Choosing a Neighborhood for a New Pizza Place in Manaus

Carlos V. S. Araujo[1]

[1]IBM Applied Data Science Capstone

vicente@icomp.ufam.edu.br

1. Introduction

I want to start this text saying that English is not my first language, so maybe some typos and language problems in general may occur, but I will try my best to minimize this errors.

For my capstone project I decide to examine the city the I was born and lives till nowadays. Brazil is passing a big financial crisis and this scenario will probably not change since the actual government is taking decisions that are making this crisis even more larger.

With the number of unemployed persons rising, a option that a lot of them are taking is to create an own business. When walking through Manaus you can often see new small markets and restaurants, but with this large amount of business emerging is getting hard to choose the best neighborhood for a new one.

In this project, I will develop a model for helping a person to chosen the best neighborhood to install a new Pizza Place in Manaus. The best place can be choose based on the number of pizza places in the surroundings and the financial status of the neighborhood.

The rest of this report is as follow: on Section 2 I talk about the data that I used on this project, on Section 3 the methodology is presented, the results I got are on Section 4 and discussed on Section 5, finally on Section 6 I make a brief conclusion and point out some future works that can be done.

2. The Data

The data I used on this project was taken from three different sources, they are:

1. From Wikipedia[1] we got the name of Manaus' neighborhoods. Since pandas can get data directly from a URL I didn't need any other package to scrap this information from the website;
2. To get the latitude and longitude for every neighborhood I used the Geocoder[2] package. With this library we can easily get the latitude and longitude with only the names of the neighborhoods and the city;
3. Finally, to get the venues on these neighborhoods I used the Foursquare API[3] that based on a latitude and longitude can return a list of venues on a given radius.

[1] https://pt.wikipedia.org/wiki/Lista_de_bairros_de_Manaus
[2] https://geocoder.readthedocs.io/
[3] https://developer.foursquare.com/

3. Methodology

On this section I will explain the step-by-step used for making this project. The first step was to making the imports, they basically are: *Numpy* and *Pandas* to handle data, *Geocoders* and *Geopy* to get latitude and longitude from locations, *Matplotlib* for colors, *KMeans* from *Scikit-learn* to create clusters based on our locations and *Folium* to generate the maps.

The names from Manaus' neighborhoods were easily got from Wikipedia using the *read_html* Pandas' function. A total of 63 neighborhoods were collected. We used *Geocoders* to get the latitude and longitude for each one of these neighborhoods. I developed a *get_latlng* function for this step. Then, with the dataframe containing the name of the neighborhoods and its geographical location, I save it as a .CSV file.

The next step was plot a map of Manaus with markers on each neighborhood. For this, we first used the *Nominatim* function from *geopy.geocoders* to get Manaus' latitude and longitude. Then with *Folium*, I ploted the map and saved it on a .html file.

To get the venues on the neighborhoods, I used the Foursquare API that can take a limit of 100 venues for each neighborhood. I used a radius of 2000 meters from the point of entrance. Remember that for making this step a Foursquare ID and Secret are needed. I save this information on an another dataframe that I also saved on a file.

Then, I made a process of one-hot-encoding of the *VenueCategory* column in the dataframe, so I group the entrances by neighborhood and I take the mean of the frequency of occurrence of each category. Then, I consulted that 55 of the 63 neighborhoods had a Pizza Place on it.

Finally, I run the K-Means algorithm with $K = 5$ and plot the map with each cluster in a different color. I also print the neighborhoods that are in each cluster.

4. Results

This text was made using LaTeX, so is likely that the images of this section were placed far away from the text, but all images are hyperlinked.

The first map that contains a marker for every neighborhood of Manaus is presented in Figure 1, the name of neighborhood appears if the marker is clicked.

On the Figure 2 is the map where the neighborhoods from Manaus are colored based on what cluster it is in. The cluster of every neighborhood was decided by the K-Means algorithm based on Pizza Place locations.

The neighborhoods presented in each cluster are presented in Figures 3, 4, 5, 6, 7 that represents clusters 0, 1, 2, 3 and 4, respectively.

5. Discussion

Based on the results presented on Section 4, the best neighborhoods to install a new Pizza Place are the ones on cluster 3. The neighborhoods on this cluster have a low number of Pizza Places or even don't have one. Alvorada, for instance, is the sixth neighborhood with most domestic residences on the city[4], so presents a greater chance for the success of a new Pizza Place.

