Starting a pharmacy business in Buenos Aires

José Ignacio Chajchir

April 26, 2021

Contents

1. Introduction

1.1 Background

Buenos Aires is the capital city of Argentina, the country where I was born and where I live. As the term “Buenos Aires” could represent different areas, the term “CABA” (stands for Ciudad Autonoma de Buenos Aires) is more accurate to differentiate the city from other places.

Large cities have a lot of diverse Neighborhoods and CABA is not an exception. If someone is interested in starting a business, location will be one of the most important factors for that business to be profitable.

1.2 Problem

The goal of this project is to, based on data, identify which are the most suitable neighborhoods in CABA to start a pharmacy business.

1.3 Target Audience

This report shall be of great interest for:

- Individual investors, especially those with background in the pharmacy business.
- Pharmacy chains interested in opening a new store

It should be useful for any organization or Specialists who perform demographic analysis.

2. Data acquisition and cleaning

2.1 Data targeting

To build a useful and neat dataframe that could lead into meaningful results, I thought about which features would be useful by asking some questions for each neighborhood.

a. How many pharmacies already exist in the neighborhood?
   The rate of persons per existing pharmacies should be a very important input parameter for the decision. If there are too many persons per pharmacies in the neighborhood, a new pharmacy will probably have clients.

b. How many persons live? How old are them?
   As mentioned in the point before, the rate of persons per existing pharmacies is very useful information, so these questions should be answered in order to calculate the rate. The age of the persons could be also valuable as older people tend to buy more medication.

c. Are these persons consumers of pharmacy products?
   If people living in the neighborhood have a low income, they will probably avoid spending money in esthetic or cosmetic products.

2.2 Data collection and cleaning

2.2.1 How many pharmacies already exists in the neighborhood?

For question “How many pharmacies already exists in the neighborhood?”, I used data from the search venue foursquare API. Although the API let you search only for pharmacies (by setting category ID) there´s a limit of 50 results per request. Therefore, data for the complete CABA area had to be collected using multiple requests, where each request belonged to a unique geographic point inside CABA area.

A grid of points was defined with certain parameters:

- The grid must represent CABA area. This area was obtained from a geojson file that contains Argentinian provinces.
- The distance within points must be defined based on:
    - Foursquare API radius parameter
    - 50 results per request limit

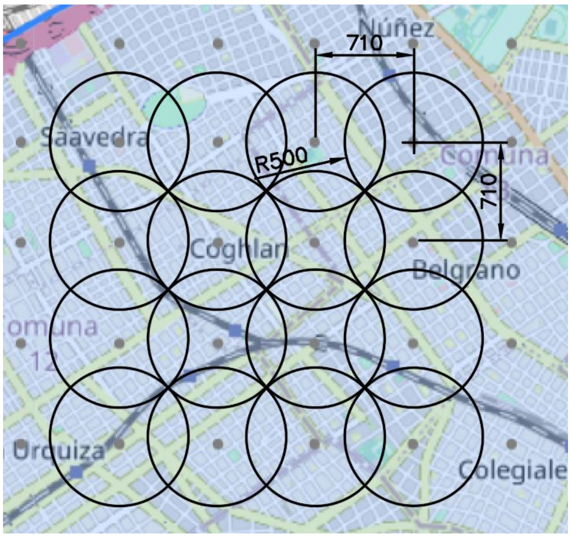

Figure 1: Shows chosen parameters for grid definition. With a radius of 500 meters and Longitude distance = Latitude distance = 710meters, the queries should find all the pharmacies in the target area. A radius of 500 meters doesn´t exceed the 50 results per query limit.

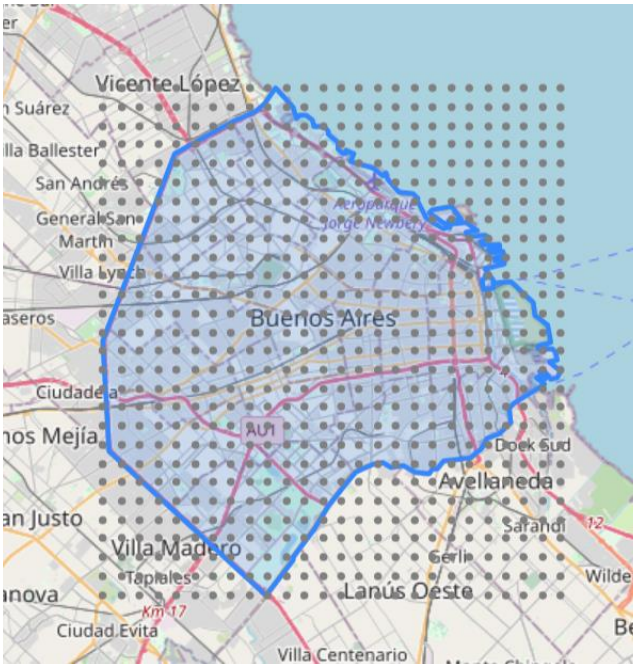

Figure 2: Grid of points and CABA polygon.

