New commercial store selling sportswear

New start up: Venues Data Analysis at Buenos Aires

A. Introduction

A.1. Description and discussion of the background

Buenos Aires is the capital and largest city of Argentina. The city is located on the western shore of the estuary of the Río de la Plata, on the South American continent's southeastern coast. The Greater Buenos Aires conurbation, which also includes several Buenos Aires Province districts, constitutes the fourth-most populous metropolitan area in the Americas, with a population of around 15.6 million.

The city of Buenos Aires is neither part of Buenos Aires Province nor the Province's capital; rather, it is an autonomous district.

Buenos Aires' quality of life was ranked 91st in the world in 2018, being one of the best in Latin America. In 2012, it was the most visited city in South America, and the second-most visited city of Latin America (behind Mexico City)

Buenos is known for its preserved Eclectic European architecture and rich cultural life.

Buenos Aires is a multicultural city, being home to multiple ethnic and religious groups. Several languages are spoken in the city in addition to Spanish, contributing to its culture and the dialect spoken in the city and in some other parts of the country. This is because in the last 151 years the city, and the country in general, has been a major recipient of millions of immigrants from all over the world, making it a melting pot where several ethnic groups live together and being considered one of the most diverse cities of the Americas.

A foreign investor is willing to open a new store selling sportswear in Buenos Aires. It has stores in the main cities of the world and has no greater knowledge of the mentioned city. For this reason, contact our study that has extensive knowledge in the field of data

science. The result of our advice will be to indicate which are the neighborhoods of Buenos Aires with the greatest potential to open this new branch. This will be an important part of your decision-making process, the other will be the qualitative analysis of the neighborhoods once these data and reports are reviewed and studied. The neighborhoods to be considered will be those in which there are public and private spaces for physical activities and with fewer stores selling sportswear. Foursquare data will be very useful for making decisions based on data on the best of those areas. This database is considered to have updated data.

Buenos Aires is a city with a high population concentrated in a few neighborhoods.
Our client expects to install a new store in one of these neighborhoods considering that they should be close to recreational areas such as sports clubs, gyms, public squares in sports training places in general.
The analysis and recommendations will focus on the neighborhoods with the largest amount of the class of establishments mentioned. Limiting the number of neighborhoods will allow us to do more research, such as finding specific sites for the installation of the new branch.
At present, it is difficult to obtain information that guides investors in this direction.
When we consider all these challenges, in order to identify the candidate neighborhoods, we propose to create a map with the neighborhoods to be evaluated along with an information table that contains statistical information of recreational places indicated by our client. Such as: sports clubs, gyms, public places in sports training places in general.
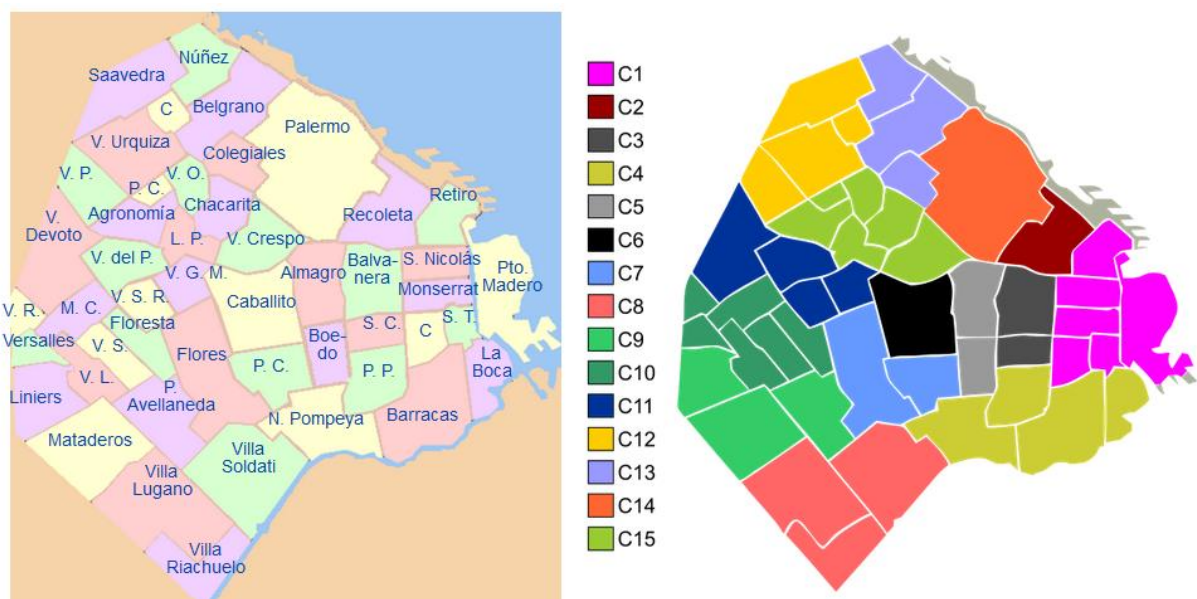
A.2 Data description

- To consider the problem, we can list the data as follows:

- From the Wikipedia I obtained a list of neighborhoods in the City of Buenos Aires.