[4] https://pt.wikipedia.org/wiki/Lista_de_bairros_de_Manaus

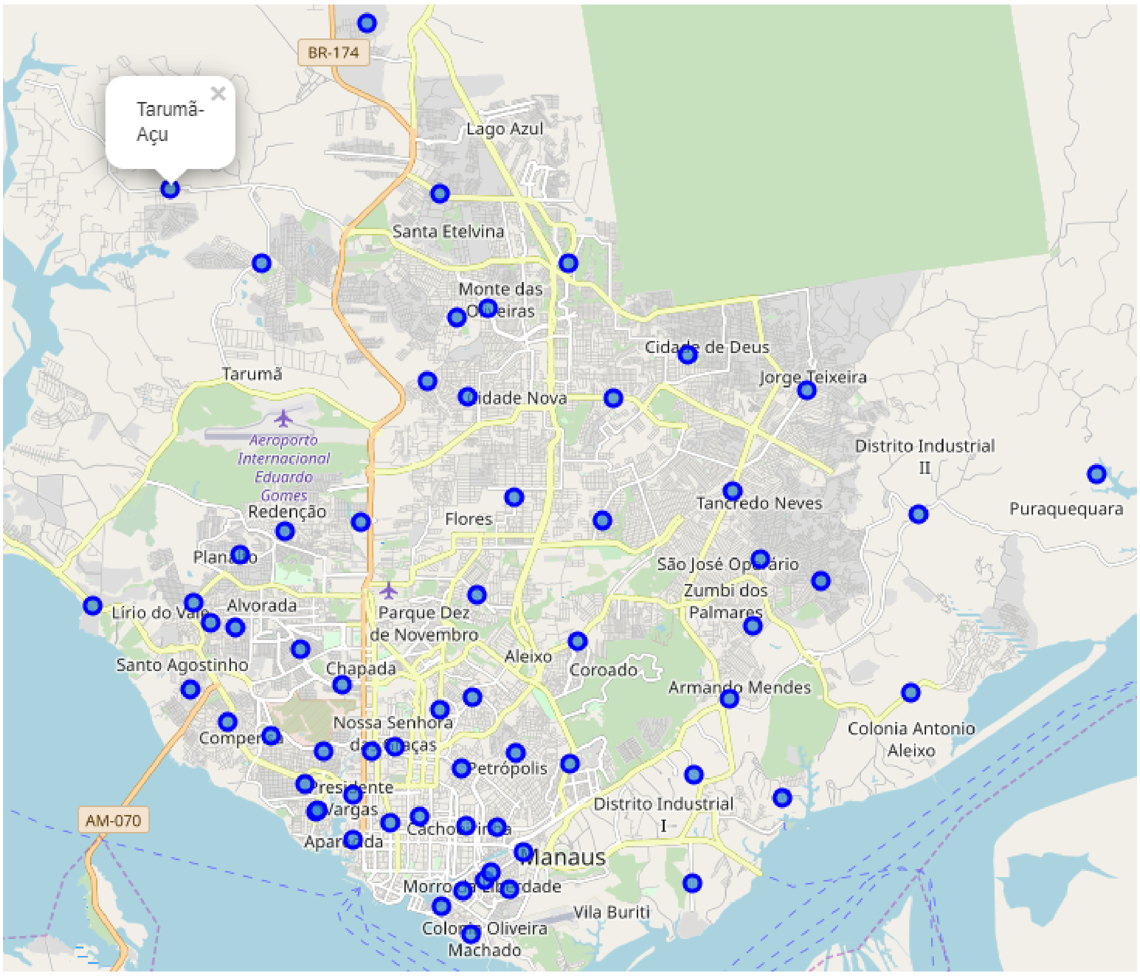

Figure 1. Map of Manaus' neighborhoods.

On the other hand, the neighborhoods from clusters 1 and 4 are the ones to be avoided, since these areas contain the large number of Pizza Places, so a new one will have a large competition from already established restaurants.

Neighborhoods on cluster 0 and 2 could be also good options for a new Pizza Place, since the competition will not be as difficult as on clusters 1 and 4. If the new place has some differential over the others, maybe it will prevail.

6. Conclusion and Future Work

On this project, I developed a model for predicting the best neighborhood to create a new Pizza Place in the city of Manaus in Brazil. To create this model, I used information from locations of others Pizza Places in the city that I got using the Foursquare API.

Using a K-Means algorithm, I put the neighborhoods of the city in five different clusters. One of these clusters contained basically neighborhoods with a low number of Pizza Places, or even with no one in it. These neighborhoods were considered to be the best options to build a new Pizza Place on the city. We also found out the neighborhoods with the largest amount of Pizza Places, so they were not recommended.

