

A Restaurant recommendation system to the city of Curitiba, Brazil

Diego Carrillo

Introduction/Business Problem

The idea of this project is to build a simple recommendation system that will return a list of neighborhoods in Curitiba (a city in south of Brazil) that better match an input list of restaurant categories. For example, if the users want to know in what neighborhoods they could find restaurants that match the categories "vegetarian" and "buffet" at the same time, they need only to pass these values to the recommendation system and they will have a list of the neighborhoods with the better results. The results are also grouped by the price range of each neighborhood. So this could be helpful not only to users wanting to find specif venues, but also to entrepreneurs that what to join the food business and need to understand in what neighborhoods of the city they have a higher chance to succeed.

Data

The first step is to obtain a list of all Curitiba's neighborhoods. It can be found at this wikipedia page (https://pt.wikipedia.org/wiki/Lista_de_bairros_de_Curitiba), but same parsing of the html file is necessary to get only the necessary information : Borough, Neighborhood, Area, Latitude, and Longitude. The area parameter is helpful to determine a radius to a circle centered in the latitude and longitude for each neighborhood, so we can fetch this information to our geographical API. The API used used to fetch the venue data for each neighborhood, and for each price range, will be the the Foursquare API. After consulting the API, we will have a dataset containing the following information for each venue: Neighborhood, Venue Name, Venue Category, and Price (An integer within the range of 1 and for 4, where 1 represents the cheaper and 4 the most expensive venue).

Methodology

Once it is obtained the correct dataset containing the category, price, and neighborhood to each city's restaurant, it can be prepared to go through a recommendation system algorithm. Recommendation systems are a collection of algorithms used to recommend items to users based on information taken from the user. These systems have become ubiquitous, and can be commonly seen in online stores, movies databases and job finders. In our case, we will use some simple matrix algebra to grade each neighborhood based on a combination of venues categories.