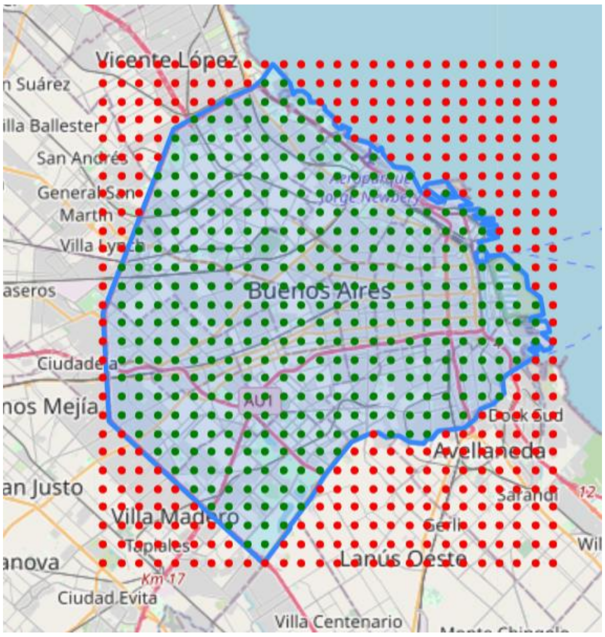

Figure 3: Grid of points classified (Green: inside CABA polygon – Red: outside CABA polygon)

Once the points inside CABA were defined, the foursquare API was tested with a few points to get familiar with the requested data.

I then created a function to obtain the data for every point inside CABA and compile the information into one single dataframe. After dropping repeated pharmacies (as showed in figure 1, intersection between circular areas were scanned twice), the dataframe looked like this:

	Point_number	Point Latitude	Point Longitude	Pharmacy name	Pharmacy Latitude	Pharmacy Longitude	Pharmacy distance to point	Pharmacy Adress	Pharmacy Address
0	2	-34.692071	-58.476496	Farmacia San Pedro	-34.690053	-58.481157	482	NaN	Cabildo
1	7	-34.685449	-58.476496	Farmacity	-34.686374	-58.476301	104	NaN	Cnel. Martiniano Chilavert 6461
2	7	-34.685449	-58.476496	Farmacia Belen	-34.685821	-58.475219	123	NaN	Cnel. Martiniano Chilavert 6364
3	7	-34.685449	-58.476496	Óptica Sacaria	-34.686685	-58.476368	138	NaN	Chilavert
4	9	-34.685449	-58.460790	Farmacia San Alberto	-34.686954	-58.461395	176	NaN	Cañada De Gomez 5201
5	14	-34.678828	-58.476496	Farmacia Inglesa	-34.676492	-58.476758	261	NaN	Somellera 5725
6	14	-34.678828	-58.476496	Optica De Betina	-34.677923	-58.474880	178	NaN	NaN
7	14	-34.678828	-58.476496	Farmacia Cientifica	-34.676375	-58.475732	281	NaN	NaN

Figure 4: Dataframe of unique pharmacies located in CABA area (total of 752 pharmacies).

Finally, I needed to find out in which neighborhood each pharmacy was located. Therefore, I decided to use a geojson file that contains CABA neighborhoods limits.

Figure 5: CABA neighborhoods geojson file.

After running a for loop, the mentioned dataframe was completed with the "Neighborhood" column, and using groupby function I finally got the answer of "How many pharmacies already exists in the neighborhood?"

	Neighborhood	counts
0	ALMAGRO	26
1	BALVANERA	52
2	BARRACAS	16
3	BELGRANO	43
4	BOCA	4
5	BOEDO	6
6	CABALLITO	42

Figure 6: A total of 45 unique neighborhoods were listed in this dataframe

2.2.2 How many persons live in the neighborhood? How old are them?

I explored several governmental websites to discover the answers. I finally found demographic statistical data of CABA with the desired information.