For future work, a good idea is to apply this methodology for different business

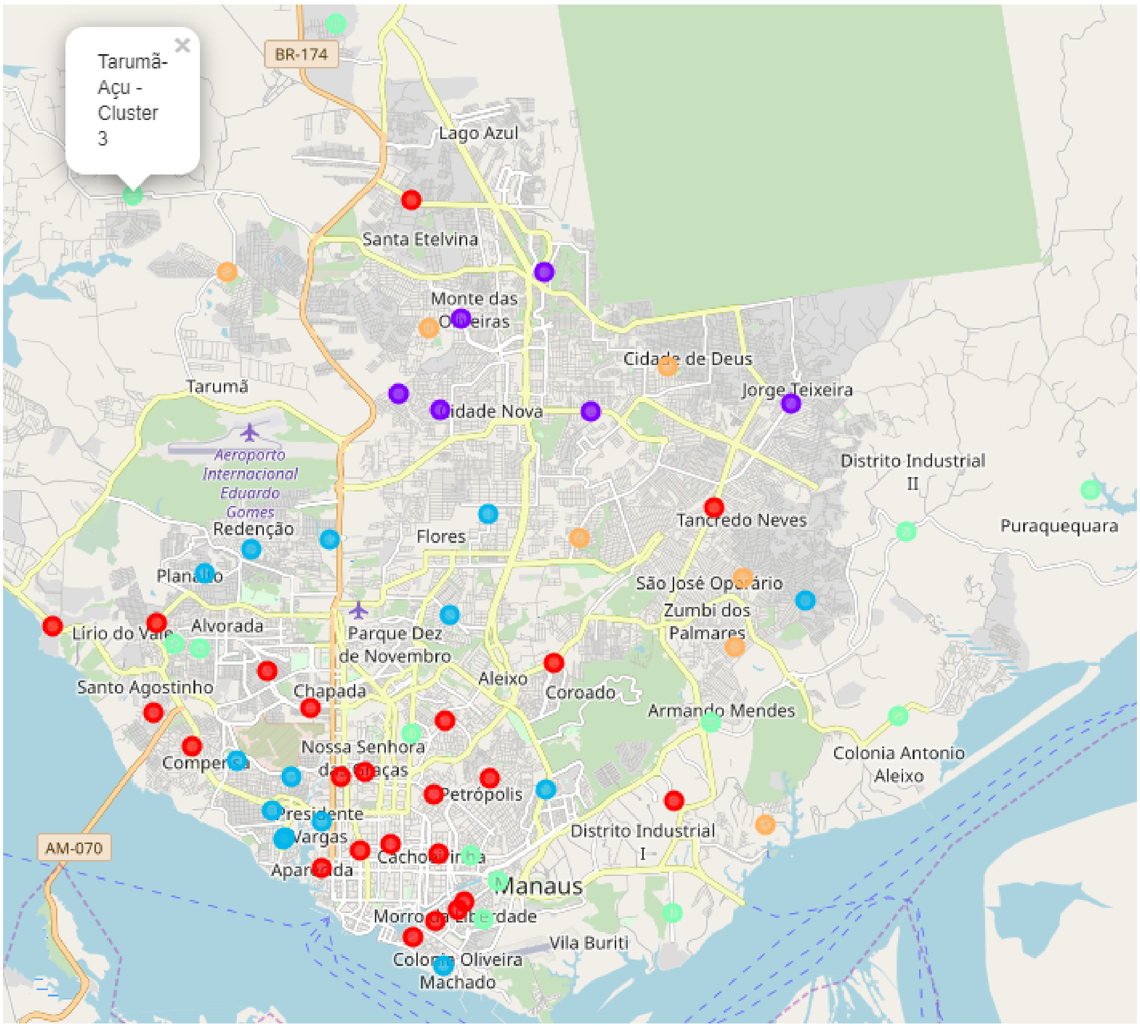

Figure 2. Map after clustering based on Pizza Place locations.

types and cities.

