Neighborhood Wars: Manaus

By Thiago Santos Figueira

May 25, 2021

Summary

Hello!

This is the report with the solution for my capstone project for this Professional Certificate.

1. Introduction: Business Problem

When moving to a new city, we face the problem of going into uncharted territory (at least, for us!). We want to choose a neighborhood that is secure, friendly, and close to what we need. Part of these requirements is fulfilled by the availability and proximity of core services, such as hospitals and schools. In this project, I will look at the neighborhoods of Manaus (in the state of Amazonas, Brazil) to determine which ones would be a nice place to live by looking at the count of available schools. This solution helps people looking to move to a neighborhood in Manaus find the one with the greatest number of schools available. It may also help the government understand which regions need more investment in the sector.

Amazon Opera House in Manaus

The main tools I will use are:

1. Web-scrapping to acquire the data
2. Geocode API to collect lat/long coordinates
3. The Foursquare API to determine the number of schools in the vicinities
4. Folium maps to visualize the region

I want to answer two main questions:

- The question we are answering is: which neighborhoods have the highest options regarding nearby schools?
- Which regions should the government look after when it comes to school availability?

2. Data Overview

The neighborhood dataset is available on this Wikipedia page. Let us have a look at it:

Lista de Bairros

Bairro[2]	Zona administrativa	Área (ha)[2]	População (estimativa 2017)[2]	Densidade Demográfica (hab./km²)	Domicílios particulares[2]
Adrianópolis	Centro-Sul	248,45	10 459	3 560,88	3 224
Aleixo	Centro-Sul	618,34	24 417	3 340,40	6 101
Alvorada	Centro-Oeste	553,18	76 392	11 681,73	18 193
Armando Mendes	Leste	307,65	33 441	9 194,86	7 402
Betânia	Sul	52,51	12 940	20 845,55	3 119
Cachoeirinha	Sul	197,71	20 035	8 572,15	5 363
Centro	Sul	426,94	39 228	7 772,29	10 828
Chapada	Centro-Sul	241,27	13 219	4 634,64	4 324
Cidade de Deus	Norte	676,76	82 919	10 364,38	19 385
Cidade Nova	Norte	1 419,38	143 201	8 534,36	34 239

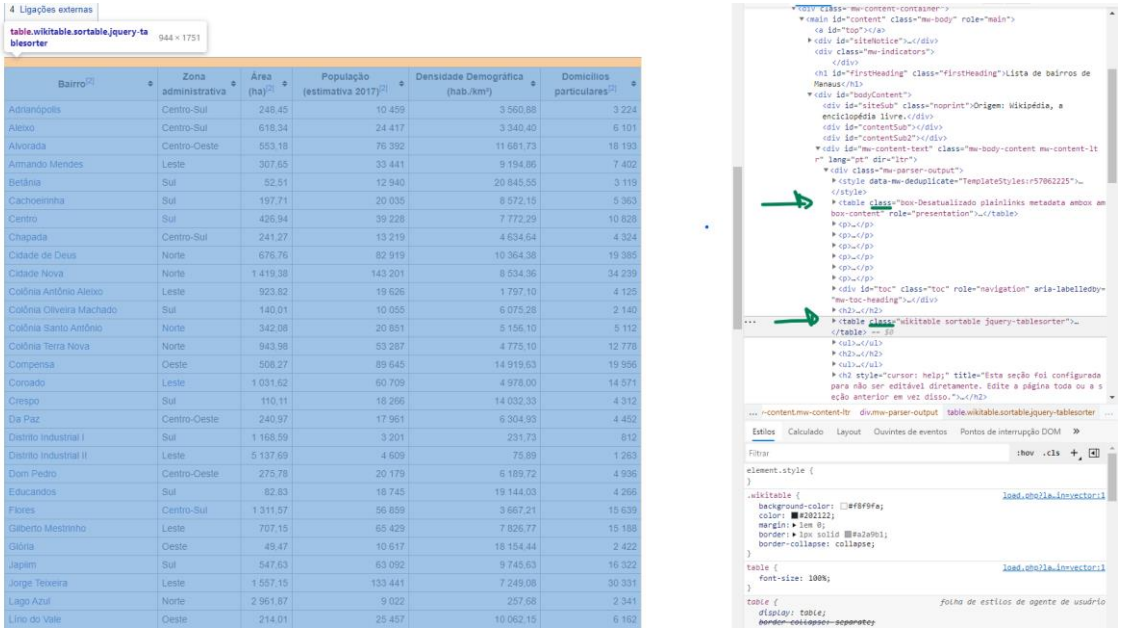
These are the first rows of the dataset. There is a total of 63 neighborhoods in Manaus. As you may have noticed, the data is in Portuguese, which is the native language of Brazil. Let us understand what each column represents in English:

Column in Portuguese	**Column in English**
Bairro	Neighborhood
Zona administrativa	Zone
Área	Area
População	Population
Densidade demográfica	Density
Domicílios particulares	Homes

Notice neighborhoods are organized in zones (South, North, East, South-Center, etc.). Some are larger than others in total area size and in demographic density. In addition to the data available, we will need to collect latitude and longitude coordinates to feed to the Foursquare API.

3. Methodology

The first step is to collect the data. For that, we will use the web-scrapping library *Beautiful Soup*. After downloading the HTML of the table, we need to find the table we want. We could retrieve the first table available, but there is the possibility the page contains more than one table, which is common in Wikipedia pages. For this reason, we must look at all tables and find the correct one. Let us have a look at the structure of the HTML.



HTML structure of the page

Notice there is indeed more than one table. In the image above, the highlighted table is the one we want to collect. Unfortunately, the tables do not have a title, but they do have a class attribute. We can use this information to pick the correct table. We use the beautiful soup library to filter the correct table according to the classes we identified (*wikitable* and *sortable*). We then create a data frame with the columns translated to English. The result is visible in the picture below:

	Neighborhood	Zone	Area	Population	Density	Homes_count
0	Adrianópolis	Centro-Sul	248.45	10459	3560.88	3224
1	Aleixo	Centro-Sul	618.34	24417	3340.4	6101
2	Alvorada	Centro-Oeste	553.18	76392	11681.73	18193
3	Armando Mendes	Leste	307.65	33441	9194.86	7402
4	Betânia	Sul	52.51	1294	20845.55	3119

We could have achieved the same result using the *Pandas* method *read_html*. This method returns a list of data frames containing html elements that satisfy our attribute specifications. For this case, we are looking for a class that contains the classes: *wikitable* and *sortable*. The thousands parameter specifies the separator used to parse thousands. The result is visible in the image below:

```
df_pandas[0].head()
```

	Bairro[2]	Zona administrativa	Área (ha)[2]	População (estimativa 2017)[2]	Densidade Demográfica (hab./km²)	Domicílios particulares[2]
0	Adrianópolis	Centro-Sul	248,45	10 459	3 560,88	3 224
1	Aleixo	Centro-Sul	618,34	24 417	3 340,40	6 101
2	Alvorada	Centro-Oeste	553,18	76 392	11 681,73	18 193
3	Armando Mendes	Leste	307,65	33 441	9 194,86	7 402
4	Betânia	Sul	52,51	12 940	20 845,55	3 119

After acquiring our dataset, we need to collect the longitude and latitude coordinates for each neighborhood. To achieve the desired result, we will import from the *Pandas* library *Geopy* the module *Nominatim*. We will create a column in the df to store the coordinates of each neighborhood. We will use the *rate limiter* from *geopy* because it allows us to perform bulk operations while gracefully handling error responses and adding delays when needed. We have the data, but it would be easier to manipulate each latitude/longitude value on its own column. We, then, use the function *apply* to transform our single column into multiple descriptive columns. Our dataframe looks like this:

	Neighborhood	Zone	Area	Population	Density	Homes_count	Full_Address	Location	Point	Latitude	Longitude	Altitude
0	Adrianópolis	Centro-Sul	248.45	10459	3560.88	3224	Adrianópolis, Manaus, Amazonas	Adrianópolis, Manaus, Microrregião de Manaus, ...	(-3.1016973, -60.0089746, 0.0)	-3.101697	-60.008975	0.0
1	Aleixo	Centro-Sul	618.34	24417	3340.40	6101	Aleixo, Manaus, Amazonas	Aleixo, Manaus, Microrregião de Manaus, Região...	(-3.0872605, -59.9900635, 0.0)	-3.087261	-59.990063	0.0
2	Alvorada	Centro-Oeste	553.18	76392	11681.73	18193	Alvorada, Manaus, Amazonas	Alvorada, Manaus, Microrregião de Manaus, Regi...	(-3.0758518, -60.0491264, 0.0)	-3.075852	-60.049126	0.0
3	Armando Mendes	Leste	307.65	33441	9194.86	7402	Armando Mendes, Manaus, Amazonas	Armando Mendes, Manaus, Microrregião de Manaus...	(-3.0940003, -59.9432246, 0.0)	-3.094000	-59.943225	0.0
4	Betânia	Sul	52.51	1294	20845.55	3119	Betânia, Manaus, Amazonas	Betânia, Manaus, Microrregião de Manaus, Regiã...	(-3.1330914, -59.9955771, 0.0)	-3.133091	-59.995577	0.0

We do not have use for all these columns, though, so we will drop some of them.

