The Battle of Neighborhoods - Week 2

Érika Soares Machado

07/06/2020

A. Introduction

A.1. Description and Discussion of the Background

Goiânia is a Brazilian city, capital of Goiás, where more than 1.5 million people live. It is an important economic center of Midwest region of Brazil. It is considered as a strategic center to areas such as industry, medicine, fashion, and agriculture. As a resident of this city, I decided to use Goiânia in my project. Currently, Goiânia has 641 boroughs, but we will work with only 210 of them. It'll be explained why later.

There are many food places in Goiânia like fast food restaurants, bakery, pizza place, sandwich place, Café, steakhouse, seafood restaurant, and others. We can also see specific Brazilian food places such as Açaí House, Pastelaria, goiano restaurant, mineiro restaurant, tapiocaria, empada house, pamonharia, and others. A specific kind of restaurant is emerging recently with great demand, Vegan restaurants. When we think of it by the investor, we expect from them to prefer the boroughs where there is a lower density of business they want to install. If an investor wants to open a Vegan restaurant in Goiânia, the best location needs to be identified. However, some information that will guide an investor in the right direction are difficult to obtain nowadays.

Following some criteria, we can create a map and guide them to find places on Goiânia that are the most promising to begin a new Vegan establishment.

A.2 Data Description

Based on the definition of our problem, the factors that will influence our decision are:

- number of existing restaurants in the neighborhood (any type of restaurant)
- number of Vegan restaurants in the neighborhood, if any
- similarity between neighborhoods
- a distance of neighborhood from the city center

Following data sources will be needed to extract/generate the required information:

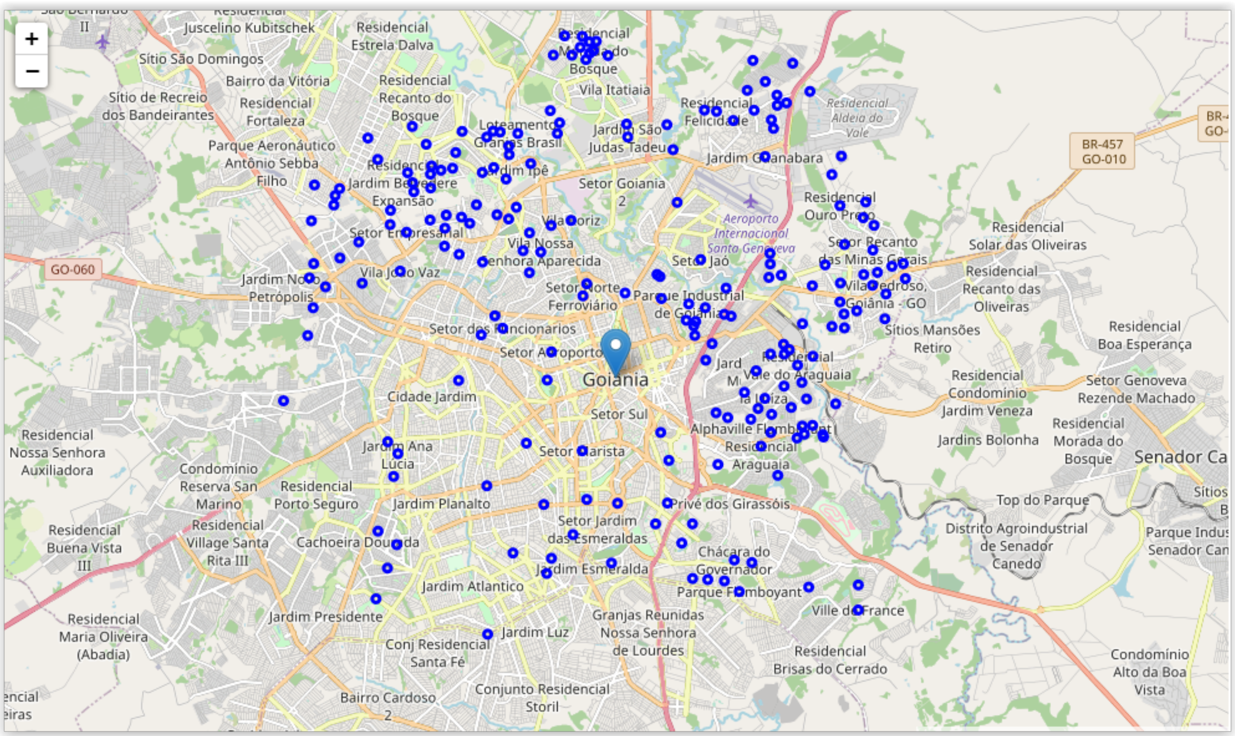
- Goiânia borough data web scraped from Open Street Map using selenium package
- Centers of candidate areas will be generated algorithmically using Geopandas package
- Number of restaurants and their type and location in every neighborhood will be obtained using Foursquare API
- Coordinate of Goiânia center will be obtained using Geopandas package

