

# Defining the best location for a new gym in the city of Campinas, Brazil

Bruno Salles  
22/12/2019

## Introduction

The city of Campinas is one of the biggest and most important cities in the state of São Paulo in Brazil. With a population over 1 million, it is the third most populous municipality in the state and the fourteenth most populous Brazilian city. The city's metropolitan area, Metropolitan Region of Campinas, is responsible for 1.8% of all Brazilian GDP and 11.4% of the State of São Paulo GDP, being Campinas the 10th richest city in Brazil, with a gross domestic product of 36.68 billion reais (2010) [1].

Due to its relevance for the local economy, the city is a popular target for entrepreneurs and companies looking for opportunities to expand their business. In this context, the current analysis aims at answering a gym owner where would be the best neighborhood to place his new gym branch focusing areas with low level of competition.

## Data Description

In order to perform the described analysis, the geographical boundaries of each neighborhood were needed, as well as demographic data. This information is available at the city's data portal [2] under the "PD2018 Unidades Territoriais Básicas (UTB) e Unidades Territoriais Rurais (UTR)" tab in the form of a shapefile (.shp extension), which contains the population and demographic density of each region as well as the geometric data that defines its boundaries. In order to be handled by the folium library, the .shp format needed to be first converted into GeoJson file. Additionally, the geospatial data provided by the city portal used the coordinate system SIRGAS 2000 UTM 23 S (EPSG:31983) which needed to be converted to the Latitude and Longitude coordinate system (EPSG:4326) which can be interpreted by the folium library. For both file extension and coordinate system conversion, the website OGRE [3] was used, resulting in a JSON file that could be imported in the python code.

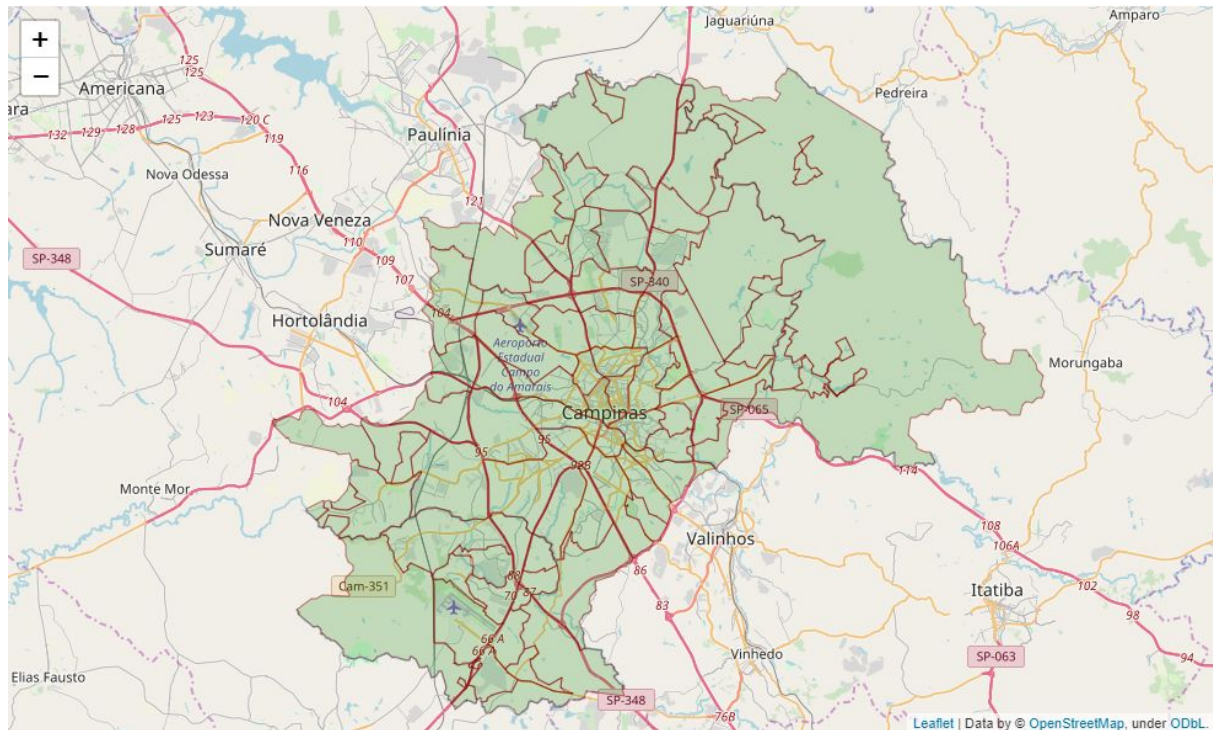
The Foursquare API was used to get the number and type of venues in each neighborhood from Campinas. This data enabled also the classification of the regions into clusters according to its venues, helping to target at the most promising neighborhood to place the new gym facility.