	Neighborhood	Zone	Area	Population	Full_Address	Latitude	Longitude
0	Adrianópolis	Centro-Sul	248.45	10459	Adrianópolis, Manaus, Amazonas	-3.101697	-60.008975
1	Aleixo	Centro-Sul	618.34	24417	Aleixo, Manaus, Amazonas	-3.087261	-59.990063
2	Alvorada	Centro-Oeste	553.18	76392	Alvorada, Manaus, Amazonas	-3.075852	-60.049126
3	Armando Mendes	Leste	307.65	33441	Armando Mendes, Manaus, Amazonas	-3.094000	-59.943225
4	Betânia	Sul	52.51	1294	Betânia, Manaus, Amazonas	-3.133091	-59.995577

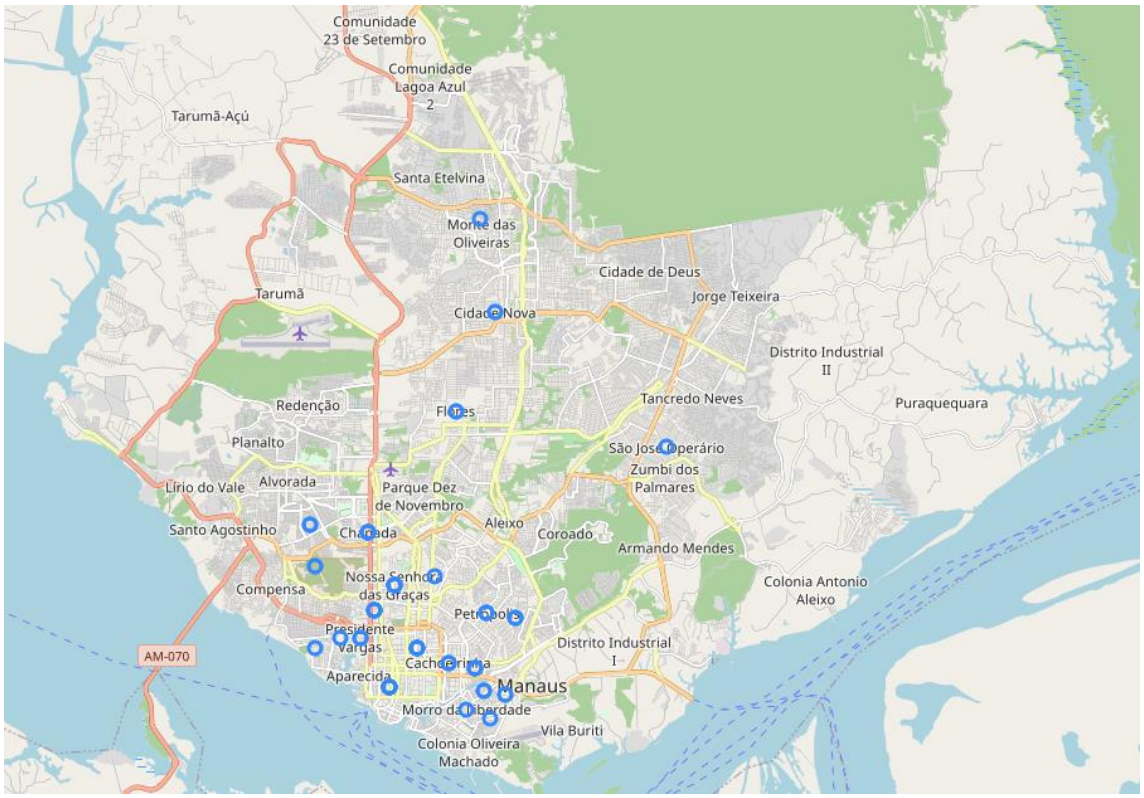
We are ready to use the Foursquare API to identify schools for each neighborhood. We created a function that searches a list of places for the given search query which is, in this case, ‘school’. This function returns a data frame with the geolocations of each identified venue:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Adrianópolis	-3.101697	-60.008975	High School Ceme	-3.109498	-60.009820	Bar
1	Adrianópolis	-3.101697	-60.008975	C.e.s.a.r School	-3.106428	-60.012660	General College & University
2	Betânia	-3.133091	-59.995577	Galileo Business School	-3.133882	-59.990531	
3	Betânia	-3.133091	-59.995577	English Winner - School	-3.137728	-59.991000	University
4	Betânia	-3.133091	-59.995577	Galileo Business School	-3.125274	-59.995067	College Academic Building

