

Best Gym Location in Buenos Aires using I.A.

~ By Mathias Hochkofler ~



INTRODUCTION & BACKGROUND



I have been acquiring skills related to data science by taking the IBM Data Science Professional Course on Coursera. The last course contains a capstone project. This project is about applying data science toolset and obtained skills to analyze a problem and creating value. My project's theme concerns a topic that I have been really interested in: Gym and health industry. My analysis was performed in Python. The details are pushed to Github.





BUSINESS PROBLEM



In recent years, there is a great boom in the healthy living industry. She is interested to opening a new unit, which will focus on offering her clients a personalized routine according to their weight, age, expectations and time. Considering the financial plan in which the gym will operate, the intention is to find an optimal location in an area of Buenos Aires.



CONSIDER:



NEARBY COMPETITORS

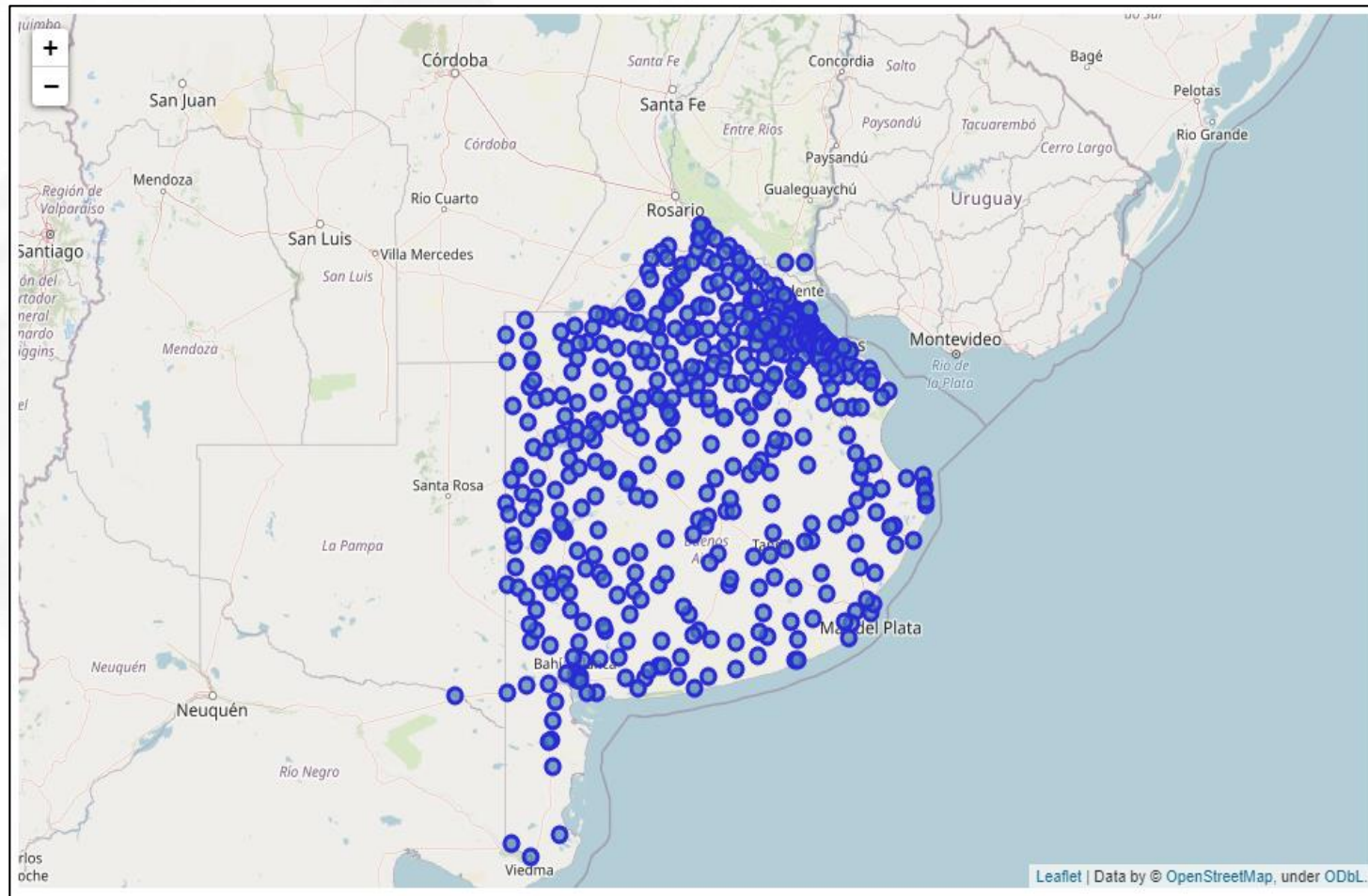


METROPOLITAN AREA

GET DATA



Geonames.org provided us a data set that contains all the postal codes / zip-codes of Argentina, categorized by state and breaking down the coordinates of each point. As shown in the next map:



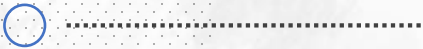
CLEANING DATA



In this way, with a simple filter we got the coordinates to use them as input parameters into Foursquare API and get a response a set of the most relevant venues around each zip-code.

