André Altoé

Capstone Project
Supplements store in Atlanta

Brasil

August, 2020

1 Introduction

1.1 Background

With the increasing number of people who cares about the health quality and performance in sports, comes the necessity of a better alimenting and supplementation in some cases. This project tries to find the best neighborhood to place a new store that covers more people around the best gyms in the city.

1.2 Problem and interest

The problem is focused in creating a new store, it can be interesting for a entrepreneur that wants the best neighborhood possible for the maximum profit, it would be very convenient for those who train or live close to this store.

2 Data acquisition and cleaning

2.1 Data gathered

- Venues location and rating (closer gyms) from Foursquare API.
- Demographic statistics from Wikipedia.
- Neighborhood coordinates from Bing maps API

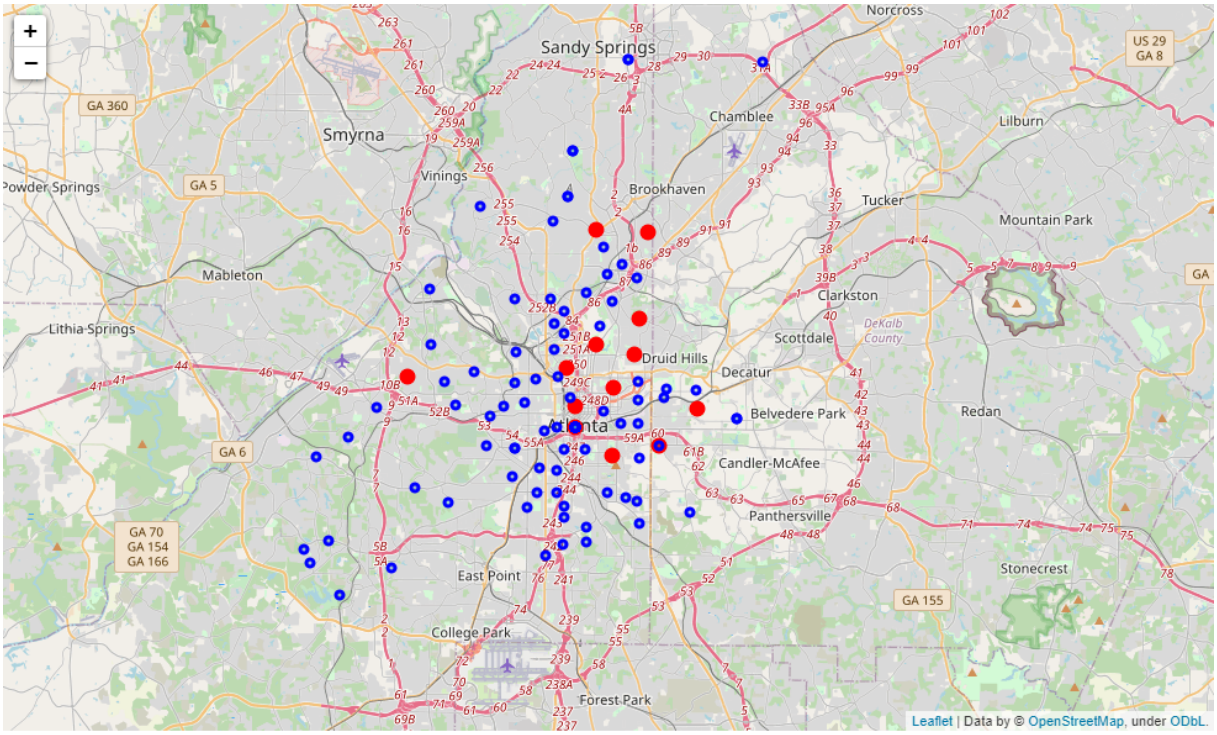
2.2 Cleaning the data

After gathering all of the necessary data, the tables were joined to show us the Neighborhood location, GYMs mean rating and population, it was necessary to understand the data and know what to do next. Unfortunately, because Foursquare is not very common today there was not so much data to gather, and the majority of the gyms rating is just the mean value.

3 Data Analysis

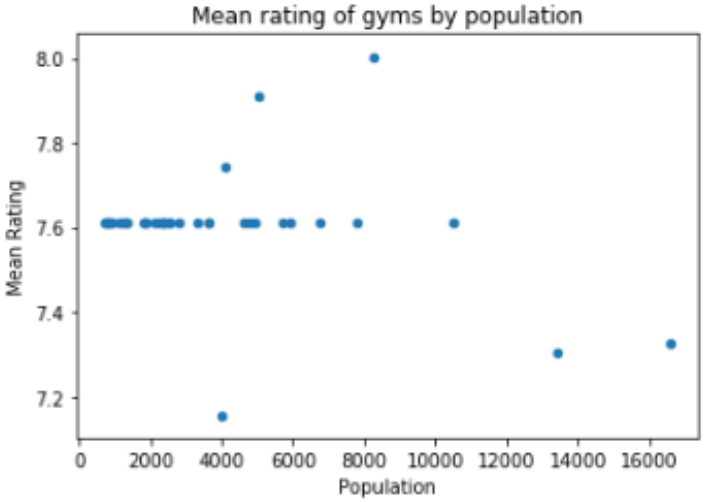
3.1 Graphics and maps

After gathering the first data about the places and location, I first made a folium map to see about the distance from one to another neighborhood and see how many does have more than 5000 residents. That was the result:



In red neighborhoods with more than 5000 residents

Unfortunately, because Foursquare is not very common today there was not so much data to gather, and making a scatter plot, that was the result.