Población total por sexo y grupo quinquenal de edad según barrio. Ciudad de Buenos Aires. Año 2010

Comuna	Barrio	Total	Total varón	Total mujer	0-4			5-9			10-14			15-19			20-24		
					Total	Varón	Mujer	Total	Varón	Mujer	Total	Varón	Mujer	Total	Varón	Mujer	Total	Varón	Mujer
Total CABA		2.890.151	1.329.681	1.560.470	165.638	84.382	81.256	156.372	79.472	76.900	150.501	76.354	74.147	167.681	83.338	84.343	228.125	110.915	117.2
Comuna 1	**Total**	**205.886**	**98.097**	**107.789**	**11.669**	**5.926**	**5.743**	**10.149**	**5.079**	**5.070**	**9.913**	**4.973**	**4.940**	**11.582**	**5.737**	**5.845**	**19.494**	**9.721**	**9.7**
	Constitucion	**44.107**	**20.666**	**23.441**	2.442	1.197	1.245	2.444	1.206	1.238	2.447	1.232	1.215	2.632	1.310	1.322	3.677	1.893	1.7
	Monserrat	**39.914**	**18.940**	**20.974**	1.862	956	906	1.694	869	825	1.680	860	820	2.139	1.054	1.085	3.769	1.933	1.8
	Puerto Madero	**6.726**	**3.611**	**3.115**	544	287	257	351	191	160	313	171	142	305	167	138	597	324	27
	Retiro	**65.413**	**31.195**	**34.218**	4.865	2.500	2.365	3.862	1.898	1.964	3.638	1.805	1.833	4.165	2.024	2.141	6.774	3.220	3.5
	San Nicolas	**29.273**	**14.046**	**15.227**	972	506	466	877	440	437	1.005	476	529	1.341	655	686	3.035	1.515	1.5
	San Telmo	**20.453**	**9.639**	**10.814**	984	480	504	921	475	446	830	429	401	1.000	527	473	1.642	836	8
Comuna 2	**Total**	**157.932**	**68.042**	**89.890**	**6.220**	**3.206**	**3.014**	**5.696**	**2.879**	**2.817**	**5.486**	**2.765**	**2.721**	**8.239**	**3.786**	**4.453**	**17.054**	**7.457**	**9.5**
	Recoleta	**157.932**	**68.042**	**89.890**	6.220	3.206	3.014	5.696	2.879	2.817	5.486	2.765	2.721	8.239	3.786	4.453	17.054	7.457	9.5
Comuna 3	**Total**	**187.537**	**85.601**	**101.936**	**9.889**	**5.078**	**4.811**	**9.227**	**4.723**	**4.504**	**8.986**	**4.570**	**4.416**	**10.211**	**5.023**	**5.188**	**16.523**	**7.945**	**8.5**
	Balvanera	**138.926**	**63.273**	**75.653**	7.262	3.715	3.547	6.694	3.415	3.279	6.385	3.284	3.101	7.521	3.715	3.806	12.881	6.168	6.7
	San Cristobal	**48.611**	**22.328**	**26.283**	2.627	1.363	1.264	2.533	1.308	1.225	2.601	1.286	1.315	2.690	1.308	1.382	3.642	1.777	1.8
Comuna 4	**Total**	**218.245**	**103.166**	**115.079**	**15.589**	**8.013**	**7.576**	**15.052**	**7.613**	**7.439**	**15.106**	**7.696**	**7.410**	**15.828**	**7.880**	**7.948**	**18.189**	**9.079**	**9.1**
	Barracas	**89.452**	**42.737**	**46.715**	7.173	3.696	3.477	6.521	3.246	3.275	6.517	3.358	3.159	6.964	3.427	3.537	8.027	4.015	4.0
	La Boca	**45.113**	**21.305**	**23.808**	3.133	1.592	1.541	3.180	1.620	1.560	3.317	1.673	1.644	3.435	1.742	1.693	3.743	1.904	1.8
	Nueva Pompeya	**42.695**	**20.232**	**22.463**	2.705	1.384	1.321	2.753	1.420	1.333	2.797	1.415	1.382	2.899	1.465	1.434	3.399	1.711	1.6
	Parque Patricios	**40.985**	**18.892**	**22.093**	2.578	1.341	1.237	2.598	1.327	1.271	2.475	1.250	1.225	2.530	1.246	1.284	3.020	1.449	1.5
Comuna 5	**Total**	**179.005**	**80.806**	**98.199**	**9.006**	**4.594**	**4.412**	**8.659**	**4.469**	**4.190**	**8.295**	**4.198**	**4.097**	**9.390**	**4.696**	**4.694**	**13.298**	**6.489**	**6.8**
	Almagro	**131.699**	**58.871**	**72.828**	6.431	3.287	3.144	5.971	3.084	2.887	5.754	2.942	2.812	6.645	3.258	3.387	9.876	4.753	5.1
	Boedo	**47.306**	**21.935**	**25.371**	2.575	1.307	1.268	2.688	1.385	1.303	2.541	1.256	1.285	2.745	1.438	1.307	3.422	1.736	1.6
Comuna 6	**Total**	**176.076**	**78.870**	**97.206**	**9.095**	**4.577**	**4.518**	**8.409**	**4.201**	**4.208**	**7.632**	**3.844**	**3.788**	**8.565**	**4.234**	**4.331**	**11.646**	**5.597**	**6.0**
	Caballito	**176.076**	**78.870**	**97.206**	9.095	4.577	4.518	8.409	4.201	4.208	7.632	3.844	3.788	8.565	4.234	4.331	11.646	5.597	6.0
Comuna 7	**Total**	**220.591**	**102.481**	**118.110**	**14.368**	**7.414**	**6.954**	**13.633**	**6.906**	**6.727**	**13.200**	**6.682**	**6.518**	**14.038**	**7.055**	**6.983**	**17.561**	**8.636**	**8.9**
	Flores	**164.310**	**76.326**	**87.984**	11.263	5.837	5.426	10.484	5.315	5.169	10.216	5.190	5.026	10.797	5.399	5.398	13.730	6.738	6.9
	Parque Chacabuco	**56.281**	**26.155**	**30.126**	3.105	1.577	1.528	3.149	1.591	1.558	2.984	1.492	1.492	3.241	1.656	1.585	3.831	1.898	1.9
Comuna 8	**Total**	**187.237**	**89.545**	**97.692**	**16.754**	**8.521**	**8.233**	**15.552**	**7.888**	**7.664**	**14.927**	**7.564**	**7.363**	**15.445**	**7.639**	**7.806**	**17.423**	**8.561**	**8.8**
	Villa Lugano	**126.374**	**60.371**	**66.003**	11.131	5.706	5.425	10.277	5.285	4.992	9.752	4.971	4.781	10.224	5.045	5.179	11.776	5.810	5.9
	Villa Riachuelo	**14.084**	**6.607**	**7.477**	836	438	398	895	453	442	851	423	428	901	438	463	1.071	525	5
	Villa Soldati	**46.779**	**22.567**	**24.212**	4.787	2.377	2.410	4.380	2.150	2.230	4.324	2.170	2.154	4.320	2.156	2.164	4.576	2.226	2.3
Comuna 9	**Total**	**161.797**	**76.207**	**85.590**	**10.091**	**5.115**	**4.976**	**10.195**	**5.220**	**4.975**	**9.985**	**5.183**	**4.802**	**10.444**	**5.207**	**5.237**	**11.672**	**5.867**	**5.8**