The github directory provided below contains the raw shapefile downloaded from the Campinas city hall data portal ([pd2018\\_utbs.zip](#)) and the converted JSON file ([Campinas\\_geo\\_data.json](#)).

[https://github.com/brunobsalles/Coursera\\_Capstone](https://github.com/brunobsalles/Coursera_Capstone)

## Methodology

The JSON file created from the shapefile contains the information that defines the limits of each neighborhood in the city. This data comprises a set of latitude and longitude pairs that when connected define the contour of each neighborhoods. These contours are shown on the picture below over the map of campinas.

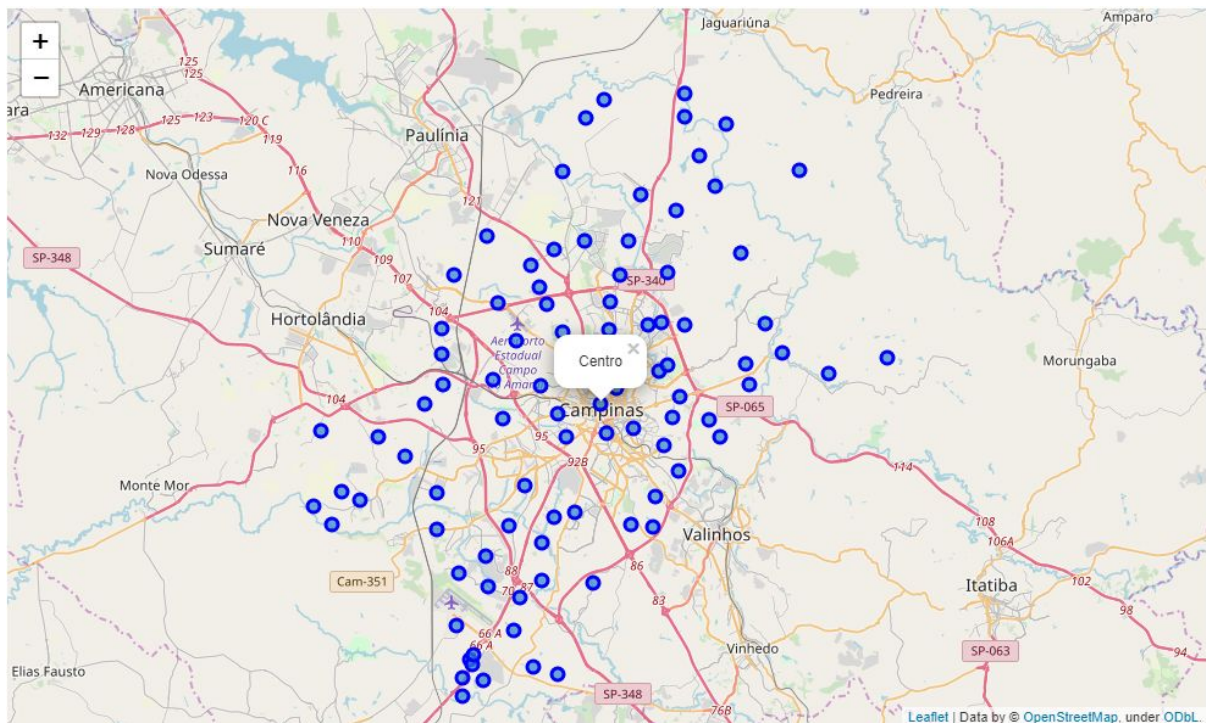


The file also contains some demographic information from each neighborhood as total population and population density. This data was used to create a dataframe as shown below, where the latitude and longitude values are the average values of the points that form the neighborhood contour.

