

Pointing the right place for an Italian restaurant in Downtown Rio de Janeiro

Marta Bandeira

February 28, 2020

1. Introduction:

1.1 Background

Rio de Janeiro is a popular metropolis where tourism and business meet each other in a very spontaneous way. It is the second large city of Brazil in terms of population, with almost 7 million inhabitants. Downtown is a neighborhood filled with stores, offices, theaters, museums and other venues where an entrepreneur would have a very good chance to be successful. It is the historical, commercial and financial center of the city. The area is mostly crowded during week days at lunch hour. And, although there are a lot of restaurants, it seems that there is still space for a good Italian place.

1.2 Problem

The proposed problem is to open an Italian restaurant in Downtown Rio de Janeiro. The hypothesis set is that there are few Italian restaurants in this area. The majority of food places are either Brazilian buffet or fast food spots. So, a nice place to eat a pasta or a pizza is missing in this lively neighborhood.

1.3 Interest

Entrepreneurs wanting to open a business in Downtown Rio de Janeiro. The data brings the names, the categories and latitude and longitude of venues in the area. Local government would also be interested for public management purposes.

2. Data acquisition and cleaning

2.1 Data sources

The first information was to gather the latitude and longitude of downtown Rio de Janeiro, in order to extract the data from Foursquare. This is the webpage where I found the information:

<https://www.adistanciaentre.com/br/centro-rio-de-janeiro-latitude-longitude-centro-rio-de-janeiro-latitude-centro-rio-de-janeiro-longit/LatLongHistoria/262432.aspx>

Next step was to make a explore call in the Foursquare API plugging in the latitude and longitude of Downtown Rio to obtain the data from the venues of the neighborhood.

2.2 Data cleaning

Foursquare API returns the call in the format of json data. So, it was necessary to transform it into a data frame that looked like this:

	name	categories	lat	lng
0	Dança CCC	Dance Studio	-22.906404	-43.181832
1	Rio Scenarium	Music Venue	-22.908255	-43.183938
2	Centro de Arte Hélio Oiticica	Art Gallery	-22.906106	-43.183425
3	CRAB - Centro Sebrae de Referência do Artesana...	Cultural Center	-22.907662	-43.183819
4	Casa do Choro	Music Venue	-22.906581	-43.180588
5	Adega do Pimenta	German Restaurant	-22.906858	-43.181992
6	Espaço Franklin	Music Venue	-22.905169	-43.182909
7	Lilia Restaurante	Brazilian Restaurant	-22.909135	-43.184018
8	Casa Cavé	Café	-22.905825	-43.178962
9	Espaço Cultural BNDES	Cultural Center	-22.908349	-43.179394
10	Confeitaria Colombo	Pastry Shop	-22.905231	-43.178757
11	Starbucks	Coffee Shop	-22.904947	-43.178935

The data frame returned 100 venues with their names, categories and geospatial attributes.

To solve the proposed problem, I needed to cluster the categories of the venues in order to analyze where the best place to open an Italian restaurant would be. To do that, I grouped the data frame by the column 'categories', which returned 52 entries and tried to run the K-means method for 5, 4, and 3 Kclusters. The best clustering number was 3, which divided the categories into Brazilian Restaurants, Middle Eastern Restaurant and other types of venues.

The clustering process demanded to apply the K-means method to the data frame grouped by categories, resulting into a new column with the Cluster Labels.

	categories	Cluster Labels	name	lat	lng
0	Art Gallery	0	2	2	2
1	Bakery	0	3	3	3
2	Bar	0	3	3	3
3	Bed & Breakfast	0	1	1	1
4	Beer Store	0	1	1	1
5	Bistro	0	1	1	1
6	Brazilian Restaurant	1	15	15	15
7	Buffet	0	1	1	1
8	Burger Joint	0	2	2	2
9	Café	2	5	5	5

