Predicting the best place to open a vegan restaurant

Danilo Pedrelli

May 16, 2019

1 Introduction: Business problem

In this project we will try to find an optimal location for a restaurant. Specifically, this report will be targeted to stakeholders interested in opening a vegan restaurant in Sao Paulo, Brazil.

Since there are lots of restaurants in Sao Paulo we will try to detect locations that are not already crowded with restaurants. We are also particularly interested in areas with no vegan restaurants in vicinity. We would also prefer locations as near as possible to the Bixiga neighborhood, which is famous for great restaurants, where many people go there just to eat.

We will use our data science powers to generate a few most promising neighborhoods based on this criteria. Advantages of each area will then be clearly expressed so that best possible final location can be chosen by stakeholders.

2 Data

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

- number of existing restaurants in the neighborhood (any type of restaurant)
- number of and distance to vegan restaurants in the neighborhood, if any distance of neighborhood from Bixiga area

We decided to use regularly spaced grid of locations, centered around city center, to define our neighborhoods.

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

- centers of candidate areas will be generated algorithmically and approximate addresses of centers of those areas will be obtained using Google Maps API reverse geocoding;

- number of restaurants and their type and location in every neighborhood will be obtained using Foursquare API coordinate of the Bixiga neighborhood in Sao Paulo will be obtained using Google Maps API geocoding.

3 Methodology

In this project we will direct our efforts on detecting areas of Sao Paulo that have an average density of restaurants that is not so big, particularly those with low number of vegan restaurants. We will limit our analysis to area ~6km around the Bixiga neighborhood.

In first step we have collected the required data: location and type (category) of every restaurant within 6km from Bixiga. We have also identified vegan restaurants (according to Foursquare categorization).

Second step in our analysis will be calculation and exploration of 'restaurant density' across different areas of Sao Paulo - we will use heatmaps to identify a few promising areas close to Bixiga with low number of restaurants in general (and no vegan restaurants in vicinity) and focus our attention on those areas.

In third and final step we will focus on most promising areas and within those create clusters of locations that meet some basic requirements established in discussion with stakeholders: we will take into consideration locations with no more than two restaurants in radius of 250 meters, and we want locations without vegan restaurants in radius of 400 meters. We will present map of all such locations but also create clusters (using k-means clustering) of those locations to identify general zones / neighborhoods / addresses which should be a starting point for final 'street level' exploration and search for optimal venue location by stakeholders.

4 Analysis

To do the analysis, first we investigate the quantity of restaurants in every area candidate, which is around 10 within a radius of 300 m. Then we calculate the distance to nearest vegan restaurant from every area candidate center (not only those within 300 m - we want distance to closest one, regardless of how distant it is). On average, vegan restaurants can be found within $\backsim$ 5000 m from every area center candidate.

Let's create a map showing heatmap/density of restaurants and try to extract some meaningfull info from that. Also, let's show few circles indicating distance of 1km, 2km and 3km from Bixiga.

Figure 1: The heatmap/density of all the restaurants.

Let's create another heatmap map showing heatmap/density of vegan restaurants only:

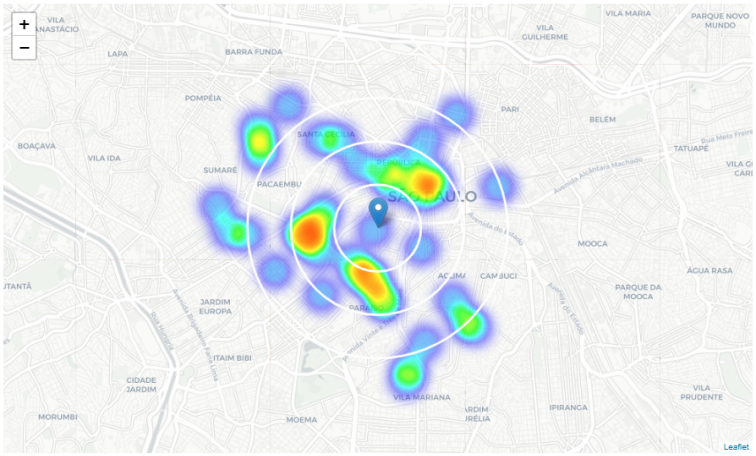

Figure 2: The heatmap/density of the vegan restaurants.