Foursquare provided us 1901 venues related to zip-codes in Buenos Aires which we drop 362 (Foursquare could not find venues related to their respective zip-code) of 548 zip-codes. So, we have 186 zip-code (neiborhoods) with 1539 related venues to analyze.

DATA HANDLING



In the case of study, the input data for the first clustering model are the frequencies of each venue category related to each of the zip-codes.

zip number	Accessories Store	American Restaurant	Argentinian Restaurant	Art Gallery	Art Museum	Asian Restaurant	Athletics & Sports	Auto Garage	Auto Workshop	...	Theme Park	Theme Park Ride / Attraction	Thrift / Vintage Store	Toy / Game Store
1601	0.0	0.0	0.066667	0.0	0.0	0.0	0.066667	0.0	0.0	...	0.0	0.0	0.0	0.0
1602	0.0	0.0	0.000000	0.0	0.0	0.0	0.000000	0.0	0.0	...	0.0	0.0	0.0	0.0
1605	0.0	0.0	0.000000	0.0	0.0	0.0	0.000000	0.0	0.0	...	0.0	0.0	0.0	0.0
1607	0.0	0.0	0.066667	0.0	0.0	0.0	0.066667	0.0	0.0	...	0.0	0.0	0.0	0.0
1609	0.0	0.0	0.000000	0.0	0.0	0.0	0.000000	0.0	0.0	...	0.0	0.0	0.0	0.0