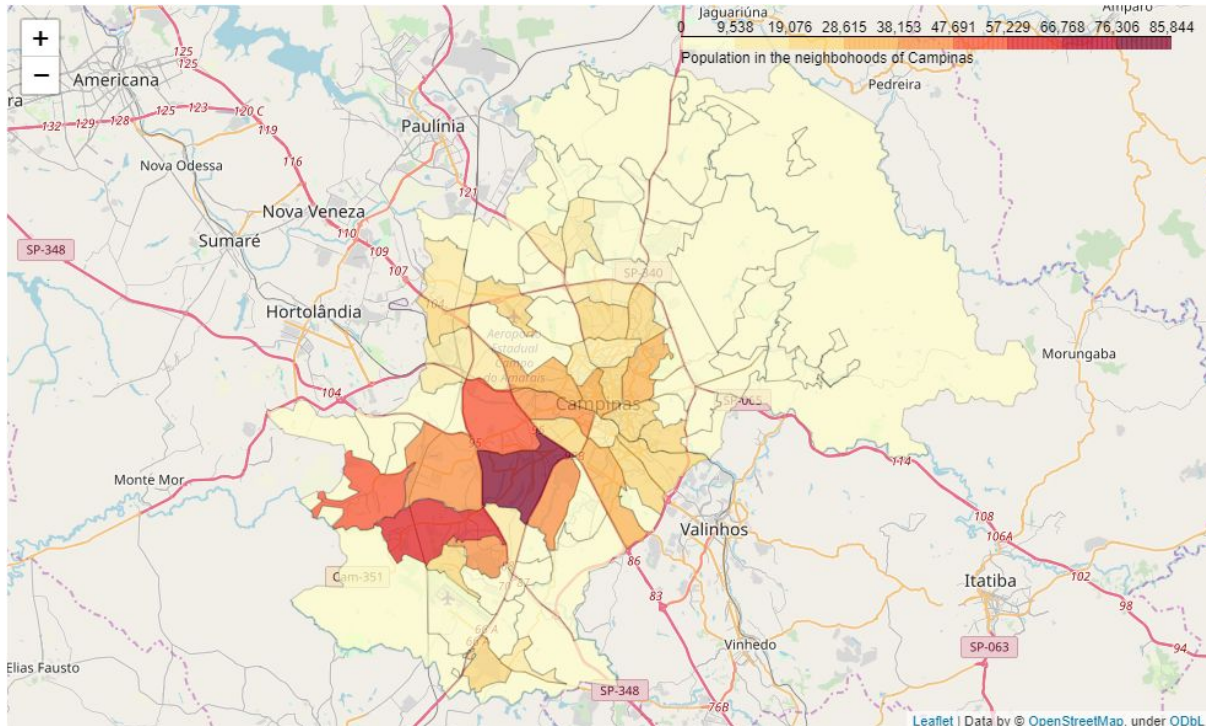
|   | Neighborhood                    | Population | Density     | Latitude   | Longitude  |
|---|---------------------------------|------------|-------------|------------|------------|
| 0 | Pq. Valenca/Pq. Itajai          | 49026.0    | 3906.320993 | -22.952906 | -47.196391 |
| 1 | Joaquim Egidio                  | 849.0      | 0.000000    | -22.886818 | -46.932617 |
| 2 | UTR - Pedra Branca              | 1883.0     | 0.000000    | -22.995581 | -47.065494 |
| 3 | Pq. Ecologico                   | 0.0        | 0.000000    | -22.909946 | -47.020844 |
| 4 | UTR - Amarais / Barao Geraldo 1 | 2664.0     | 39.063210   | -22.753775 | -47.069453 |

Using folium library each neighborhood has marked on the map, where the names are displayed in a pop-up when hovering over a marker.