This map is not so 'hot', but it indicates higher density of existing vegan restaurants directly northeast and southwest from Bixiga center, with closest pockets of low vegan restaurant density positioned on the remaining areas.

Based on this we will now focus our analysis on areas south-west, south, south-east and east from Berlin center - we will move the center of our area of interest and reduce it's size to have a radius of 2.5km. This places our location candidates mostly around the Paulista Avenue, where a lot people walk around, including tourists.

The Paulista Avenue

Analysis of popular travel guides and web sites often mention the Paulista Avenue as interesting, rich with culture, and different people selling their art. Below is an overview of it given by Wikipedia [1]:

"Paulista Avenue (Avenida Paulista in Portuguese, Paulista being the demonym for those born in São Paulo state) is one of the most important avenues in São Paulo, Brazil. It stretches 2.8 kilometres (1.7 mi) and runs northwest to southeast. Its northwest point is Praça Marechal Cordeiro de Farias at its intersection with Rua da Consolação and its southeast point is Praça Oswaldo Cruz at its intersection with Treze de Maio, Bernardino de Campos, Desembargador Eliseu Guilherme, and Dr. Rafael de Barros avenues. Major crossroads on the street are Rua Augusta, Rua Haddock Lobo and Avenida Brigadeiro Luis Antonio. Parallel to it are Cincinato Braga, Joaquim Eugenio de Lima on the Bela Vista/Paraíso side and Alameda Santos and Coronel Oscar Freire on the Jardins side. Paulista Avenue crosses sections of the neighborhoods of Paraíso, Bela Vista, Jardim Paulista, Cerqueira César and Jardim América, ending in Higienópolis.

The headquarters of a large number of financial and cultural institutions are located on Paulista Avenue. As a symbol of the center of economic and political power of São Paulo it has been the focal point of numerous political protests beginning in 1929 and continuing into the 21st century. It is also home to an extensive shopping area and to South America's most comprehensive fine-art museum, MASP. Being one of the highest points in São Paulo, it is clustered with radio and TV stations antennae, most notably that of Rede Gazeta. Paulista Avenue is a major hub of the subway and bus lines of the city."

Popular with tourists, alternative and bohemian but booming and trendy, relatively close to city center and well connected, ths region appears to justify further analysis.

We define new, more narrow region of interest, which will include low-restaurant-count parts around Paulista Avenue closest to Bixiga.

Now let's calculate two most important things for each location candidate: number of restaurants in vicinity (we'll use radius of 250 meters) and distance to closest vegan restaurant:

[1] https://en.wikipedia.org/wiki/Paulista_Avenue

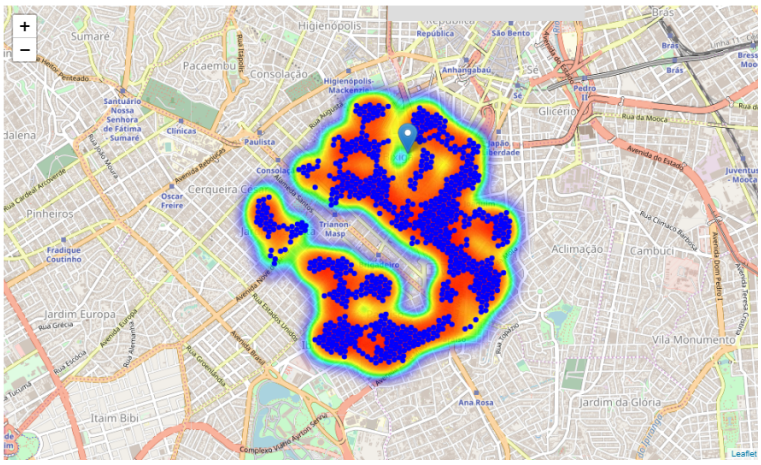

Figure 3: Zones with no vegan restaurants nearby.