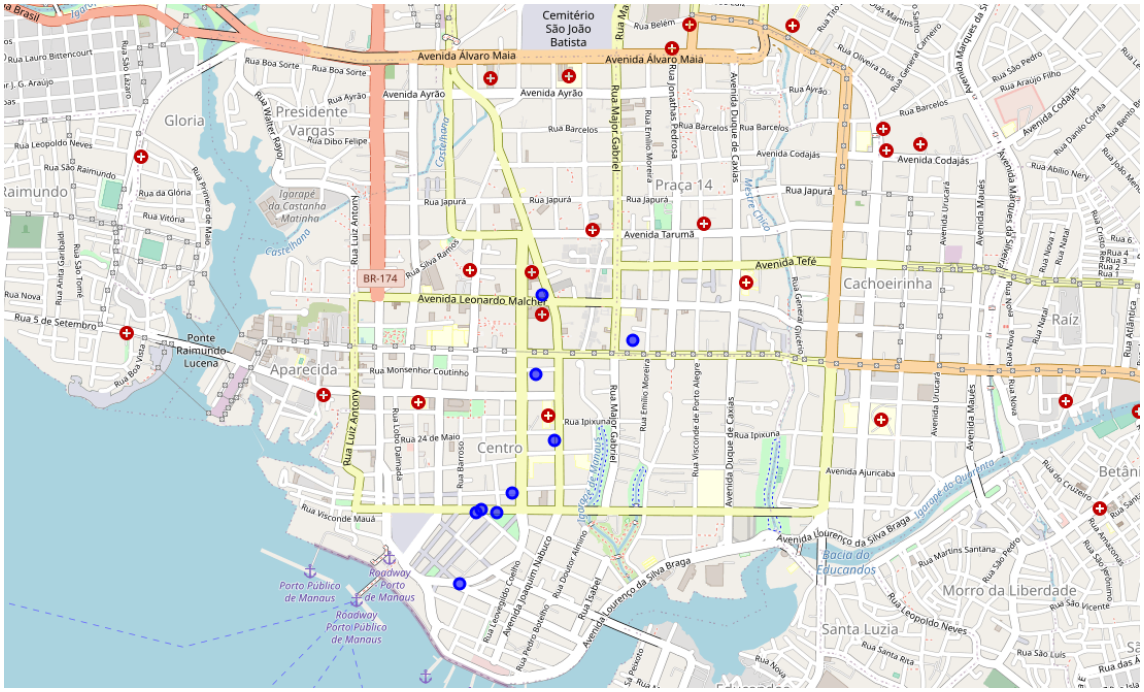
Observe that even though some venues have the word *school* in the title, they belong to a category different than the one we want. If we look at the values in this column, we notice values that do not fit with the education category; other venues have no category identified, so we will drop these entries. We can now look at the remaining categories:

```
College Academic Building          14
Student Center                     10
School                              6
University                          6
High School                         5
Trade School                        4
General College & University        4
College Communications Building     2
College Classroom                   2
College Arts Building               1
Community College                   1
Language School                     1
Office                              1
Private School                      1
Design Studio                       1
Name: Venue Category, dtype: int64
```