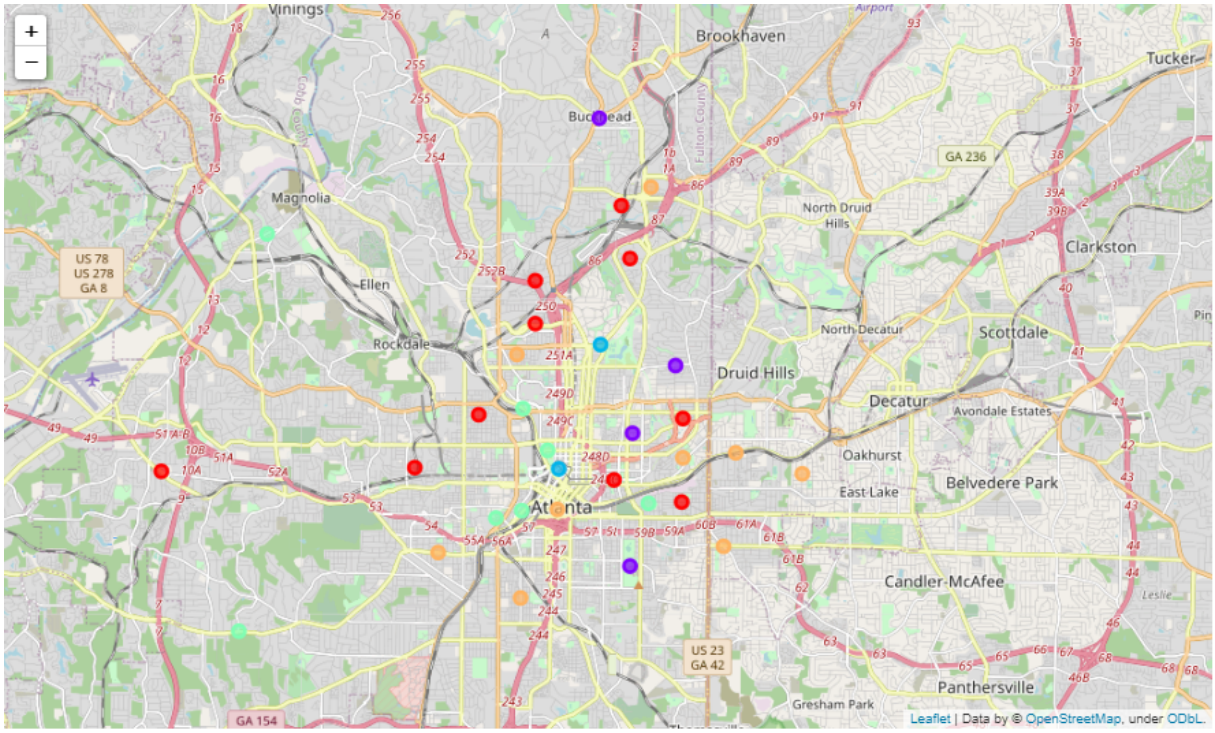
As we can see there is just one best value and it is from North Buckhead with a rating of 8,004 and a decent amount of population above the mean.

4 Clustering

Then, I wanted to cluster the neighborhoods to see which ones would be the best and their cluster differences. The final results gave me this table:

	Neighborhood	Cluster Labels	rating	Population (2010)
26	North Buckhead	1	8.004444	8270
12	East Atlanta	4	7.910000	5033
21	Inman Park	4	7.745333	4098
0	Adamsville	0	7.613333	2403
1	Arlington Estates	3	7.613333	776
22	Kirkwood	4	7.613333	5897
24	Marietta Street Artery	3	7.613333	745
27	Old Fourth Ward	1	7.613333	10505
28	Peachtree Heights West	4	7.613333	4767
29	Peachtree Hills	0	7.613333	2785
30	Piedmont Heights	0	7.613333	2323
31	Pittsburgh	4	7.613333	3658
32	Poncey-Highland	0	7.613333	2133
33	Reynoldstown	0	7.613333	2550
34	Riverside	3	7.613333	1341
35	Sweet Auburn	0	7.613333	1882
36	The Villages at Castleberry Hill	3	7.613333	864
20	Hunter Hills	0	7.613333	2223
19	Home Park	4	7.613333	4941
18	Heritage Valley	3	7.613333	910
9	Castleberry Hill	3	7.613333	1285
2	Atlanta University Center	4	7.613333	5703
4	Audubon Forest	3	7.613333	813
6	Brookwood	0	7.613333	1834
8	Cascade Heights	3	7.613333	1124
37	Virginia-Highland	1	7.613333	7800
14	English Avenue	0	7.613333	3309
15	Grant Park	1	7.613333	6771
16	Greenbriar	4	7.613333	3628
17	Hammond Park	0	7.613333	2565
7	Cabbagetown	3	7.613333	1247
23	Lindbergh	4	7.613333	4598
5	Baker Hills	3	7.613333	757
3	Atlantic Station	0	7.613333	1888
10	Chastain Park	0	7.613333	2398
25	Midtown	2	7.328000	16569
11	Downtown	2	7.304444	13411
13	Edgewood	4	7.156667	3983

It is possible to see that the best cluster is the number 1 that has a medium population and the best ratings, also the cluster 2 was not appropriate because the rating was low, even with more population. Here is the map:



Purple: Cluster 1

Red: Cluster 0

Orange: Cluster 4

Blue: Cluster 2

Green: Cluster 3

5 Conclusions

5.1 Conclusion

Unfortunately, the foursquare API is old and is not used by so many people anymore, and there is not so much information and nothing up to date, but, for learning purposes it is interesting and gives some idea about the project, the best place to put the store with those data would be North Buckhead or the places within the Cluster 1

5.2 Future Directions

To improve this model it would be necessary some recent data and more data, much more, also, it would be interesting with the price data, that was planned to be used but was not possible because foursquare did not returned so much about it.