Steakhouse Branch Neighborhood Selection in Rio de Janeiro

Felipe N. Santos

January 28, 2021

1. Introduction

1.1 Background

Rio de Janeiro, is the second more populous city in Brazil and the sixth-more in the Americas (~6.750.000 citizens). With an area of 1,221 km² (486.5 sq mi). This impressive data gives the dimension of the local market and the consequent opportunities, especially in food and restaurants business.

Even as Rio is a very big city, with 162 neighborhoods, while the the most populous one has more than 300.000 citizens, the regions of the city are very diverse in every sense, including income.

1.2 Problem

In this context, one of the well established local businesses are the Brazilian "all you can eat'' style steakhouses, that are world famous, but this recognition has its price: concurrency.

Meat is an expensive commodity. Used not to be, as Brazil is one of the worlds production leaders, but recent rise in China imports (billed in us dollars) more attractive than internal market, so the local prices suffered a steep rise, growing from US$25 to US$43 for the "arroba"(15Kg).

That change represents a challenge to the steakhouses to maintain the same sales volume (and not close out and fire employers), and, above that, to expand and grow their leverage opening another branch. This became a real difficult and critical decision in the actual scenario for this market.

The objective of this research is to select the best neighborhood to open a new steakhouse, taking in account both low concurrency and population income, as this business typically has an expensive ticket, now more yet.

2. Data Acquisition and processing

2.1 Data sources

I will use two data sets and two APIs for this project:

- https://www.data.rio
-

This is the source of neighbourhood list and geometry.

- https://pt.wikipedia.org/wiki/Lista_de_bairros_do_Rio_de_Janeiro_por_IDH
-

The source of the family income by neighbourhood

- Geolocator

This is for latitude and longitude.

And Foursquare for the venues to discover, as required.

2.2 Data processing

2.2.1 Concurrency

The main idea is to investigate which neighborhoods have not a big steakhouse and at the same time have a high house income.

Foursquare does not has all the steakhouses in Rio, just the big branded ones, This is not a big problem to the scope of this research, as the idea is to serve as reference to the brands that already has many filial, and in this case small local restaurants does not offer the same experience and appeal.

Considered this point, builded a clean list of the neighborhoods in Rio, and visualize in the map.

	Neighbourhood	Latitude	Longitude
0	Paquetá, Rio de janeiro, RJ, Brazil	-22.758926	-43.109199
1	Freguesia (Ilha), Rio de janeiro, RJ, Brazil	-22.785060	-43.169453
2	Bancários, Rio de janeiro, RJ, Brazil	-22.791759	-43.180966
3	Galeão, Rio de janeiro, RJ, Brazil	-22.807506	-43.235521
4	Tauá, Rio de janeiro, RJ, Brazil	-22.797749	-43.186748
5	Portuguesa, Rio de janeiro, RJ, Brazil	-22.799960	-43.206817
6	Moneró, Rio de janeiro, RJ, Brazil	-22.796495	-43.197406
7	Vigário Geral, Rio de janeiro, RJ, Brazil	-22.809533	-43.309704
8	Cocotá, Rio de janeiro, RJ, Brazil	-22.805005	-43.180546

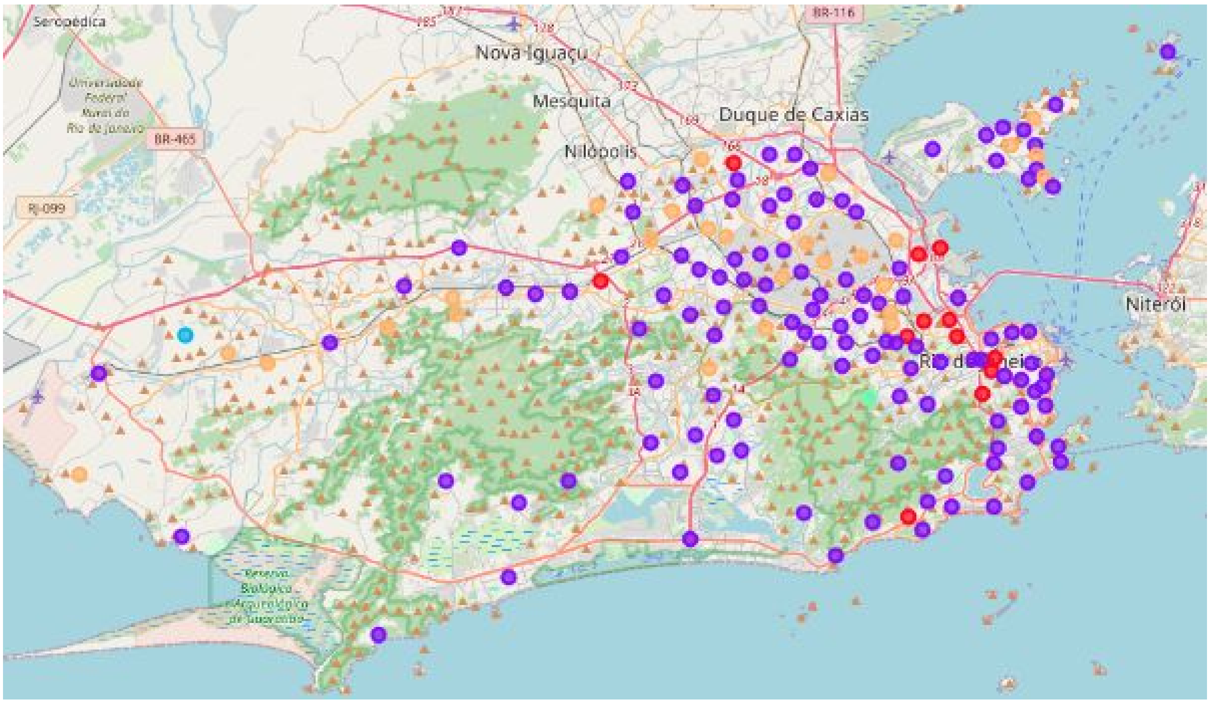
After this initial stage, retrieved the venues data using the Foursquare API.