B. Methodology

As a database, I used the Github repository in my study. My initial data has the main components Local name, Distance, Lat, Long (the lat and long data are with values switched, but this is corrected later). The column Distance is the distance between the center city and the center of the borough.

	Local name	Distance	Lat	Long
0	Aldeia do Vale	8.7	-49.199667	-16.605009
1	Alphaville Flamboyant Residencial Araguaia	3.8	-49.213159	-16.698653
2	Bairro Capuava	6.6	-49.323221	-16.655652
3	Bairro Feliz	3.0	-49.228554	-16.662100
4	Bairro Santo Hilário	6.1	-49.195476	-16.650797

All maps plotted were constructed with folium library. As I said before, Goiânia has 641 boroughs, but I worked with only those are 10km near to the center city, so the number os boroughs reduced to 210. A map with the center boroughs was created to visualize the initial data and it's showed below:



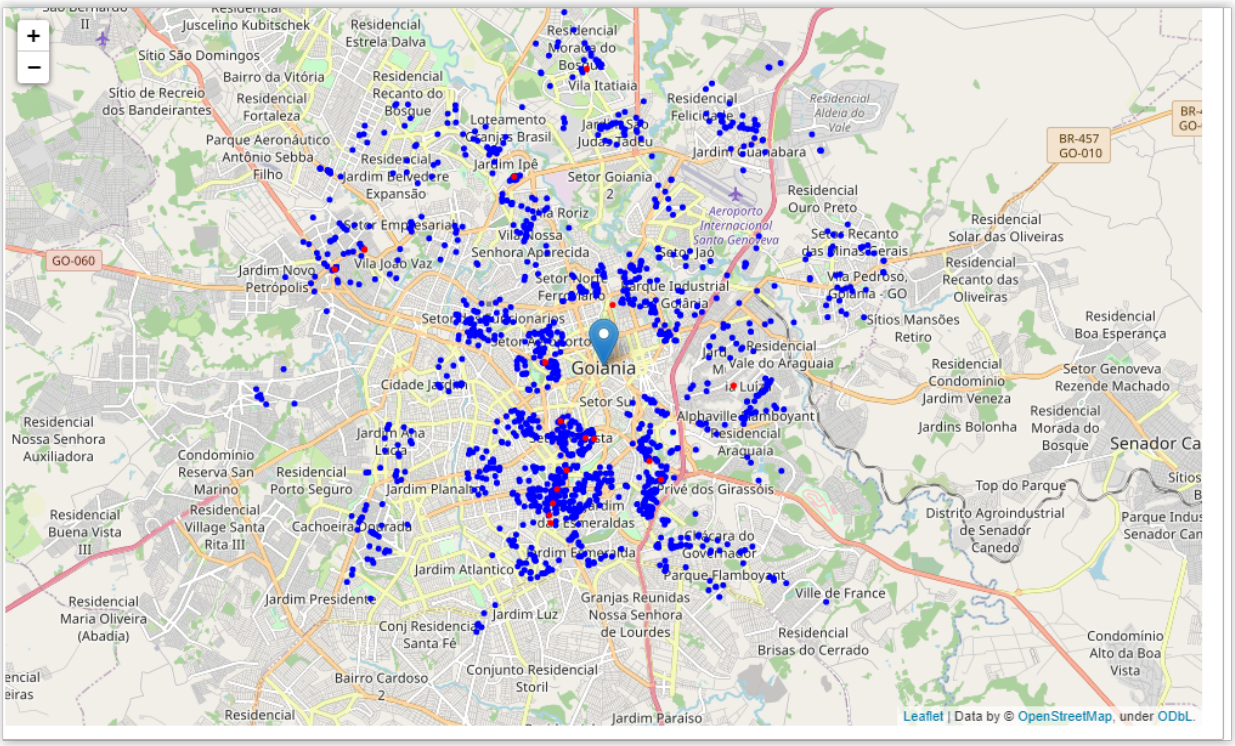
I utilized the Foursquare API to explore the boroughs and segment them. I designed the limit as 100 venues, the radius 700 meters for each borough from the venue, and I select the category food places. You can see the first 5 rows of the obtained data bellow :