After that, I merged the data frame with the Cluster Labels with the first data frame that contained the names, the categories the latitude and longitude of Downtown Rio venues. The resulting data frame looked like this:

	name_x	categories	lat_x	lng_x	Cluster Labels	name_y	lat_y	lng_y
0	Dança CCC	Dance Studio	-22.906404	-43.181832	0	2	2	2
1	Rio Scenarium	Music Venue	-22.908255	-43.183938	0	3	3	3
2	Centro de Arte Hélio Oiticica	Art Gallery	-22.906106	-43.183425	0	2	2	2
3	CRAB - Centro Sebrae de Referência do Artesana...	Cultural Center	-22.907662	-43.183819	0	2	2	2
4	Casa do Choro	Music Venue	-22.906581	-43.180588	0	3	3	3

The geospatial information is duplicated and the counting of the venues' is not useful for the purpose of the study. So, I dropped these columns and renamed the others to exclude the _x label. The final data frame is presented below:

	name	categories	lat	lng	Cluster Labels
0	Dança CCC	Dance Studio	-22.906404	-43.181832	0
1	Rio Scenarium	Music Venue	-22.908255	-43.183938	0
2	Centro de Arte Hélio Oiticica	Art Gallery	-22.906106	-43.183425	0
3	CRAB - Centro Sebrae de Referência do Artesana...	Cultural Center	-22.907662	-43.183819	0
4	Casa do Choro	Music Venue	-22.906581	-43.180588	0
5	Adega do Pimenta	German Restaurant	-22.906858	-43.181992	0
6	Espaço Franklin	Music Venue	-22.905169	-43.182909	0
7	Lilia Restaurante	Brazilian Restaurant	-22.909135	-43.184018	1
8	Casa Cavé	Café	-22.905825	-43.178962	2
9	Espaço Cultural BNDES	Cultural Center	-22.908349	-43.179394	0
10	Confeitaria Colombo	Pastry Shop	-22.905231	-43.178757	0

This data frame made possible to create a map with the 100 most popular venues in Downtown Rio according to the Foursquare API. The information on the map is useful to analyze the best spot for an Italian restaurant.

3. Methodology

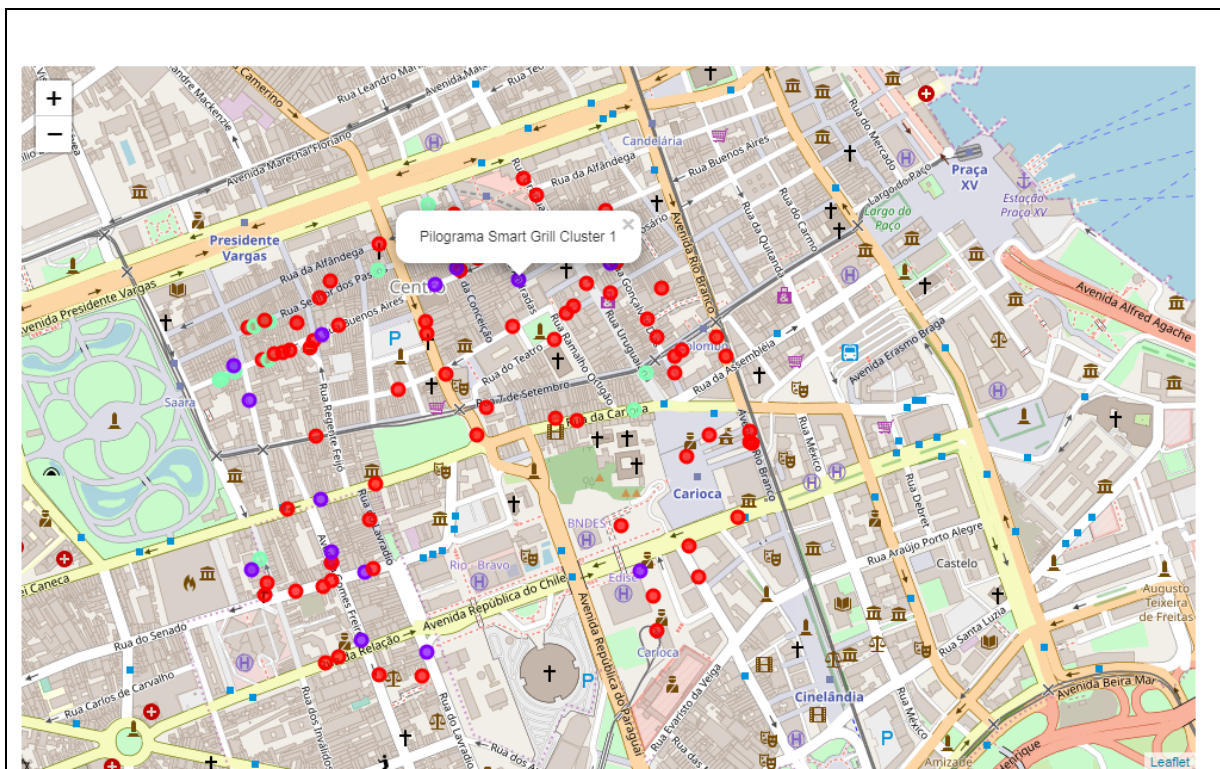
The methodologies used to solve the problem of finding the right place in Downtown Rio to open an Italian restaurant involved data cleaning, clustering and mapping the data to pinpoint each venue according to their cluster. The data acquired in Foursquare was given in the json format, so first I transformed it into a data frame that contained the venues' names, categories and geospatial information.