- Rename columns, I cleaned the data and reduced the city of Buenos Aires, where I used them to create the map of the city's neighborhoods.

- As I could not obtain latitude and longitude of each of the neighborhoods, I used a function that provides this information.

- I used from foursquare.com set API's to obtain recreational places from each neighborhood with their respective frequency. Also, to obtain the amount of sportswear stores already established therein.

B. Methodology

As a database, I used GitHub repository in my study. My master data which has the main components: comuna, name, latitude and longitude of each of the neighborhoods.
Comuna: the city of Buenos Aires is administratively divided into fifteen comunas, administratively divided into fifteen comunas

Data Analysis and Location Data:

- Foursquare location data will be leveraged to explore or compare neighborhoods around Buenos Aires.
- Data manipulation and analysis to derive subsets of the initial data.
- Identifying the high traffic areas using data visualization and statistical analysis.

Visualization:

- Analysis and plotting visualizations.
- Data visualization using various mapping libraries.

	name	categories	lat	lng
0	Feria del Productor al Consumidor	Farmers Market	-34.593981	-58.483098
1	Club Arquitectura	Athletics & Sports	-34.589630	-58.484929
2	Vivero Agronomía	Garden Center	-34.591700	-58.488838
3	Social Parrilla	BBQ Joint	-34.588955	-58.484677
4	Corredor Aeróbico de Agronomía	Trail	-34.592877	-58.483940
5	Parada Línea 80	Bus Stop	-34.592044	-58.489767
6	Túnel Gustavo Cerati	Tunnel	-34.592892	-58.490347

I used python **folium** library to visualize geographic details of Buenos Aires and its boroughs and I created a map of Buenos Aires with boroughs superimposed on top. I used latitude and longitude values to get the visual as below:
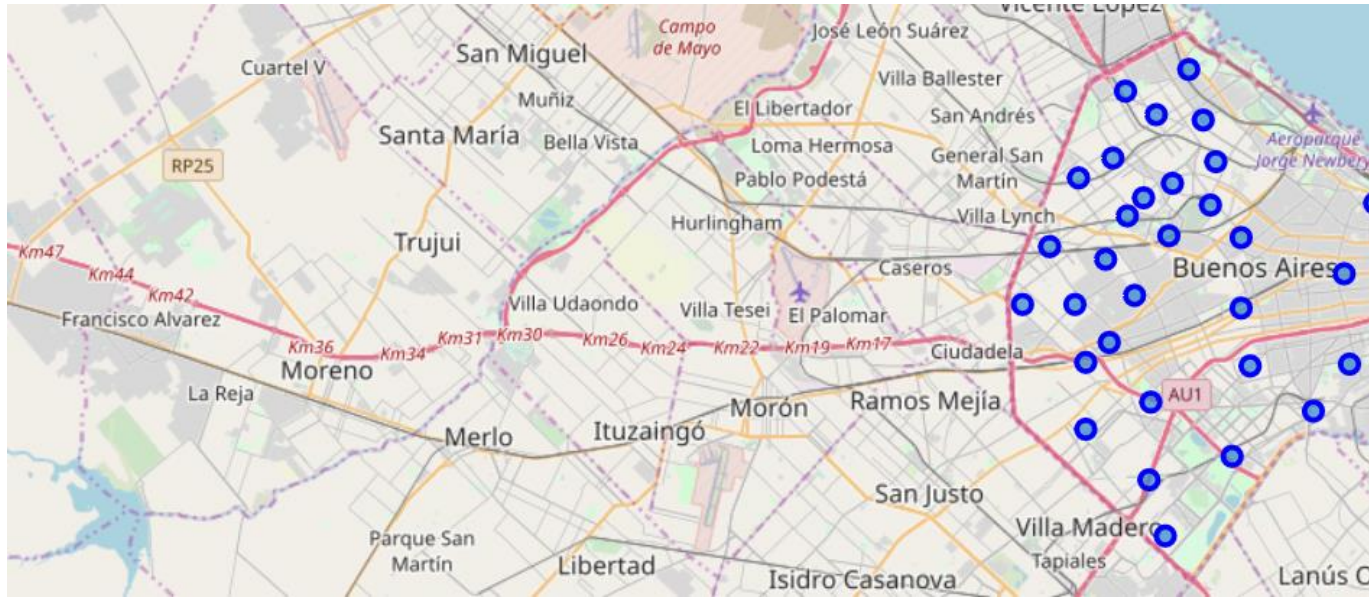
In summary of this graph **158** unique categories were returned by Foursquare, then I created a table which shows list of top 10 venue category for each borough in below table.

