. To decide this I used the information from the same Wikipedia table as before.

	Neighborhood	Population Density	Latitude	Longitude
27	Recoleta	34.959	-34.587540	-58.397170
2	Balvanera	34.950	-34.610110	-58.406020
1	Almagro	33.960	-34.603240	-58.420430
24	Parque Chas	28.195	-34.584130	-58.478140
6	Caballito	25.830	-34.622620	-58.446220
30	San Cristóbal	23.803	-34.623700	-58.401490
35	Villa Crespo	23.647	-34.596690	-58.442250
9	Colegiales	21.922	-34.575020	-58.447770
10	Constitución	21.838	-34.621210	-58.386550
4	Belgrano	20.433	-34.561530	-58.457020
32	San Telmo	19.976	-34.620390	-58.368490
17	Montserrat	19.800	-34.613510	-58.381740