Figure 7: Excel file showing CABA population by sex, age group and neighborhood.

I then converted this excel file into a dataframe and performed cleaning operations. I also added a calculated column to get population above 65 years old. The desired dataframe was obtained:

	Neighborhood	Total	Total_more_65
0	CONSTITUCION	44107.0	6515.0
1	MONSERRAT	39914.0	5868.0
2	PUERTO MADERO	6726.0	490.0
3	RETIRO	65413.0	8336.0
4	SAN NICOLAS	29273.0	4325.0
5	SAN TELMO	20453.0	3590.0
6	RECOLETA	157932.0	31265.0
7	BALVANERA	138926.0	22096.0
8	SAN CRISTOBAL	48611.0	7932.0
9	BARRACAS	89452.0	9724.0
10	LA BOCA	45113.0	5661.0
11	NUEVA POMPEYA	42695.0	6617.0
12	PARQUE PATRICIOS	40985.0	6118.0
13	ALMAGRO	131699.0	23199.0
14	BOEDO	47306.0	7601.0

Figure 8: "Barrio" means Neigborhood in Spanish. "Total" represents the total population and "Total_more_65" represents the population aged more than 65.

2.2.3 Are these persons consumers of pharmacy products?

I searched online data of population income per neighborhood but unfortunately, I couldn´t find well-ordered and reliable information. What I did find instead, is real state information of price per square meter of each neighborhood. For the matter of this project, income and price per square meter were considered correlated.