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Alphaville Flamboyant Residencial Araguaia	-16.698653	-49.213159	Megaburger's	-16.694582	-49.213653	Fast Food Restaurant
1	Alphaville Flamboyant Residencial Araguaia	-16.698653	-49.213159	Panificadora Skina do Pão	-16.692809	-49.212366	Bakery
2	Alphaville Flamboyant Residencial Araguaia	-16.698653	-49.213159	Pizza 10	-16.693148	-49.212926	Pizza Place
3	Alphaville Flamboyant Residencial Araguaia	-16.698653	-49.213159	Pit Dog XGordo	-16.696231	-49.213708	Burger Joint
4	Alphaville Flamboyant Residencial Araguaia	-16.698653	-49.213159	Restaurante e Pizzaria Tan Tan	-16.693422	-49.211480	Brazilian Restaurant

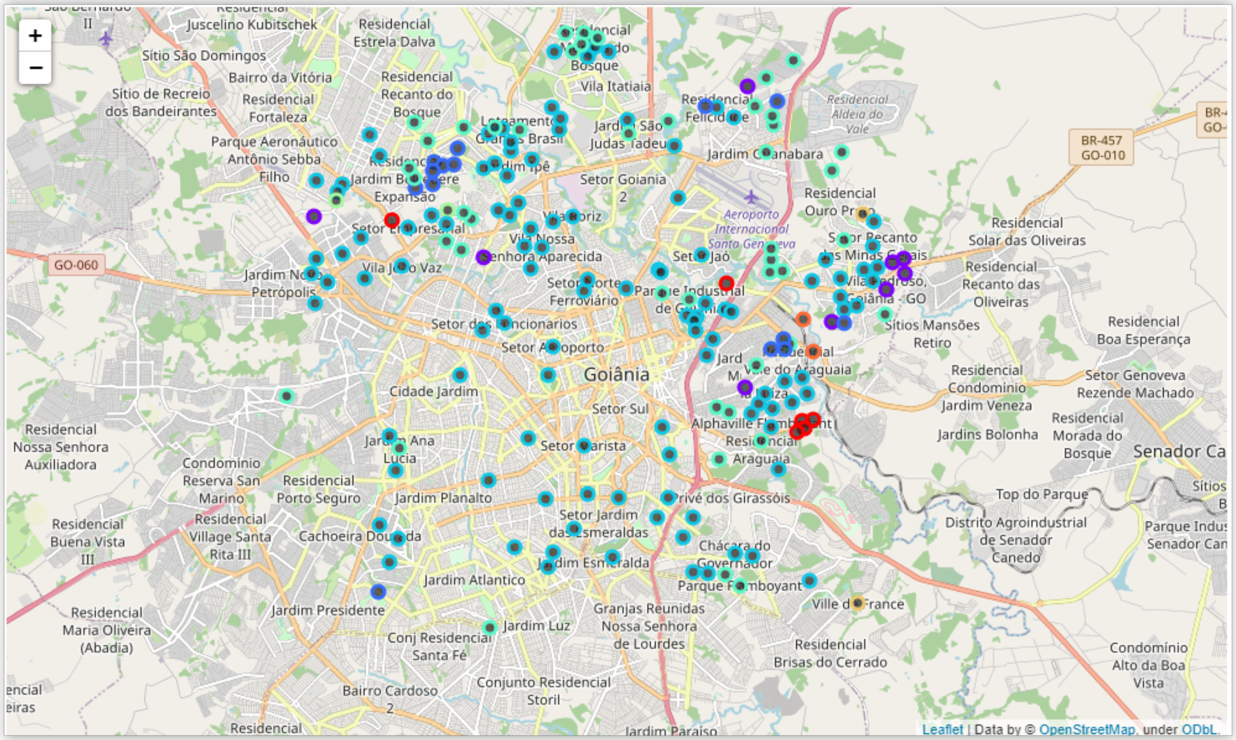
In summary of this data 2482 venues were returned by Foursquare with 66 unique categories. Here is a top 10 neighborhoods and frequency venues table with decrease values. We can notice that none of the boroughs reached the 100 limits of venues that I searched in Foursquare API. The result doesn't mean that inquiry run all the possible results in boroughs. Actually, the Foursquare app is not widely used in Brazil, so tha data couldn't match with the real data, this could be one of the causes. Also, it was used one borough coordinate, maybe using more coordinates from the boroughs, the number will increase. Or, another explanation is that the boroughs have few restaurants.