Barrio	Barrio Latitud	Barrio Longitud	Venue	Venue Latitud	Venue Longitud	Venue Categoria
Agronomia	7	7	7	7	7	7
Balvanera	14	14	14	14	14	14
Belgrano	42	42	42	42	42	42
Caballito	42	42	42	42	42	42
Chacarita	22	22	22	22	22	22
Coghlan	15	15	15	15	15	15
Colegiales	31	31	31	31	31	31
Constitucion	8	8	8	8	8	8
La Paternal	4	4	4	4	4	4
Mataderos	3	3	3	3	3	3

	Barrio	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Agronomia	Athletics & Sports	Trail	Garden Center	Farmers Market	Bus Stop
1	Balvanera	Fast Food Restaurant	Café	Hotel	BBQ Joint	Gym / Fitness Center
2	Belgrano	Coffee Shop	Bookstore	Argentinian Restaurant	Ice Cream Shop	Café
3	Caballito	Café	Bakery	Pizza Place	Gym	Argentinian Restaurant
4	Chacarita	Pizza Place	Argentinian Restaurant	Bus Stop	Bakery	Coffee Shop
5	Coghlan	Café	Bakery	Pizza Place	Restaurant	Gym / Fitness Center
6	Colegiales	Pizza Place	Argentinian Restaurant	Café	Brewery	Beer Bar
7	Constitucion	Hotel	Plaza	Farmers Market	Bar	Pharmacy

Based on our qualitative data, we suggest that the best locations to open a new branch are places near recreational areas such as sports clubs, gyms, public squares in general sports training places. The analysis and recommendations will focus on the neighborhoods with the highest amount of the class of establishments mentioned. Limiting the number of neighborhoods will allow us to carry out further research, such as finding specific sites for the installation of the new branch.

Taking advantage of the ability of data science, investors can save countless hours of analysis for each neighborhood, consulting many real estate agents with their respective costs and perhaps making a wrong decision. Data will provide better answers and better solutions to their task at hand.

C. Results

The criteria of types of business agreed with the client:

- Soccer Stadium
- Athletics & Sports
- Fitness Center
- Outdoors & Recreation
- Gym
- Plaza
- Sports Clubs
- Tennis Court
- Yoga Studio
- Soccer Field

Frequency distribution for the top 10 venue categories for each neighborhood
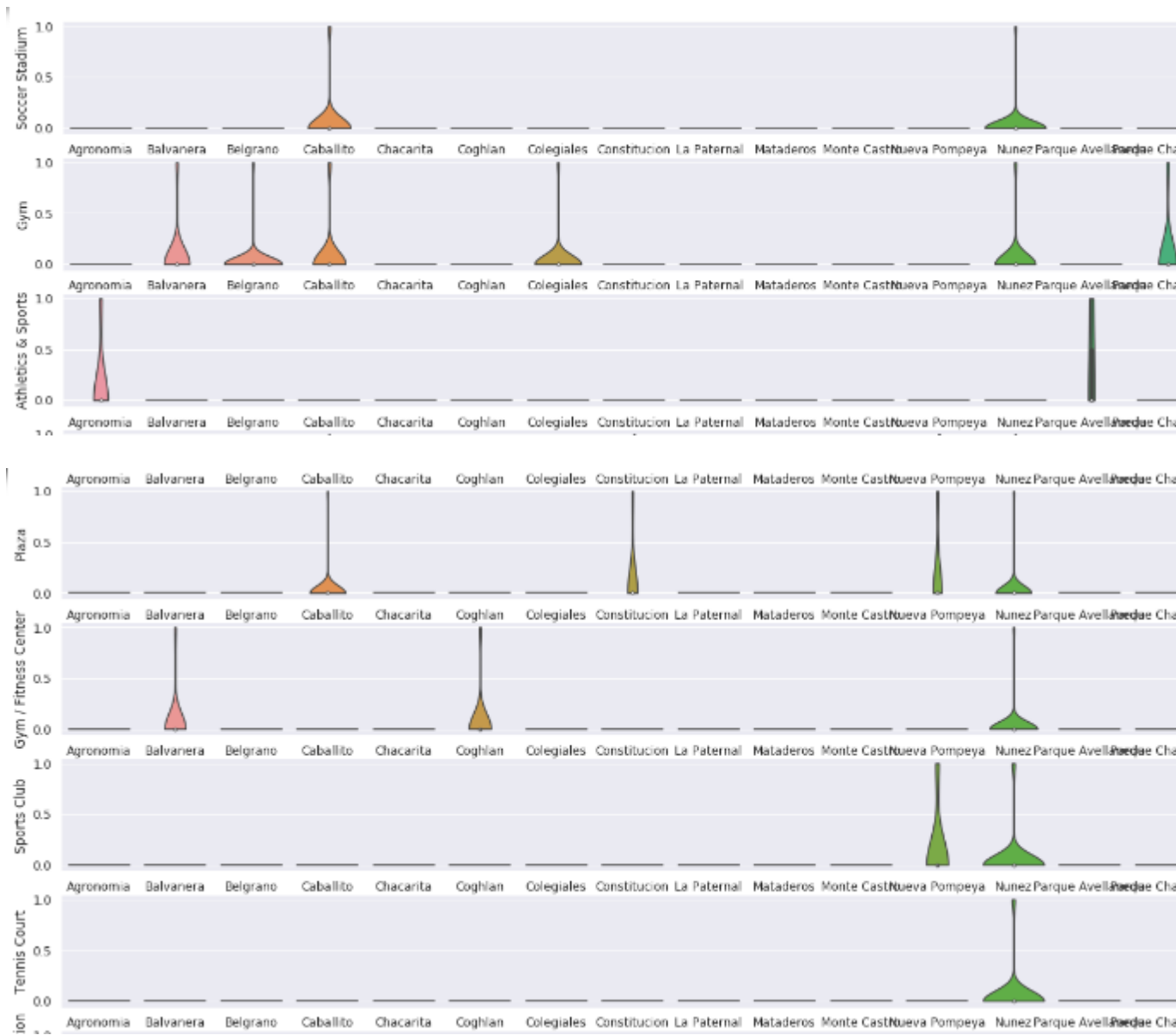
Neighborhoods