	Neighborhood	USD/m2
0	Puerto Madero	5786
1	Palermo	3313
2	Belgrano	3164
3	Nuðez	3039
4	Recoleta	2973
5	Retiro	2926
6	Colegiales	2888
7	Villa Urquiza	2801
8	Coghlan	2690
9	Chacarita	2643

Figure 9

2.3 Data sources

Multiple datasets and geodata where needed to perform section 2.2:

- How many pharmacies already exists in the neighborhood? (2.2.1)
  - Zip file containing CABA shape file: https://infra.datos.gob.ar/catalog/modernizacion/dataset/7/distribution/7.34/download/provincias.zip
  - Existing Pharmacies in CABA: Foursquare API
  - CABA neighborhoods geojson file: https://cdn.buenosaires.gob.ar/datosabiertos/datasets/barrios/barrios.geojson
- How many persons live? How old are them? (2.2.2)
  - Statistics about neighborhood population: https://www.estadisticaciudad.gob.ar/eyc/?p=28008/PB_barrio_ARIP_CNP2010.xls
- Are these persons consumers of pharmacy products? (2.2.3)
  - Neighborhoods Price per Square meter: https://www.zonaprop.com.ar/noticias/zpindex/

2.4 Data preparation

In this section I will describe how dataframes obtained in section 2.2 where merged and modified so as to get the appropriate data needed to feed a model intended to answer the main question of this project.

At first, output data of section 2.2.2 was merged with output data of section 2.2.1. Since they had 48 and 45 neighborhoods respectively, a left join merge was performed.

Secondly, the obtained dataframe was merged with output data from section 2.2.3. After performing cleaning operations and adding calculated columns, the following table was obtained:

	Neighborhood	Total_pop	Total_pop_+65	Pharmacies	USD/m2	pop_per_pharma	+65pop_per_pharma
0	CONSTITUCION	44107.00	6515.00	6.00	1960	7351.17	1085.83
1	MONSERRAT	39914.00	5868.00	30.00	2083	1330.47	195.60
2	PUERTO MADERO	6726.00	490.00	4.00	5786	1681.50	122.50
3	RETIRO	65413.00	8336.00	22.00	2926	2973.32	378.91
4	SAN NICOLAS	29273.00	4325.00	48.00	2163	609.85	90.10
5	SAN TELMO	20453.00	3590.00	8.00	2417	2556.62	448.75
6	RECOLETA	157932.00	31265.00	88.00	2973	1794.68	355.28
7	BALVANERA	138926.00	22096.00	52.00	2043	2671.65	424.92
8	SAN CRISTOBAL	48611.00	7932.00	6.00	2015	8101.83	1322.00
9	BARRACAS	89452.00	9724.00	16.00	2364	5590.75	607.75
10	LA BOCA	45113.00	5661.00	4.00	1781	11278.25	1415.25
11	NUEVA POMPEYA	42695.00	6617.00	5.00	1872	8539.00	1323.40
12	PARQUE PATRICIOS	40985.00	6118.00	7.00	2022	5855.00	874.00

Figure 10

Columns “pop_per_pharma” and “+65pop_per_pharma” formulas:

- pop_per_pharma=Total_pop/Pharmacies
- +65pop_per_pharma=Total_pop_+65/Pharmacies

As mentioned in section 2.1, the rate of persons per existing pharmacies is a very important feature. I decided to add +65pop_per_pharma as older people are more likely to make pharmacies purchases, especially medication. So, if neighborhood A has a similar “pop_per_phama” as neighborhood B but neighborhood A has a better “+65pop_per_pharma” than B Neighborhood A should be one step ahead than B to start a new pharmacy.

3. Methodology

3.1 Exploratory analysis

The 3 main features were plotted in a Bar chart.