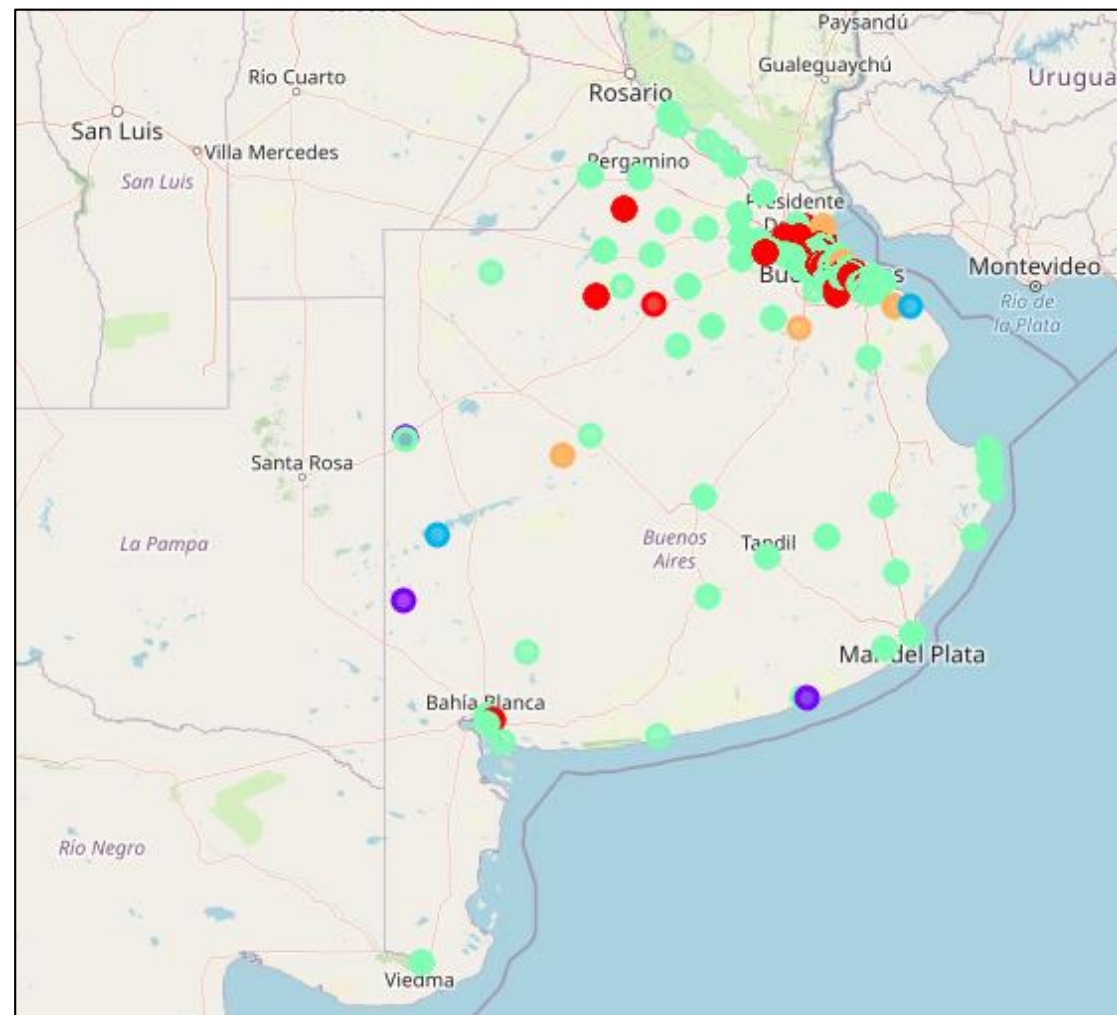
CLUSTER NEIGHBORHOODS BY VENUES FREQUENCY



The model processes the data input and responds by classifying each zip-code according to the frequencies sent.

The result can be seen in the map.

This model is not enough to decide, that is why a second clustering model will be carried out in which the input data will be related in the competition with other gyms.



GET & PREPARE DATA



The data necessary to carry out the model were obtained through the Foursquare API, which we set the central coordinates (latitude and longitude) of Buenos Aires and a base radius (10.000 meters) as input, we get every gym's location inside defined radius.

Comparing the data given (latitude, longitude and rating) with each zip-code, the following measurements were calculated:

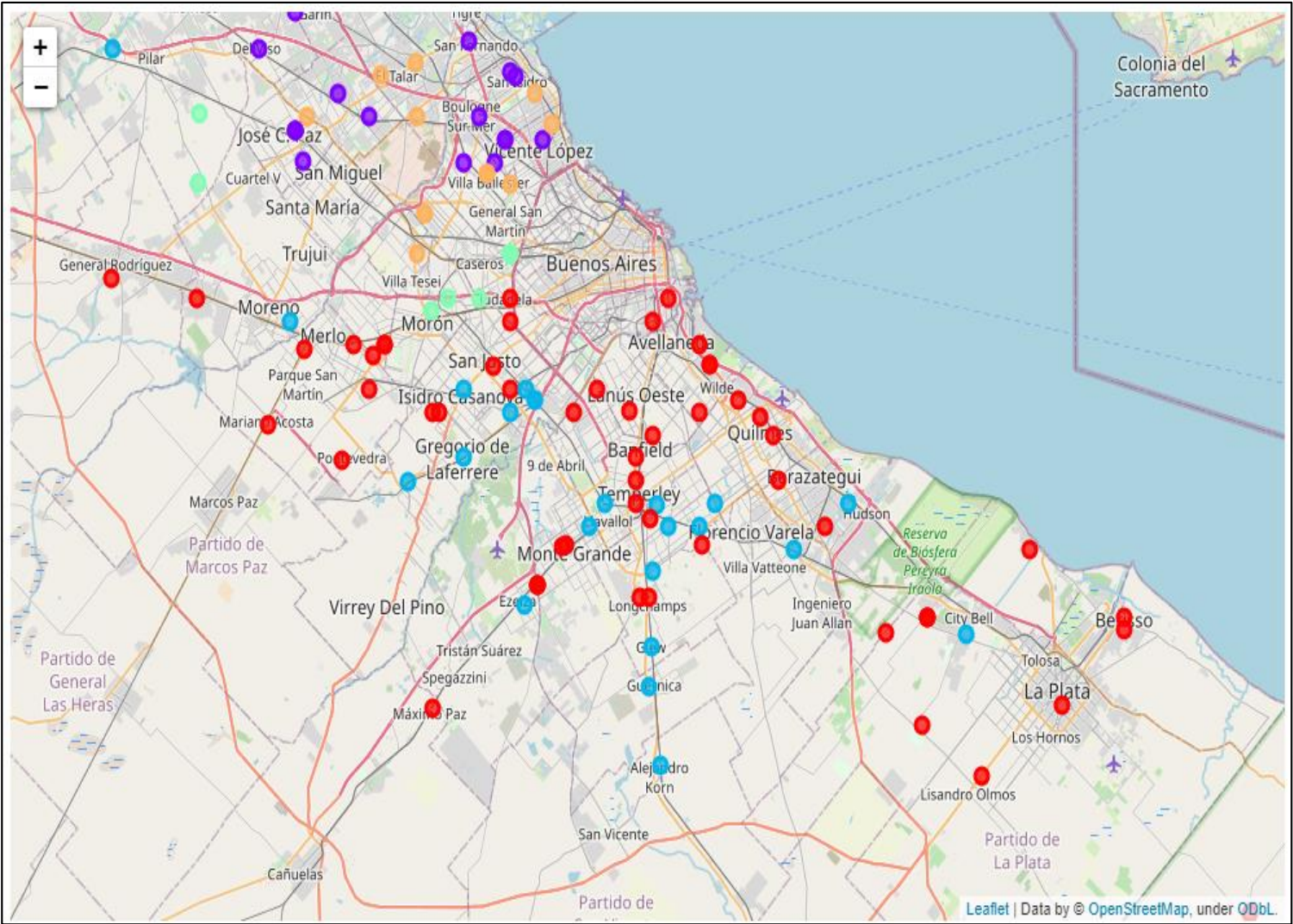
- **Nearby gyms walking:** Gymnasiums less than 2,000 meters to the zip-code
- **Nearby gyms bike:** Gymnasiums at a distance between 2000 and 5000 meters to the zip-code
- **Nearby gyms car:** Gymnasiums at a distance between 5000 and 10000 meters to the zip-code
- **Rating:** social rating 1 to 10 of each gym.