The idea was to cluster the venues' categories in a manner that could illustrate on the Folium map their location in different colors representing each cluster. With this information is possible to find a place where there are few restaurants and many venues. The right spot would be somewhere with a considerable movement of people and few places to eat. This was possible to find in the resulting map.

4. Results

The final data frame resulted in the map with the venues pinpointed by their cluster. I chose to separate the categories in 3 clusters:

- i) Brazilian Restaurants (purple)
- ii) Middle Eastern Restaurants and Cafes (green)
- iii) Others (red)



restaurants in this zone. Therefore, it would be the best indication to open the Italian restaurant.

5. Discussion

The first observation that it need to be pointed out is that Foursquare does not provide a great volume of data on Downtown Rio de Janeiro. The API only returned 100 venues in the neighborhood. So, this report it is only an exercise to apply the methodologies learned in the Data Science Professional Certificate courses.

The same analysis can be done with a lot more data, which would return better results. The idea was to find a place in Downtown Rio, a very lively neighborhood, full of business and commerce, to open an Italian Restaurant.

Based on the 100 venues returned by Foursquare, it was possible to cluster them by categories and to show them on the map. Visualizing the geospatial information, it is possible to find a place close to venues and with few restaurants around.

Another observation that appeared in the data is the lack of other type of restaurants, including Italian food. Among the most frequent places, the categories related to food were only Brazilian, Middle Eastern and Coffee Shops. In the red cluster (others) the food places were sparsed and diverse.

The circling area number 2 seemed the best place to open the Italian restaurant because it is close to many venues and with few restaurants around. There is only one Middle Eastern restaurant.

6. Conclusions

Downtown Rio de Janeiro is a very agitated neighborhood where entrepreneurs have a good chance to be successful when opening a business. Since it is the commercial and financial center of the city, its lunch hour demands a good offer of dining spots. A good Italian restaurant is always a good choice for those either closing a big deal or for those who are shopping and making tourism.

The methodologies learned in the Data Science Professional Certificate courses made possible to create a data frame from the json data extract from Foursquare and cluster them by the categories of the venues.

The Folium map made possible the visualization of the clusters on the map in order to suggest the best spot for the Italian restaurant.

Github Repository - notebook

https://github.com/MartaBandeira/Coursera_Capstone/blob/master/Downtown%20Rio%20Restaurants.ipynb

IBM Cloud - notebook

https://dataplatform.cloud.ibm.com/analytics/notebooks/v2/27fe5fb6-e986-40d5-b160-6c40ef6b95bc/view?access_token=-