Looking good. What we have now is a clear indication of zones with low number of restaurants in vicinity, and no vegan restaurants at all nearby.

Let us now cluster those locations to create centers of zones containing good locations. Those zones, their centers and addresses will be the final result of our analysis.

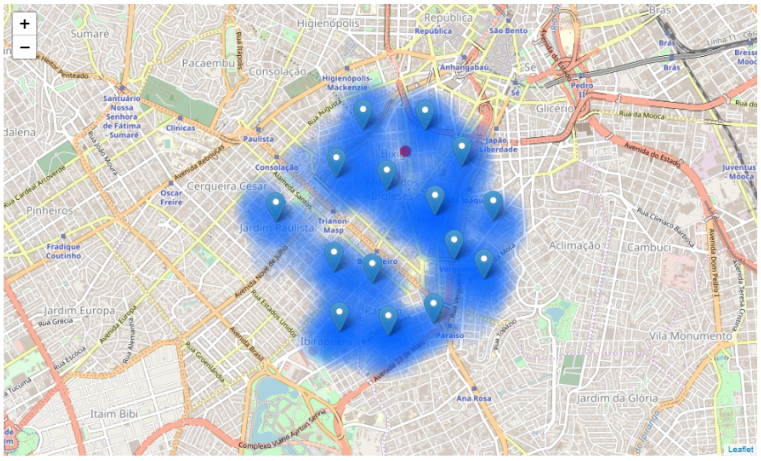

Figure 4: Indication of good places to open a vegan restaurant.

This concludes our analysis. We have created 15 addresses representing centers of zones containing locations with low number of restaurants and no vegan restaurants nearby, all zones being fairly close to the Bixiga area (all less than 4km from Bixiga, and about half of those less than 2km from it). Although zones are shown on map with a radius of ~500 meters (green circles), their shape

is actually very irregular and their centers/addresses should be considered only as a starting point for exploring area neighborhoods in search for potential restaurant locations. Most of the zones are located around the Paulista Avenue, which we have identified as interesting due to being popular with tourists, fairly close to Bixiga and well connected by public transport.

5 Results and Discussion

Our analysis shows that although there is a great number of restaurants around Bixiga, there are pockets of low restaurant density fairly close to this area. Highest concentration of restaurants was detected northeast and southwest from Bixiga, so we focused our attention to areas south, south-east and east, corresponding to area around the Paulista Avenue.

After directing our attention to this more narrow area of interest, we first created a dense grid of location candidates (spaced 100m appart); those locations were then filtered so that those with more than two restaurants in radius of 250m and those with a vegan restaurant closer than 400m were removed.

Those location candidates were then clustered to create zones of interest which contain greatest number of location candidates. Addresses of centers of those zones were also generated using reverse geocoding to be used as markers/starting points for more detailed local analysis based on other factors.

Result of all this is 15 zones containing largest number of potential new restaurant locations based on number of and distance to existing venues - both restaurants in general and vegan restaurants particularly. This, of course, does not imply that those zones are actually optimal locations for a new restaurant! Purpose of this analysis was to only provide info on areas close to Bixiga but not crowded with existing restaurants (particularly vegan) - it is entirely possible that there is a very good reason for small number of restaurants in any of those areas, reasons which would make them unsuitable for a new restaurant regardless of lack of competition in the area. Recommended zones should therefore be considered only as a starting point for more detailed analysis which could eventually result in location which has not only no nearby competition but also other factors taken into account and all other relevant conditions met.

6 Conclusions

Purpose of this project was to identify Sao Paulo areas close to Bixiga neighborhood with low number of restaurants (particularly vegan restaurants) in order to aid stakeholders in narrowing down the search for optimal location for a new vegan restaurant. By calculating restaurant density distribution from Foursquare data we have first identified general boroughs that justify further analysis (Paulista Avenue), and then generated extensive collection of locations which satisfy some basic requirements regarding existing nearby restaurants. Clustering of those locations was then performed in order to create major zones

of interest (containing greatest number of potential locations) and addresses of those zone centers were created to be used as starting points for final exploration by stakeholders.

Final decision on optimal restaurant 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.