With the population information contained in the GeoJSON file a Choropleth map was created, showing that the most populous areas are located around the city center and on the southwestern area.



These areas will be considered as primary targets for the new gym branch. However, the quantity of already installed gyms is an important factor to define the most promising areas. This information can be obtained with the Foursquare API. The API was used to get the

venues in each Neighborhood up to a limit of 200 and within a radius of 700 meter around the latitude and longitude coordinates of each area. The table below shows the head of the dataframe built with the Foursquare API, which contains the venues names, coordinates and venue's category.

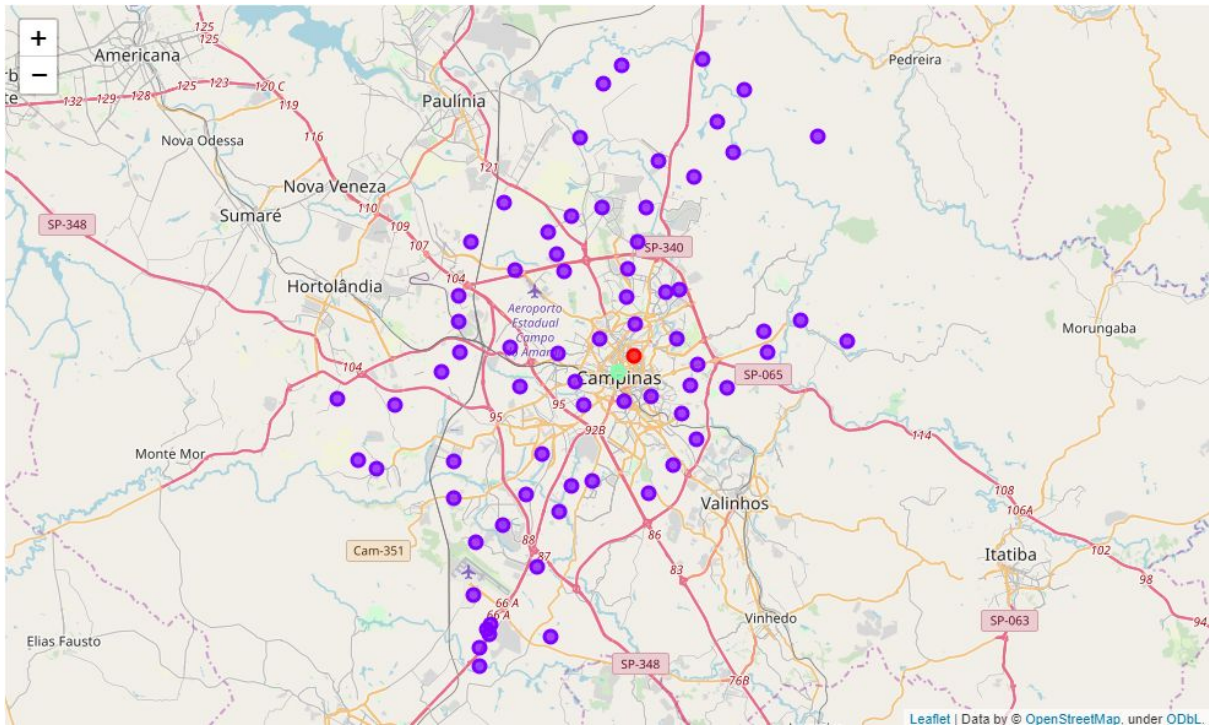
|   | Neighborhood           | Neighborhood Latitude | Neighborhood Longitude | Venue                         | Venue Latitude | Venue Longitude | Venue Category       |
|---|------------------------|-----------------------|------------------------|-------------------------------|----------------|-----------------|----------------------|
| 0 | Pq. Valenca/Pq. Itajai | -22.952906            | -47.196391             | Fran Bazar                    | -22.954944     | -47.197322      | Arts & Crafts Store  |
| 1 | Pq. Valenca/Pq. Itajai | -22.952906            | -47.196391             | Toca Do Pastel                | -22.952035     | -47.193625      | Dumpling Restaurant  |
| 2 | Pq. Valenca/Pq. Itajai | -22.952906            | -47.196391             | Tia Da Sorveteria             | -22.956122     | -47.199500      | Ice Cream Shop       |
| 3 | Joaquim Egidio         | -22.886818            | -46.932617             | Restaurante Rancho Vô Joaquim | -22.888362     | -46.934044      | Brazilian Restaurant |
| 4 | Pq. Ecologico          | -22.909946            | -47.020844             | D2                            | -22.910333     | -47.018909      | Badminton Court      |

