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

Marta Bandeira

Fevereiro/2020

IBM Data Science Professional Certificate

Capstone Project

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.

Data acquisition and cleaning

Data sources - Foursquare

Data cleaning – transforming json data into data frame

	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

Clustering by Categories

Grouping by categories

K – means clustering

Kclusters = 3 (best value, 4 and 5 were tested)

	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

Merging the data frames

Merging the first data frame with all the venues with the clustered by categories data frame
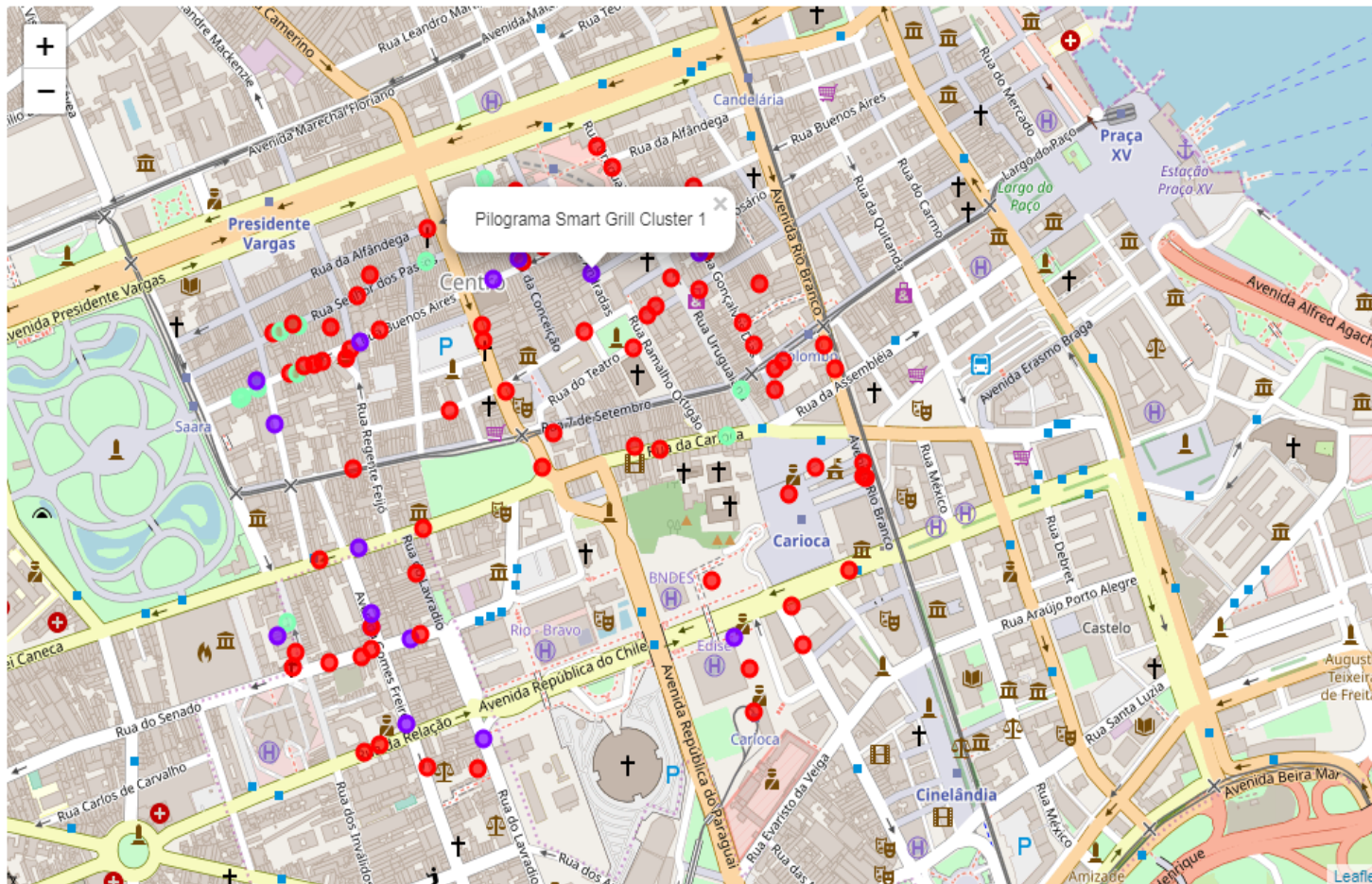Cleaning the useless columns
Final Data Frame:

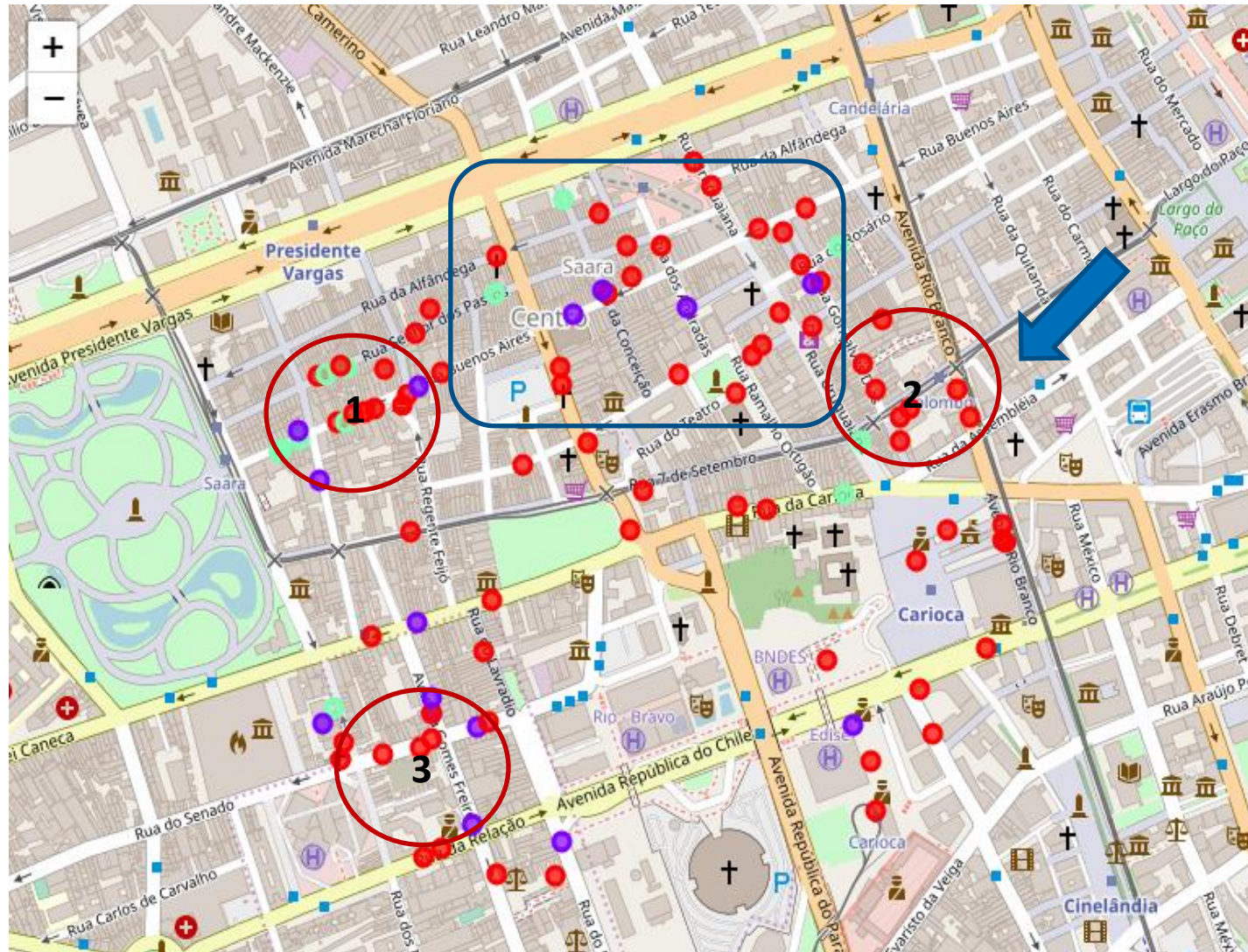|	name	categories	lat	lng	Cluster Labels
0	Dança CCC	Dance Studio	-22.906404	-43.181832	0
1	Rio Scenarium	Music Venue	-22.908255	-43.183938	0
2	Centro de Arte Hélio Oiticica	Art Gallery	-22.906106	-43.183425	0
3	CRAB - Centro Sebrae de Referência do Artesana...	Cultural Center	-22.907662	-43.183819	0
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

Creating the map

Final Data Frame allowed to create a Folium map with the venues pointed out by their clusters with different colors

Suggested place to open the Italian Restaurant – circling area number 2

Results

Suggested place to open the Italian Restaurant – circling area number 2

- The circling region number 1 is the most crowded with venues. It is a very popular commercial place called Saara. Thera are many stores and it is very busy, especially at lunch hour. But it is also a place full of Brazilian and Middle Eastern restaurants.

- In the circling region number 3 there is also a concentration of venues, but there are 5 Brazilian restaurants.

- The blue square area the venues are more spread, but it is also a region with a lot of restaurant offer.

- The circling area number 2 is close to the venues of the blue square region and to Largo da Carioca, which is a very busy place. The Foursquare data is showing few

Conclusions

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.

Obrigado

mybandeiradefreitas@yahoo.com