Capstone Project - The Battle of the Neighborhoods

Applied Data Science Capstone by IBM/Coursera

Luciano Pinheiro

jun 2021

Business problem

Belém is an poor Amazon city and capital of Pará State, in Brazil. It has about 1.5 million residents. On the last decades, the city had a exponential growth. But this growth was not organized. So, while some places had a good development, with lots of new big and small business, others were simply forgotten or occupied in a illegal way.

The idea here is to explore the city trying to find the distribution of a particular business over the city, trying to find some gaps, which may help entrepreneurs to find places to start new business.

We will demonstrate the solution with pharmacies. But you may notice that with just small changes, it will work with any sort of business.

Data

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

- number of existing pharmacies in the neighborhood
- the size of district/neighborhood

We will get the names of districts on an Wikipedia page. Then, the geolocation will be returned by geocoder python library.

Another source of data is the list of pharmacies from Foursquare API.

On that API, there is no district information, so we must apply a function to link the venue/place to some neighborhood.

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

- Wikipedia
- Foursquare API
- geocoder python library

Methodology

Having the main data, we can see the distribution, the density and the gaps on the map. We may now intersect the amount of pharmacies on each district and their population to find the best places where an entrepreneur may use the data. With luck, we may find a good way to represent that.
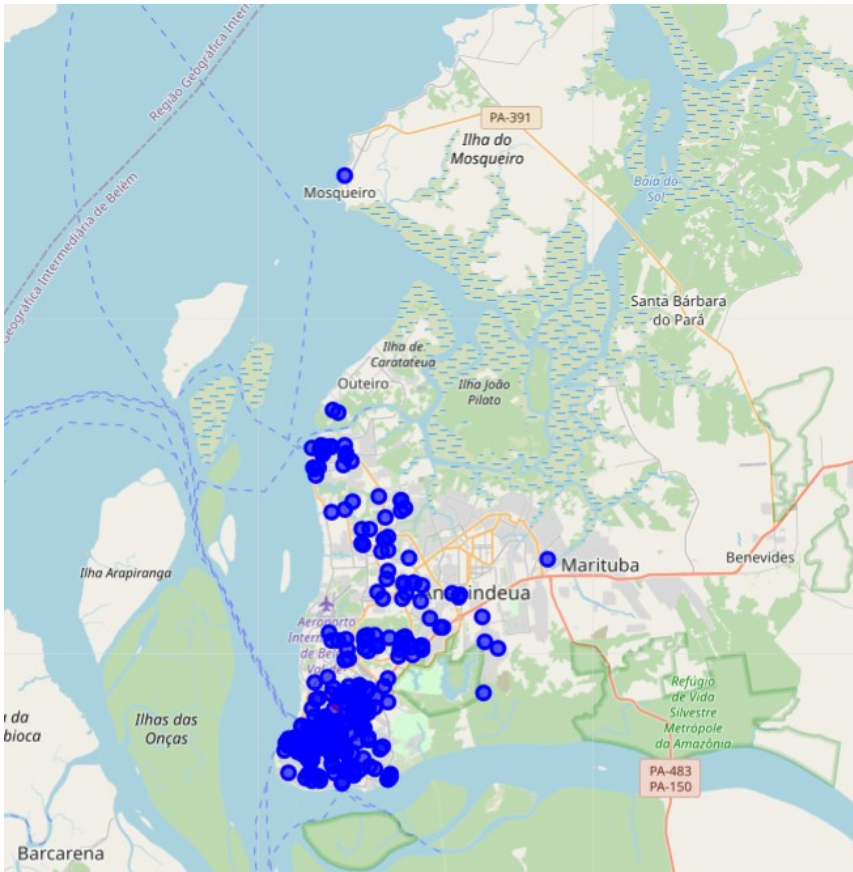
Analysis

Image 1: Distribution of pharmacies in Belém

We may now intersect the amount of pharmacies on each district and their population, to find the best places where an entrepreneur may use the data. Best places here are places with high density of people and low density of pharmacies.