Once the data was organized into the above dataframe, the neighborhoods were clustered into 3 groups based on the amount and type of venues using the k-means algorithm. The first rows of the resulting dataframe is shown below, where the cluster of each neighborhood is displayed along with the most common venues categories.

|   | Neighborhood                  | Neighborhood Latitude | Neighborhood Longitude | Cluster Labels | 1st Most Common Venue | 2nd Most Common Venue | 3rd Most Common Venue   | 4th Most Common Venue | 5th Most Common Venue | 6th Most Common Venue | 7th Most Common Venue | 8th Most Common Venue |
|---|-------------------------------|-----------------------|------------------------|----------------|-----------------------|-----------------------|-------------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| 0 | Aeroporto de Viracopos        | -23.017966            | -47.142582             | 1              | Airport Terminal      | Rental Car Location   | Yoga Studio             | Drugstore             | Flea Market           | Fishing Spot          | Fast Food Restaurant  | Farmers Market        |
| 1 | Alto da Nova Campinas/Gramado | -22.911052            | -46.999863             | 1              | Restaurant            | Yoga Studio           | Drugstore               | Flea Market           | Fishing Spot          | Fast Food Restaurant  | Farmers Market        | Farm                  |
| 2 | Bairro das Palmeiras          | -22.899202            | -47.016513             | 1              | Tennis Court          | Pool                  | Lake                    | Brazilian Restaurant  | BBQ Joint             | Diner                 | Restaurant            | Volleyball Court      |
| 3 | Bosque das Palmeiras          | -22.793723            | -47.038488             | 1              | Pizza Place           | Fishing Spot          | Pet Store               | Yoga Studio           | Drugstore             | Flea Market           | Fast Food Restaurant  | Farmers Market        |
| 4 | Bosque/ Jd. Proenca           | -22.915268            | -47.043093             | 1              | Bar                   | Japanese Restaurant   | Health & Beauty Service | Plaza                 | Park                  | Restaurant            | Drugstore             | Campground            |

The clusters were plotted with different colors on the city's map as depicted below. It can be seen that two clusters comprise only one neighborhood, they are Cambuí (red mark) and City Center (green mark). These neighborhoods are known by their high concentration of venues, far above the other neighborhoods. Since the target of the gym owner is to locate the new branch in an area of lower competition, these two areas would not be considered as possibilities, meaning that all other marks on the map below will be considered, except the green and red marks.