setting the described measures and the result of the previous cluster model into a new cluster model calculate the next table:

	Cluster Labels	nearby gym walking	nearby gym bike	nearby gym car	min gym distance	mean gym distance	mean ratings nearby gyms
0	0	0	7	19	2172.0	6073.538462	5.380769
1	0	0	3	19	3826.0	6976.409091	5.450000
2	0	2	3	20	1275.0	6978.080000	5.396000
3	0	0	7	19	2172.0	6073.538462	5.380769
4	0	3	9	15	802.0	4906.407407	5.366667

CLUSTER BY NEARBY GYMS

○



● Cluster 0:

● Cluster 1:

● Cluster 2:

● Cluster 3:

● Cluster 4:

CLUSTERS 0 AND 2



The main characteristic is that they do not have gyms near at short, medium or long distance. That is why making an investment in these points is a great risk because there is no certainty that there is a potential demand in those areas.

CLUSTER 1



This group is characterized by being several gymnasiums at a medium and long distance, being more precise, the minimum distance to a gym is 1.800 meters and an average distance is 6,000 meters. This group represents a good market opportunity, although there is strong competition over long distances.

CLUSTER 3



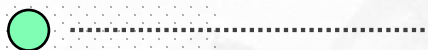
It is located on the edges of the growing metropolitan area; It is characterized by having no offer at short-medium distance and very low offer at long distance, on average it has 3 nearby gyms at of 8,300 meters and a the nearest gym is 7,700 meters away. So that It is close to demand but with enough room to grow.

CLUSTERS 4



It is characterized by being a small group with a high density of gyms, having an average of 5 gyms at short distance, 9 gyms at medium distance and 11.6 gyms at long distances. Given its high level of supply, investing in this group is risky due to the high level of competition.

BEST OPTION



In order to maximize opportunities with the least possible risk, the study indicates that the best option is in Cluster 3, with the following addresses:

- Club Newman, Benavídez, Partido de Tigre, Buenos Aires, 1621, Argentina
- Barrio Santa Isabel, Ingeniero Maschwitz, Partido de Escobar, Buenos Aires, B1623, Argentina
- Presidente Derqui, Partido del Pilar, Buenos Aires, 1635, Argentina
- Isla Nazar Anchorena, Primera Sección, Partido de Tigre, Buenos Aires, B1644BHH, Argentina
- 879, 411 - Beazley, Sáenz Peña, Partido de Tres de Febrero, Buenos Aires, B1674AVJ, Argentina
- Luis Antonio Beruti, Morón, Partido de Morón, Buenos Aires, B1708KCH, Argentina
- Boca Ratón Golf Club, Pilar Sur, Partido del Pilar, Buenos Aires, Argentina

CONCLUSION



Purpose of this project was to identify Buenos Aires areas close to center with low number of gyms in order to aid stakeholders in narrowing down the search for optimal location for a Gym and fitness center. By calculating Gym density distribution from Foursquare data, we have first identified general boroughs that justify further analysis, and then generated extensive collection of locations which satisfy some basic requirements regarding existing nearby gyms. Clustering of those locations was then performed in order to create major zones of interest (containing greatest number of potential locations).

Final decision on optimal gym location will be made by stakeholders based on specific characteristics of neighborhoods and locations in every recommended zone, taking into consideration additional factors like attractiveness of each location (proximity to park or water), levels of noise / proximity to major roads , real estate availability, prices, social and economic dynamics of every neighborhood etc.

CONCLUSION

..... ○

Final decision on optimal gym location will be made by stakeholders based on specific characteristics of neighborhoods and locations in every recommended zone, taking into consideration additional factors like attractiveness of each location (proximity to park or water), levels of noise / proximity to major roads , real estate availability, prices, social and economic dynamics of every neighborhood etc.

A stylized, handwritten signature in blue ink, appearing to read 'Signature'.

MATHIAS HOCHKOFER



THANKS FOR YOUR ATTENTION

-EVER Template