After that, I used the Foursquare API again to search only for the 'Vegetarian / Vegan restaurant' category. I did this to use a bigger ratio since the density of this restaurant kind is lower, so I try to get more information possible in a ratio of 1km. I found 25 Vegan venues.

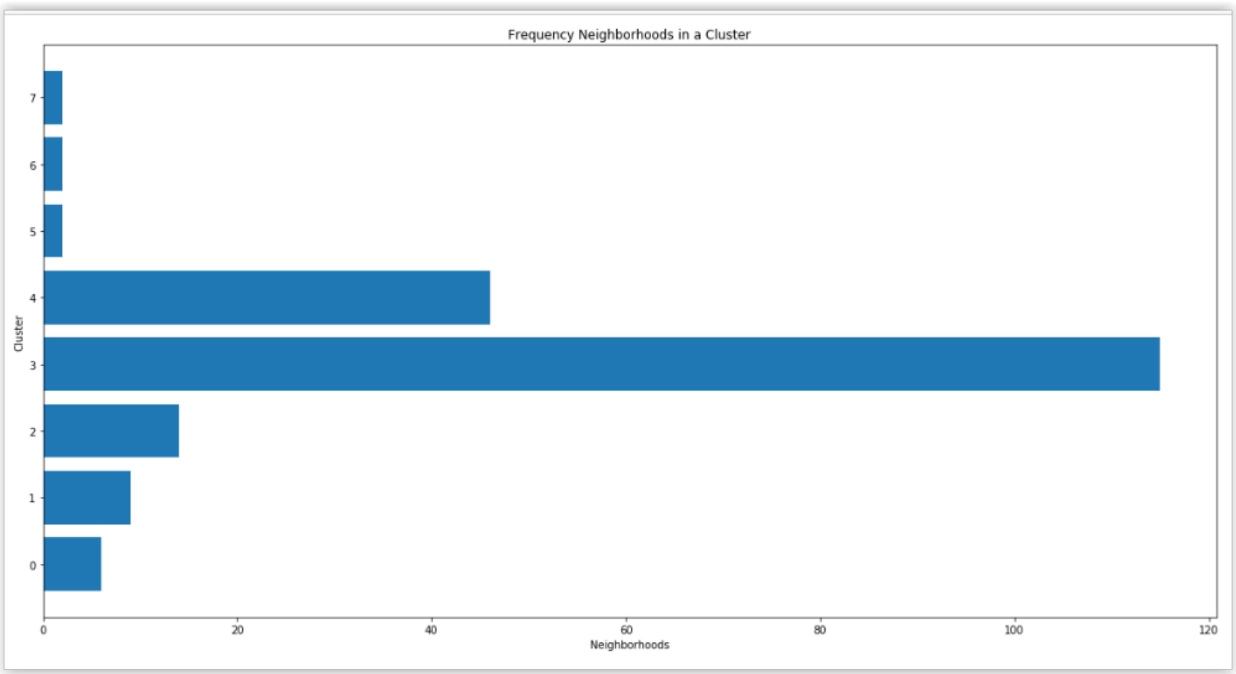
The map below shows us the venue data researched when the red points are the Vegan restaurants and the blue ones are from other kinds:



We can see that there is a concentration of restaurants in the south zone. To calculate the similarity of the boroughs I used K-means algorithm, which is one of the simplest and popular unsupervised machine learning algorithms for clusterization. A cluster refers to a collection of data points aggregated together because of certain similarities. First, I used the Elbow Method to find the optimum k of the K-Means. But it didn't work very well in this case. So I decided to use k = 8, and the result was as shown below:



The cluster with most neighborhoods is number 3



Second I labeled the Vegan data to find in which cluster the Vegan restaurants mostly are in it. As expected, there are more Vegan Restaurants in cluster 3. That way, we will choose the Neighborhoods in cluster 3 to continue our discussion, because we can conclude that the neighborhoods in that are more likely to accept a new Vegan restaurant due to the similarity with the neighborhoods that there already is this kind of establishment.

Then I reduced our data to only neighborhoods in cluster 3 and calculate the restaurants' density a distance

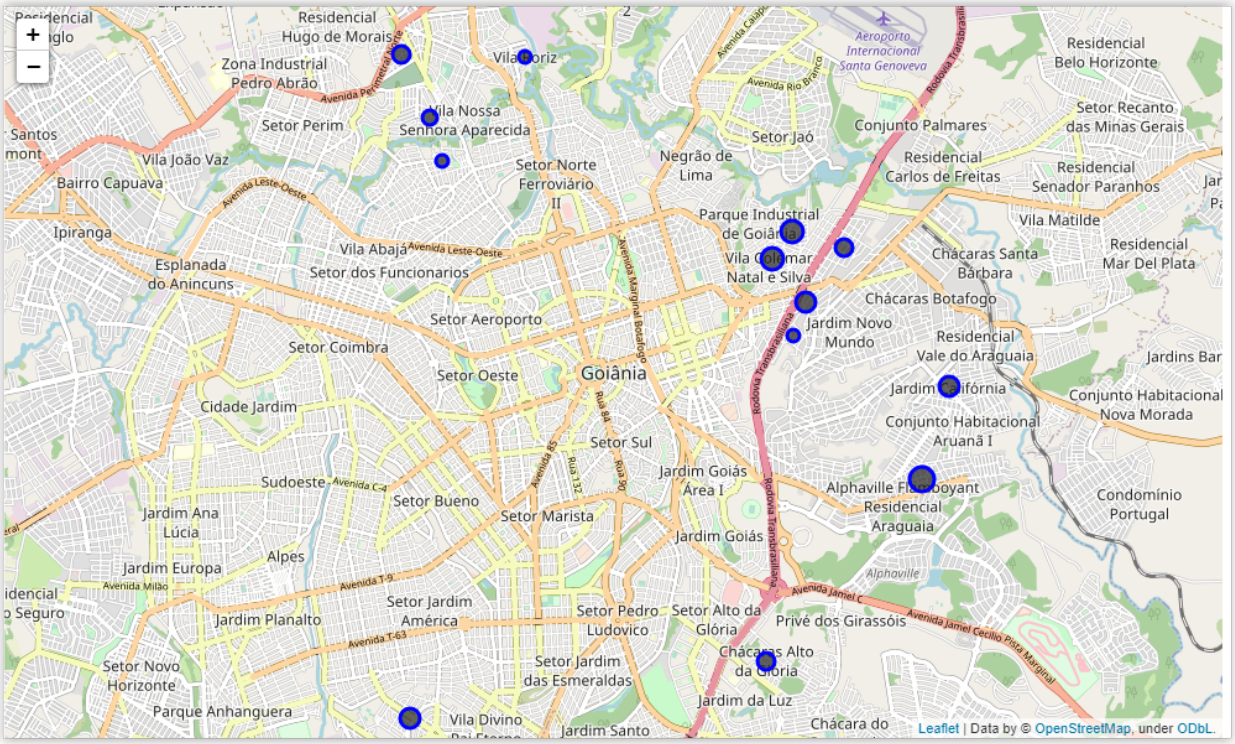
to the city center. This resulted in 113 boroughs to work with.