	Neighborhood	Pizza Place	Cluster Labels	Latitude	Longitude
31	Morro da Liberdade	0.030000	0	-3.13664	-59.99987
33	Nossa Senhora das Graças	0.030000	0	-3.10729	-60.01957
28	Lírio do Vale	0.020000	0	-3.07531	-60.06433
51	São Francisco	0.020000	0	-3.11206	-60.00492
49	Santo Agostinho	0.020000	0	-3.09457	-60.06487
48	Santa Luzia	0.020000	0	-3.13906	-60.00441
47	Santa Etelvina	0.028571	0	-2.98487	-60.00960
21	Educandos	0.020000	0	-3.14273	-60.00937
20	Dom Pedro	0.030000	0	-3.08579	-60.04079
52	São Geraldo	0.030000	0	-3.10819	-60.02475
18	Distrito Industrial I	0.031250	0	-3.11344	-59.95334
14	Compensa	0.030303	0	-3.10161	-60.05682
39	Petrópolis	0.030000	0	-3.10865	-59.99303
15	Coroado	0.026667	0	-3.08387	-59.97915
6	Centro	0.030000	0	-3.12399	-60.02063
1	Aleixo	0.030000	0	-3.09622	-60.00263
42	Praça 14 de Janeiro	0.030000	0	-3.12252	-60.01424
4	Betânia	0.020000	0	-3.13487	-59.99843
5	Cachoeirinha	0.020000	0	-3.12461	-60.00391
7	Chapada	0.030000	0	-3.09341	-60.03136
32	Nossa Senhora Aparecida	0.020000	0	-3.12783	-60.02900
57	Tancredo Neves	0.015873	0	-3.05077	-59.94501
41	Ponta Negra	0.017857	0	-3.07591	-60.08657

Figure 3. Neighborhoods present in the cluster 0.

	Neighborhood	Pizza Place	Cluster Labels	Latitude	Longitude
12	Colônia Santo Antônio	0.098765	1	-3.02968	-60.00359
37	Novo Israel	0.098039	1	-3.02629	-60.01250
34	Nova Cidade	0.096774	1	-3.00026	-59.98115
26	Jorge Teixeira	0.111111	1	-3.02839	-59.92836
30	Monte das Oliveiras	0.086957	1	-3.01041	-59.99912
8	Cidade Nova	0.091837	1	-3.03013	-59.97116

Figure 4. Neighborhoods present in the cluster 1.

	Neighborhood	Pizza Place	Cluster Labels	Latitude	Longitude
38	Parque 10 de Novembro	0.040000	2	-3.07368	-60.00146
43	Presidente Vargas	0.040000	2	-3.11788	-60.02900
40	Planalto	0.050000	2	-3.06484	-60.05415
22	Flores	0.034483	2	-3.05193	-59.99318
50	Santo Antônio	0.040000	2	-3.11538	-60.03959
25	Japiim	0.040000	2	-3.11095	-59.98095
24	Glória	0.040000	2	-3.12127	-60.03669
23	Gilberto Mestrinho	0.047619	2	-3.07046	-59.92513
17	Da Paz	0.040000	2	-3.05772	-60.02716
53	São Jorge	0.050000	2	-3.10827	-60.03556
11	Colônia Oliveira Machado	0.035294	2	-3.14878	-60.00275
56	São Raimundo	0.040000	2	-3.12169	-60.03723
61	Vila da Prata	0.050000	2	-3.10480	-60.04709
46	Redenção	0.040000	2	-3.05946	-60.04406

Figure 5. Neighborhoods present in the cluster 2.

	Neighborhood	Pizza Place	Cluster Labels	Latitude	Longitude
55	São Lázaro	0.01	3	-3.13889	-59.99412
59	Tarumã-Açu	0.00	3	-2.98379	-60.06931
60	Vila Buriti	0.00	3	-3.13729	-59.95393
0	Adrianópolis	0.00	3	-3.09907	-60.00971
44	Puraquequara	0.00	3	-3.04676	-59.86425
35	Nova Esperança	0.01	3	-3.07978	-60.06053
27	Lago Azul	0.00	3	-2.94707	-60.02602
19	Distrito Industrial II	0.00	3	-3.05580	-59.90360
16	Crespo	0.01	3	-3.13042	-59.99135
10	Colônia Antônio Aleixo	0.00	3	-3.09518	-59.90549
3	Armando Mendes	0.00	3	-3.09647	-59.94543
2	Alvorada	0.01	3	-3.08071	-60.05521
45	Raiz	0.01	3	-3.12525	-59.99695

Figure 6. Neighborhoods present in the cluster 3.

	Neighborhood	Pizza Place	Cluster Labels	Latitude	Longitude
36	Novo Aleixo	0.083333	4	-3.05731	-59.97373
29	Mauazinho	0.076923	4	-3.11864	-59.93371
54	São José Operário	0.068182	4	-3.06567	-59.93851
13	Colônia Terra Nova	0.083333	4	-3.01232	-60.00587
9	Cidade de Deus	0.062500	4	-3.02062	-59.95474
58	Tarumã	0.076923	4	-3.00029	-60.04924
62	Zumbi dos Palmares	0.071429	4	-3.08040	-59.94052

Figure 7. Neighborhoods present in the cluster 4.