	Neighbourhood	Neighbourhood Latitude	Neighbourhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Paquetá, Rio de janeiro, RJ, Brazil	-22.758926	-43.109199	Praia da Moreninha	-22.757123	-43.110195	Beach
1	Paquetá, Rio de janeiro, RJ, Brazil	-22.758926	-43.109199	Praia dos Tamoios	-22.761351	-43.107832	Beach
2	Paquetá, Rio de janeiro, RJ, Brazil	-22.758926	-43.109199	Hotel Lido	-22.762042	-43.110205	Boarding House
3	Paquetá, Rio de janeiro, RJ, Brazil	-22.758926	-43.109199	Tia Leleta Bar	-22.760765	-43.108334	Brazilian Restaurant
4	Paquetá, Rio de janeiro, RJ, Brazil	-22.758926	-43.109199	Quintal da Regina	-22.760820	-43.108288	Bar

Using "one-hot encoding'' processing, the most common venue types for each neighborhood were calculated. Above that a clustering process was made with K-Means algorithm, grouping the neighborhoods with similar venue types concentration.

	Neighbourhood	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
140	Leme, Rio de janeiro, RJ, Brazil	-22.961704	-43.166904	3.0	Hotel	Scenic Lookout	Beach	Seafood Restaurant	Pizza Place
56	Tomás Coelho, Rio de janeiro, RJ, Brazil	-22.867698	-43.306088	3.0	Paintball Field	Factory	Food Truck	Train Station	Gym / Fitness Center
66	Manguinhos, Rio de janeiro, RJ, Brazil	-22.879676	-43.250888	3.0	Train Station	Snack Place	Gym	Deli / Bodega	Restaurant
48	Engenho da Rainha, Rio de janeiro, RJ, Brazil	-22.862650	-43.293787	0.0	Brazilian Restaurant	Music Venue	Bakery	Food & Drink Shop	Pastry Shop
118	Pechincha, Rio de janeiro, RJ, Brazil	-22.928988	-43.353419	3.0	Plaza	German Restaurant	Market	Breakfast Spot	Brazilian Restaurant

The clusters were plotted to visualization.



Custer analysis shows very different constitution of local commerce.

With the criteria of significant population of restaurants in three most common venues, the concentration data is:

Cluster	Restaurants / Venues	Steakhouses and BBQ
0	21 / 36 (58,3%)	1
1	87 / 159 (54,7%)	5
2	1 / 3 (33,3%)	0
3	1 /3 (33,3%)	0
4	19 / 81 (23,4%)	0

Clusters 0 and 1 are clearly the ones with greater concentration of restaurants (> 50%) between the 3 most common columns.

So, in the final intersection of the two methods, will select the neighborhoods from remaining clusters (2, 3, and 4) with high house income, to do a combined funnel.

This was recorded as the "candidate list'' to be crossed with the clustering filtered results, funneling to the more specific neighborhood advice objective.

Bancários	Jardim Carioca	Pavuna	Cordovil	Praia da Bandeira
Zumbi	Guadalupe	Parque Anchieta	Honório Gurgel	Rocha Miranda
Ramos	Vila Kosmos	Paciência	Engenho da Rainha	Complexo do Alemão
Senador Camará	Cavalcante	Higienópolis	Quintino Bocaiúva	Senador Vasconcelos
Cosmos	Jacarezinho	Inhoaíba	Jacaré	Tanque
Grajaú	Sepetiba	Deodoro	Jabour	

2.2.2 Social Filtering

As the great part of the list is of poor locals (so the low restaurants count), is necessary to apply a social filtering, both eliminating the mentioned "false positive'' of low concurrency by low demand and selecting high potential locals).

3. Result

Loading the social data and creating a cut with the third quarter of family income statistics, the resulted list returned just one option: "Grajaú ".

4. Enhancing Possibilities

Is common sense in data science “that the quality of the prevision is only as good as the data set’’. So, with a more extensive API like Google Maps, which has a lot more venues registered, maybe a different approach can be used and compared, utilizing only steakhouse, BBQ etc. for the cluster selection. This possibly can be more precise and return more options.

4. Conclusion

With the data available and a combination of two selection methods (clustering and filtering), the best neighborhood to open a steakhouse branch in Rio de Janeiro is Grajaú.