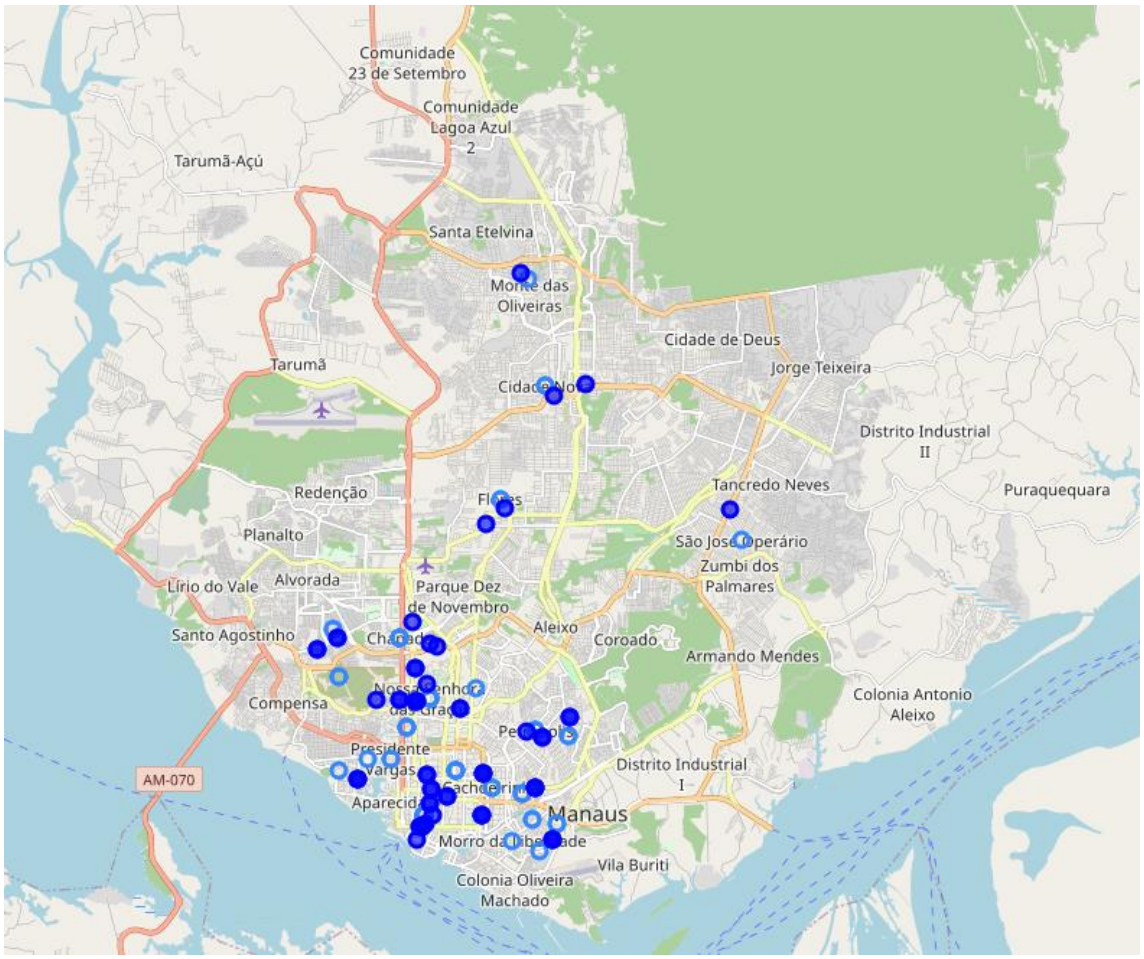
We are now ready to visualize the neighborhoods using the Foursquare API. We first look at neighborhoods that have one or more schools in a radius less than 1 km.



If we verify which neighborhood has the highest number of schools available, we find Centro, which stands for downtown in English. This neighborhood has nine schools available in the radius of 1 kilometer.



Notice some schools are closer than others. We can also look at all neighborhoods and schools combined in a single map.



The points filled with blue represent the schools, while the unfilled points represent neighborhoods.

4. Results

The main results are the answers to both questions asked initially. In other words, we accomplished our goal to use the data to find a small solution for real-world problems. There are a few possible improvements to the current approach: one of them is considering schools in the local language (escolas, in Portuguese). Another improvement could involve considering a custom radius depending on the size of each neighborhood.

5. Discussions

In the first map, we observe some neighborhoods lack nearby schools. In other words, students must commute for more than 1 km in search of educational opportunities. We also notice most schools are located around downtown. In fact, the downtown neighborhood ('Centro') is the one with the most available schools.

To answer the questions posed initially, the best neighborhoods to live in, if you want to move to Manaus and study in a nearby school, are close to the downtown. Besides, the government should establish a plan to increase the number of schools in the 40 neighborhoods lacking them.

6. Conclusions

In this project, we:

1. Explored two ways to collect the data from the Wikipedia page.
2. Identified the longitude and latitude coordinates for each neighborhood.
3. Used the Foursquare API to identify nearby schools (at most 1 km distant)
4. Visualized which neighborhoods have nearby schools.
5. Visualized the neighborhood with the greatest number of nearby schools.
6. Discussed the results.