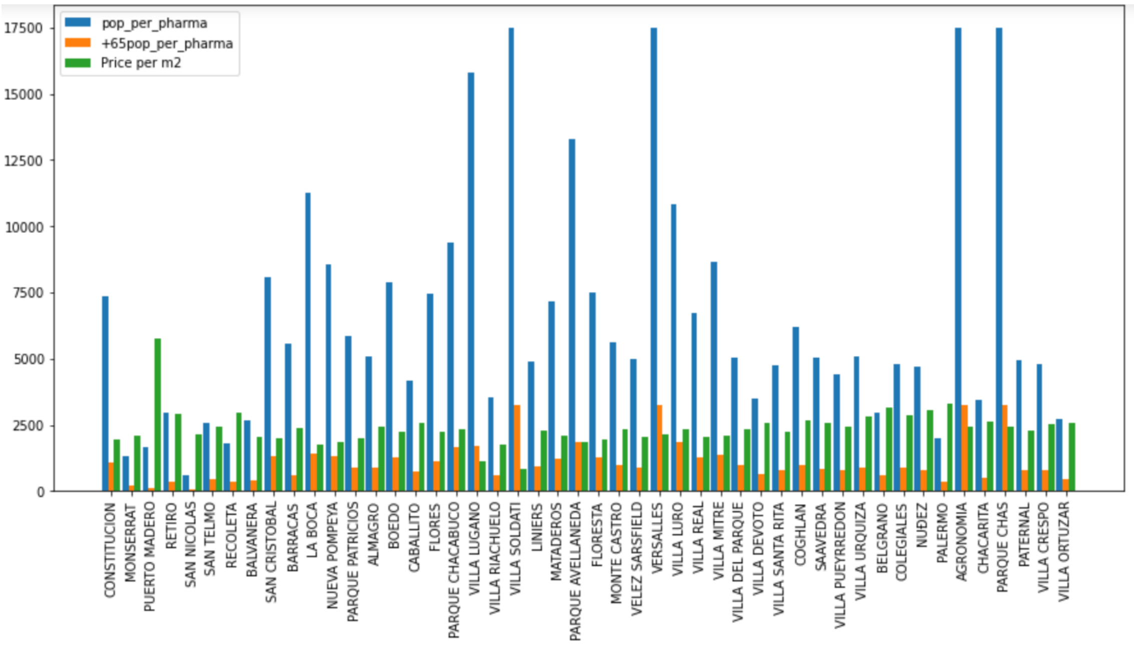

Figure 11

At a first glance, we can observe there´s plenty of variation over the 46 neighborhoods.

Although with this plot it is somehow possible to identify possible neighborhoods where to start the business, I decided that grouping the neighborhoods would be the best approach to build the outcome of this project.

3.2 Model

As my goal was to identify groups of neighborhoods based on their similarity, I decided to use K-means. Despite its simplicity, K-means is vastly used for clustering in many data science applications, especially useful if you need to quickly discover insights from unlabeled data.

From Data showed in figure 10, I decided to slice it into 2 different datasets and run separate K-means clustering algorithms

- k_means_data_A: USD/m2 - pop_per_phama
- k_means_data_B: USD/m2 - +65pop_per_phama

The purpose of this division was to evaluate the results considering two groups of residents: total residents and residents age 65-plus.
Once the slicing operations were performed, datasets A and B were scaled using StandardScale. Before running K-means, a K=5 was defined almost randomly as it appeared to a be reasonable number based on to the quantity of neighborhoods.

After running both algorithms, dataframe from figure 10 was completed with the results of the clustering processes:

	Neighborhood	Total_pop	Total_pop_+65	Pharmacies	USD/m2	pop_per_pharma	+65pop_per_pharma	cluster_A	cluster_B
0	CONSTITUCION	44107.0	6515.0	7.0	1960	6301.000000	930.714286	0	3
1	MONSERRAT	39914.0	5868.0	30.0	2083	1330.466667	195.600000	2	3
2	PUERTO MADERO	6726.0	490.0	4.0	5786	1681.500000	122.500000	3	2
3	RETIRO	65413.0	8336.0	22.0	2926	2973.318182	378.909091	1	0
4	SAN NICOLAS	29273.0	4325.0	48.0	2163	609.854167	90.104167	2	3

Figure 12

4. Results

From results in figure 12, two pivot charts were created to understand the properties of the clusters generated from dataset A and dataset B

	pop_per_pharma					USD/m2				
	my25	median	my75	mean	std	my25	median	my75	mean	std
cluster_A										
0	7391.366162	7993.083333	8853.729167	8320.793771	1537.371980	1952.25	2075.0	2229.75	2079.583333	178.458275
1	2949.558140	3501.105263	4777.363636	3802.563851	1318.784161	2591.00	2801.0	2973.00	2829.615385	243.788343
2	3308.663462	4862.336601	5034.950175	4129.154186	1564.791434	2074.75	2282.0	2376.50	2235.375000	202.111809
3	1681.500000	1681.500000	1681.500000	1681.500000	NaN	5786.00	5786.0	5786.00	5786.000000	NaN
4	16219.812500	17489.000000	17489.000000	16510.000000	1708.806433	1304.25	2000.5	2354.75	1802.166667	677.952629

Figure 13: pivot chart results dataset A clustering

As shown in figure 13, after running K-means over dataset A we can identify that cluster 4 is the best performer for feature "pop_per_pharma". Neighborhoods in this cluster have an average of 16.510 persons per pharmacy, almost twice the rate that the 2nd performer has (8.320).