So, as we can see in the analysis, there are 6 candidate neighborhoods, Nunez stands out, according to the criteria agreed with our client with great frequency.
They are the following neighborhoods:

- Nunez, is the main candidate
- Villa Real
- Villa Devoto
- Caballito
- Villa Lugano
- Villa Crespo

Let's take this further with some exploration and inferential analysis

We have the 6 neighborhoods that include all the criteria of place category. But if we include the following categories of place:

- Clothing Store
- Sporting Goods Shop

We could make some inferences based on data and knowledge of the domain of marketing and industry, to focus the list. That is, we should analyze whether there are other stores selling sportswear.

Frequency of Clothing and Sporting Goods Shops for each neighborhood

D. Discussion

The inferential analysis that uses the data, as well as knowledge of the domain of retail and marketing, allow the list to focus on only 3 neighborhoods of the 6 selected.
The reasoning is that they have fulfilled most of the requirements presented by our client:

- Soccer stadium
- Athletics & Sports
- Fitness center
- Outdoors & Recreation
- Gym
- Square
- Sports clubs
- Tennis court
- Yoga Studio
- Soccer field

Then, the 3 final candidate neighborhoods to open the new store where most of the requirements are met They are:

- Nunez
- Villa Devoto
- Villa Real

Nunez is the one that meets most of the needs posed by our client.

Let's visualize our neighborhoods on a map of Buenos Aires.

	Comuna	Barrio	Latitud	Longitud
0	Comuna 13	Nunez	-34.545348	-58.462149
1	Comuna 11	Villa Devoto	-34.600994	-58.515516
2	Comuna 10	Villa Real	-34.618943	-58.525877

The three selected neighborhoods are residential with ample spaces for sports and outdoor activities. From this visualization, it is clear that, on a practical level, without data on which to base decisions, the number of neighborhoods to be analyzed is very large and investigating and then visiting them all would be a daunting and slow task.

We have significantly reduced the search area to only 3 that should adapt to our client's business.

We have made inferences from the data when making location recommendations, but that is exactly the point. There is no right or wrong answer or conclusion for the task at hand. The job of data analysis here is to run a course for the selection of new store locations

- to meet the criteria initially established by our client places where sports practices abound.
- Reduce the search to only a few of the main areas that best fit the criteria.

F. Conclusions

There are many ways in which this analysis could have been done based on different methodologies and perhaps different data sources. The method used is a direct way to reduce the options, complying with the initial directives of our client. The analysis and the results is not conclusive, it is a starting point that will guide the next part of the process to find the location of specific stores. The next part will involve knowledge of the domain of the industry, and perhaps, of the city itself. But data analysis and the resulting recommendations have greatly reduced the best data-based options and what we can infer from them.

Without taking advantage of the data to make specific decisions, the process could have been extended and resulted in the opening of a new store in An incorrect area. The data has helped provide a better strategy and a way forward, these data-based decisions will lead to a better solution in the end.

G. References

[1] Buenos Aires – Wikipedia

[2] Foursquare API

[