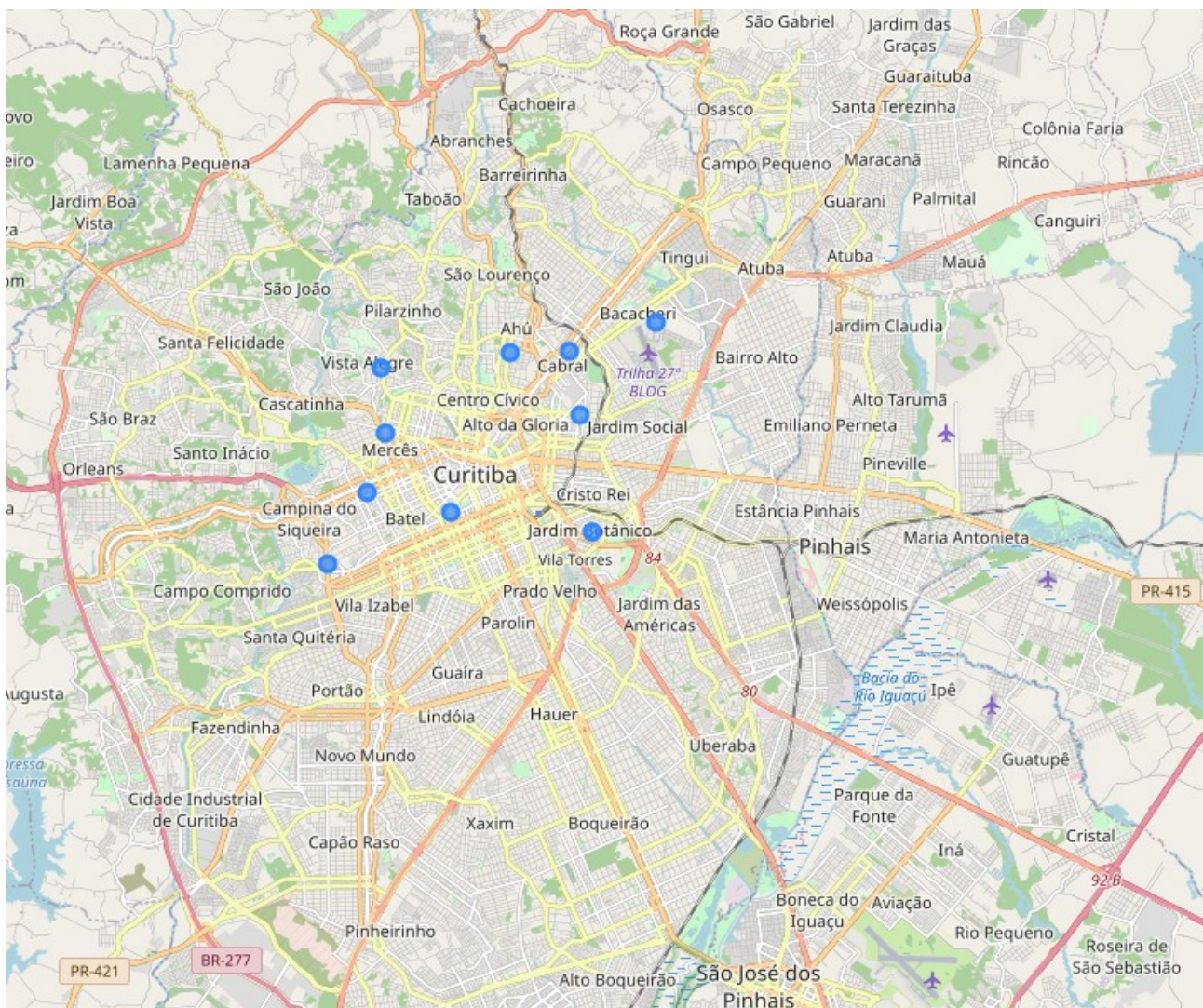
Results

To test the recommender implementation, let's input the list ['Buffet', 'Vegetarian / Vegan Restaurant'] to the algorithm, so it can generate an order dataset with the neighborhoods that most contain restaurants with the profile of a vegetarian or vegan buffet.

The resulting dataset contains all 74 neighborhoods of the city, and most of them have a recommendation grade of 0, since none have a single vegetarian/vegan buffet restaurant. Let's check the top 10 graded neighborhoods for this scenario. We can notice in the table below that we also get a price profile for each row.

Neighborhood	Recomendation grade	Price
Vista Alegre	0.07142857142857142	0.3809523809523808
Seminário	0.06521739130434782	0.4492753623188407
Jardim Botânico	0.04	0.38666666666666666
Bigorrião	0.037037037037037035	0.6172839506172839
Hugo Lange	0.034482758620689655	0.5977011494252874
Bacacheri	0.03076923076923077	0.4307692307692307
Cabral	0.030303030303030304	0.4242424242424242
Mercês	0.029411764705882353	0.4607843137254902
Centro	0.022222222222222223	0.7111111111111111
Ahú	0.022222222222222223	0.5925925925925924

Based on this results, it is helpful to generate a geographic map showing where this neighborhoods are located:



Discussion

Based on the resulting dataset and on the geographical point, we can notice some similarities with the results. Geographically, we can see that the neighborhoods containing vegetarian and vegan buffets are close to the same area in the city, a central one, and also regions that are usually more expensive. From our dataset, we get that the mean restaurant price for all Curitiba's neighborhoods is 0.422507, while the mean price of the top 10 neighborhoods for vegetarian buffets is 0.505138. Based on that, we can infer that this profile of restaurant is above the general mean of the city in terms of price. This kind of information can be very relevant to entrepreneurs that wish to understand the city's restaurant environment before opening a new venue.

Conclusions

Finally, it is possible to conclude that this simple recommendation algorithm, that uses only free and open data, can be very valuable to everyone who wishes to understand Curitiba's restaurant pattern. It is important to clarify that the developed algorithm could be used to any combination of restaurant category available at the foursquare API. Also, if anyone wants to make the same analysis to other cities, they would need only to create another dataset containing the name and locations of the neighborhoods, and apply the same functions that were used here.