Regarding price per m2, which represents neighborhood income as we mentioned on section 2.2.3, we can notice that cluster 4 is the worst performer.

Following this type of analysis, we can describe each cluster:

Cluster_A_0:

- Medium rate of persons per pharmacy
- Low income
- Cluster_A_0 neighborhoods: ['CONSTITUCION', 'SAN CRISTOBAL', 'LA BOCA', 'NUEVA POMPEYA', 'BOEDO', 'FLORES', 'PARQUE CHACABUCO', 'MATADEROS', 'FLORESTA', 'VILLA LURO', 'VILLA REAL', 'VILLA MITRE']

Cluster_A_1:

- Low rate of persons per pharmacy
- Medium income
- Cluster_A_1 neighborhoods:
  ['RETIRO', 'RECOLETA', 'CABALLITO', 'VILLA DEVOTO', 'COGHLAN', 'SAAVEDRA', 'VILLA URQUIZA', 'BELGRANO', 'COLEGIALES', 'NUĐEZ', 'PALERMO', 'CHACARITA', 'VILLA ORTUZAR']

Cluster_A_2:

- Low rate of persons per pharmacy
- Medium income
- Cluster_A_2 neighborhoods:
  ['MONSERRAT', 'SAN NICOLAS', 'SAN TELMO', 'BALVANERA', 'BARRACAS', 'PARQUE PATRICIOS', 'ALMAGRO', 'VILLA RIACHUELO', 'LINIERS', 'MONTE CASTRO', 'VELEZ SARSFIELD', 'VILLA DEL PARQUE', 'VILLA SANTA RITA', 'VILLA PUEYRREDON', 'PATERNAL', 'VILLA CRESPO']

Cluster_A_3:

- Very low rate of persons per pharmacy
- Very high income
- Cluster_A_3 neighborhoods:
  ['PUERTO MADERO']

Cluster_A_4:

- Very high rate of persons per pharmacy
- Low income
- Cluster_A_4 neighborhoods:
  ['VILLA LUGANO', 'VILLA SOLDATI', 'PARQUE AVELLANEDA', 'VERSALLES', 'AGRONOMIA', 'PARQUE CHAS']

The same analysis was carried out with results from dataset B

	+65pop_per_pharma					USD/m2				
	my25	median	my75	mean	std	my25	median	my75	mean	std
cluster_B										
0	454.281250	713.090226	836.709091	661.099453	222.061422	2588.75	2745.5	2961.25	2808.000000	247.794580
1	1267.333333	1339.325000	1663.000000	1455.698611	233.443425	1864.50	2040.5	2146.50	1977.500000	324.094964
2	122.500000	122.500000	122.500000	122.500000	NaN	5786.00	5786.0	5786.00	5786.000000	NaN
3	582.500000	799.000000	916.000000	718.061710	288.909237	2050.00	2261.0	2357.00	2201.823529	191.114898
4	3277.000000	3277.000000	3277.000000	3277.000000	0.000000	1828.25	2289.5	2422.75	1961.500000	760.791036

Figure 13: pivot chart results dataset B clustering

Cluster_B_0:

- Very low rate of persons per pharmacy
- Medium income
- Cluster_B_0 neighborhoods:
  ['RETIRO', 'RECOLETA', 'CABALLITO', 'VILLA DEVOTO', 'COGHLAN', 'SAAVEDRA', 'VILLA URQUIZA', 'BELGRANO', 'COLEGIALES', 'NUĐEZ', 'PALERMO', 'CHACARITA', 'VILLA CRESPO', 'VILLA ORTUZAR']

Cluster_B_1:

- Medium rate of persons per pharmacy
- Low income
- Cluster_B_1 neighborhoods:
  ['SAN CRISTOBAL', 'LA BOCA', 'NUEVA POMPEYA', 'BOEDO', 'PARQUE CHACABUCO', 'VILLA LUGANO', 'MATADEROS', 'PARQUE AVELLANEDA', 'FLORESTA', 'VILLA LURO', 'VILLA REAL', 'VILLA MITRE']

Cluster_B_2:

- Low rate of persons per pharmacy
- Very high income
- Cluster_B_2 neighborhoods:
  ['PUERTO MADERO']

Cluster_B_3:

- Low rate of persons per pharmacy
- Medium income
- Cluster_B_3 neighborhoods:
  ['CONSTITUCION', 'MONSERRAT', 'SAN NICOLAS', 'SAN TELMO', 'BALVANERA', 'BARRACAS', 'PARQUE PATRICIOS', 'ALMAGRO', 'FLORES', 'VILLA RIACHUELO', 'LINIERS', 'MONTE CASTRO', 'VELEZ SARSFIELD', 'VILLA DEL PARQUE', 'VILLA SANTA RITA', 'VILLA PUEYRREDON', 'PATERNAL']