C. Results

After that, it was calculated the distance between the center of each borough and the venues. I did this because with the distance it was possible to do the restaurant frequency in a determined ratio. The criteria used for the borough selection was:

- Restaurants frequency between 1º quantile and 2º quantile of the distribution frequency
- Vegan restaurants frequency < 3 Vegan restaurants in 1km meters near the neighborhood
- Distance < 6km to center city

It resulted in 13 promising area



This concludes my analysis. We have 13 addresses representing the center of zones containing locations with a low number of restaurants and with no Vegan restaurants nearby, all zones being fairly close to the city center (all less than 6 km). Most of them are located in peripheral neighborhoods.

D. Discussion

My analysis shows that there is a great number of restaurants in Goiânia (~2400 in our initial area of interest which was 10x10km around the center city). Highest concentration of restaurants was detected South Zone that we can consider as Goiânia's gastronomic pole.

After identifying the number of restaurants in the whole city, we decided to devise the borough in clusters (k = 8) to identify the similarity between the neighborhoods. With the data clustered it was possible to recognize which cluster had more vegan restaurants. It was cluster 3 than the neighborhoods in there were classified as those ones with high similarity in terms of restaurants. In this way, we see that the new restaurant will be more successful in one of these boroughs. In this cluster, there are 113 boroughs.

After directing our attention to this group, we first created a dense grid of restaurants per borough (amount of restaurants in a radius of 600 m from the borough center) and also a dense grid of vegan restaurants per borough (amount of vegan restaurants in a radius of 1 km from the borough center). With this information, we used 3 criteria to filter the data:

- Restaurants frequency between 1º quantile and 2º quantile: We believe that those locations that have few or no one restaurant are too risky to begin this kind of restaurant because it has a very specific public. On the other hand, The localities that there are a large number of restaurants (we consider large amount more than the median) are risky because the competitiveness is high and it may compromise the new establishment.
- Vegan restaurant frequency has to be less than 3 Vegan restaurants in 1km meters near the neighborhood.
- The distance to the center city has to be less than 6km.

Those criteria resulted in 13 zones with the largest potential to allocate the new restaurant. But it's important to examine deeper these resulted positions. It's essential to research about security if there are social places nearby and other economic indicators. We also have 2 recommendations for the start, each one based on their owns characteristics:

- **Jardim Brasil:** It's near to government agencies, leisure places (Flamboyant park, Oscar Niemeyer, Flamboyant Shopping Mall, Serra Dourada Stadium, Goiânia Arena), and it is near to luxury condominiums.
- **Parque Amazonia:** It's in the same area as the gastronomic hub (South zone), but has the advantage of not having so many restaurants in the region.

E. Conclusion

The purpose of this project was to identify Goiânia areas close to center with low number of restaurants (particularly Vegan restaurants) in order to aid stakeholders in narrowing down the search for the optimal location for a new Vegan restaurant. But in this case that we are working with Vegan restaurant it's also necessary to review the criteria used, because common restaurants may not be considered direct competitors to Vegan restaurants.

We recommend 13 locations, in specialty 2, but the final decision will be made by stakeholders based on specific characteristics of neighborhoods and locations in every recommended zone taking into consideration additional factors like security, the 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.