The next table show the neighborhoods, the number of pharmacies and the number of pharmacies per 1000 inhabitants.

district	inhabitants	pharmacies	pharmacies per 1000
Sucurijuquara	1074	0.0	0.000000
Souza	13190	0.0	0.000000
Murubira	1519	0.0	0.000000
Mangueirão	36224	0.0	0.000000
Natal do Murubira	1098	0.0	0.000000
Águas Negras	6555	0.0	0.000000
Água Boa	8553	0.0	0.000000
Itaiteua[8][9]	1939	0.0	0.000000
Maracajá	3345	0.0	0.000000
São João do Outeiro	12134	0.0	0.000000
Ariramba	1942	0.0	0.000000

Pratinha	22589	0.0	0.000000
São Clemente	7714	0.0	0.000000
Aeroporto	1170	0.0	0.000000
Maracacueira	9819	0.0	0.000000
Miramar	515	0.0	0.000000
Bonfim	776	0.0	0.000000
Caruará	794	0.0	0.000000
Cabanagem	27781	1.0	0.035996
Campina de Icoaraci	26722	1.0	0.037422
Barreiro	26003	1.0	0.038457
Telégrafo	42953	2.0	0.046563
Sacramenta	44413	3.0	0.067548
Benguí[nota 1]	29379	2.0	0.068076
Cidade Velha	12128	1.0	0.082454
Coqueiro	51776	5.0	0.096570
Paracuri	9934	1.0	0.100664
Parque Verde	39126	4.0	0.102234
Montese	61439	7.0	0.113934
Condor	42758	5.0	0.116937
Tapanã	66669	8.0	0.119996
Jurunas	64478	8.0	0.124073
Guamá	94610	12.0	0.126836
Tenoné	30429	4.0	0.131454
Maracangalha	30534	5.0	0.163752
Águas Lindas	17520	3.0	0.171233
Cremação	31264	6.0	0.191914
Pedreira	69608	14.0	0.201126
Marambaia	66708	15.0	0.224861
Curió-Utinga	16642	4.0	0.240356
Agulha	19712	5.0	0.253653
Marco	65844	17.0	0.258186
Ponta Grossa	13245	4.0	0.302001
Castanheira	24424	8.0	0.327547
Brasília	6019	2.0	0.332281
Mangueiras	2851	1.0	0.350754
Aurá	1827	1.0	0.547345
Umarizal	30090	18.0	0.598205

Nazaré	20504	13.0	0.634023
Fátima	12385	8.0	0.645943
Cruzeiro	11369	8.0	0.703668
Una	6724	5.0	0.743605
São Brás	19936	15.0	0.752408
Batista Campos	19136	17.0	0.888378
Val-de-Cans	7032	7.0	0.995449
Reduto	6336	9.0	1.420455
Guanabara	1919	3.0	1.563314
Campina	6156	12.0	1.949318
Universitário	2557	7.0	2.737583

On the following map, we combined the size of neighborhood (green) with the number of pharmacies per 1000 inhabitants (blue). The red circles are districts with no pharmacies.

We may see that some big districts have a small number of pharmacies, and that may be some a good opportunity. In general, the bigger the green aura, the better the opportunity.

Image 1: Neighborhoods and pharmacies

We can see visually in the map that some big neighborhoods have a small number of pharmacies. So, we can create an Opportunity Index based on that observation. The index may use the size of neighborhood and the number of pharmacies per capita.

The formula is as follow:

OI = inhabitants/4000 – (3+ Pharmacies_per_1000]*8)

The 10 best districts are:

district	inhabitants	pharmacies	Pharmacies per 1000	Opportunity index
Guamá	94610	12.0	0.126836	19.637808
Pedreira	69608	14.0	0.201126	12.792990
Tapanã	66669	8.0	0.119996	12.707284
Jurunas	64478	8.0	0.124073	12.126913
Marambaia	66708	15.0	0.224861	11.878115
Montese	61439	7.0	0.113934	11.448277
Marco	65844	17.0	0.258186	11.395512
Coqueiro	51776	5.0	0.096570	9.171441
Sacramenta	44413	3.0	0.067548	7.562868
Telégrafo	42953	2.0	0.046563	7.365750

Results and discussion

The data shows that Belém is short of pharmacies. Some districts have 0 activities. Based only at location, we could say that any of that districts may be eligible for a new business. Visually, we have 4 highlights: the Island of Mosqueiro, the Island of Caratateua, Mangueirão district and around Pratinha neighborhood.

Other interesting view is the last map, which show the relationship between the size of the district and the number of pharmacies. We even made a formula to show numerically this relationship. In this way, Guamá, Pedreira e Tapanã are very dense districts with relative low amount of pharmacies.

We must at last say that this analysis worked only in two dimensions: the number of pre-existent business and the population of districts. We know for experience that this is a violent city and maybe the low density areas are too risky to open a business. Some neighborhoods may have no basic services, like paved streets or bus line nearby.

For further work, we could make the use of spatial security data, like police stations location overall real state price. It is also interesting to play with Opportunity Index formula. For example, in another combination of parameters, Mangueirão was the second best due to lack of pharmacies in a 32K inhabitants in a small area.

Conclusion

The main goal of this work is to find opportunities to a pharmacy business in the city of Belém. Due to low density of competitors, it's not so hard to find a good place, although, as said before, the location and population are not the only dimensions to evaluate.