Cluster_B_4:

- Very high rate of persons per pharmacy
- Low income
- Cluster_B_4 neighborhoods:
  ['VILLA SOLDATI', 'VERSALLES', 'AGRONOMIA', 'PARQUE CHAS']

5. Discussion

Both sets of results (Cluster_A and Cluster_B) show a clear inverse correlation between persons per pharmacy rate and income. But which of them is more important to select the right cluster? To answer these questions a deeper research should be carried out.

However, with our current results, **I would recommend clusters Cluster_A_4 and Cluster_B_4 for starting a pharmacy**. Although they have the lowest value for the income feature (represented as USD/m2) they are not far away from the average of the cluster values:

Average= 2.950 USD/m2

Cluster_A_4= 1.802 USD/m2

Cluster_B_4= 1.961 USD/m2

In addition, pharmacies sell mostly medication which are necessity goods, so people are likely to consume them despite their level of income.

6. Conclusion

6.1 Final recommendation

Finally, Cluster_A_4 and Cluster_B_4 were merged to check their differences and similarities. Both clusters have these 4 neighborhoods in common:

['VILLA SOLDATI', 'VERSALLES', 'AGRONOMIA', 'PARQUE CHAS']

Cluster_A_4 has 2 additional neighborhoods **['VILLA LUGANO', 'PARQUE AVELLANEDA']**.

This led to the final recommendation of this study:

- Highly recommended neighborhoods:

  ['VILLA SOLDATI', 'VERSALLES', 'AGRONOMIA', 'PARQUE CHAS']

- Alternative recommended neighborhoods:

  ['VILLA LUGANO', 'PARQUE AVELLANEDA'].

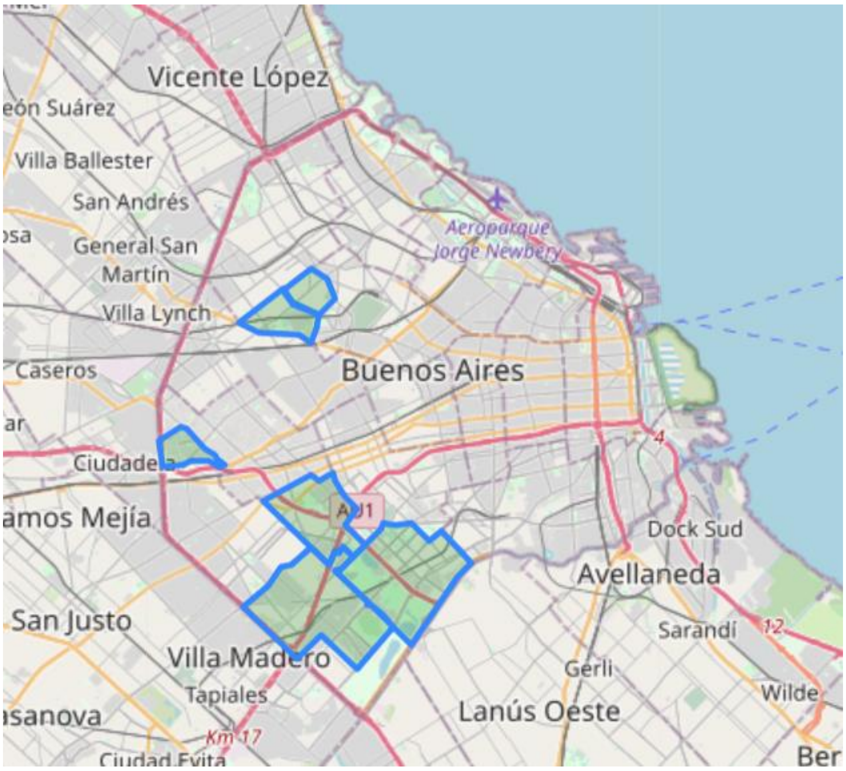

Figure 14: Recommended neighborhoods

6.2 Future enhancements

This model could be improved by boosting accuracy of existing data and getting additional information to generate new features.

Boosting accuracy:

- Using real Income data instead of price per square meters should be considered. Probably this kind of data could be provided by the Government of CABA city.

Getting additional data:

- Data of nearby hospitals and medical centers.
- Data of nearby commercial areas/shopping malls.