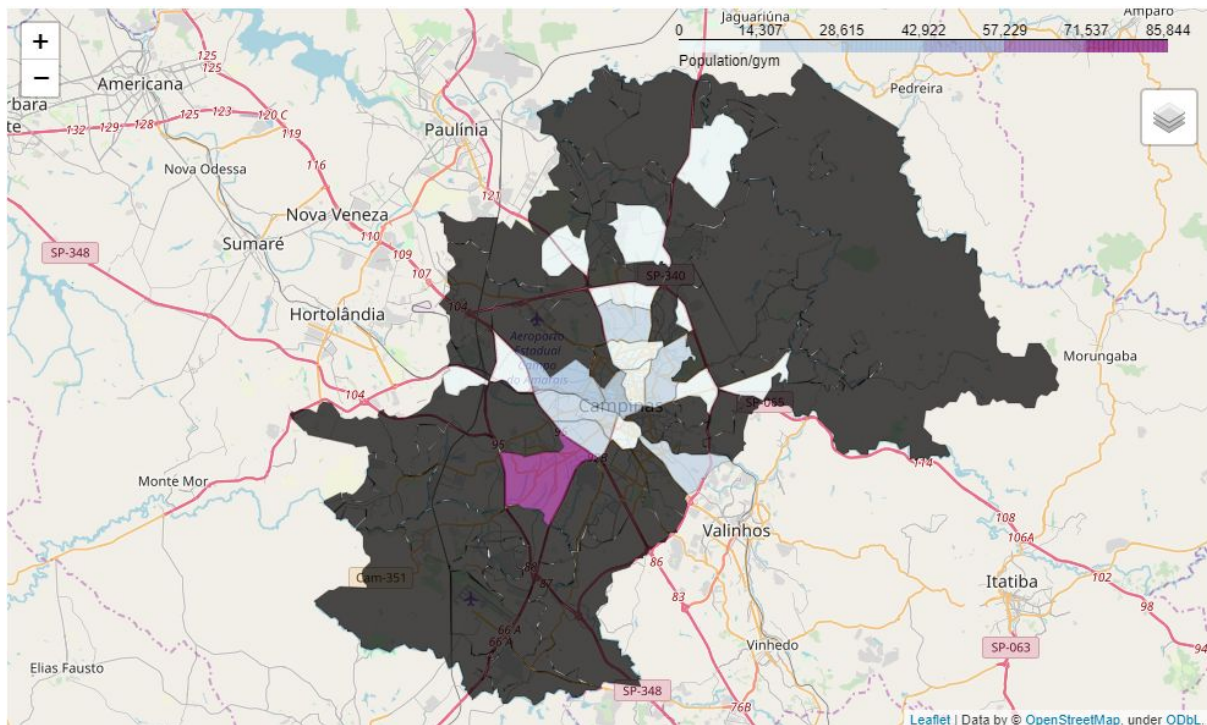


## Results

To define the best neighborhood to place the new branch, the parameter population by number of gyms was calculated by simply dividing the population of each neighborhood by the sum of venues with category 'Gym / Fitness Center' or 'Gym'. The result is displayed on the dataframe below, being the rows displayed in descending order of the parameter population/gym.

|    | Neighborhood                              | Population | Density   | Latitude   | Longitude  | Gym / Fitness Center | Gym | Total Gyms | Pop/gym |
|----|---|------------|-----------|------------|------------|----------------------|-----|------------|---------|
| 69 | Jd. Santa Lucia/ V. Uniao/ Jd. do Lago    | 85844.0    | 0.000000  | -22.945369 | -47.104134 | 1                    | 0   | 1.0        | 85844.0 |
| 58 | Centro                                    | 34961.0    | 7.578698  | -22.902441 | -47.061105 | 0                    | 1   | 1.0        | 34961.0 |
| 59 | Jd. Eulina/ Jd. Chapadao/ Bonfim          | 31037.0    | 4.447642  | -22.893458 | -47.094633 | 0                    | 1   | 1.0        | 31037.0 |
| 6  | Swift/Jd.Sao Vicente/Jd.Esmeraldina       | 25858.0    | 64.858324 | -22.937642 | -47.017600 | 1                    | 0   | 1.0        | 25858.0 |
| 25 | V. Costa e Silva/ Primavera/ Pq. Taquaral | 23896.0    | 36.744077 | -22.864328 | -47.056631 | 0                    | 1   | 1.0        | 23896.0 |

The same data can be better visualized in the form of a Choropleth map, where it is easily seen that the neighborhood Jd. Santa Lucia/V. Uniao/Jd. do Lago has the highest population per gym ratio among all neighborhoods in Campinas.



## Discussion

When calculating the Population/gym ratio, the neighborhoods that returned no venues with category 'Gym / Fitness Center' or 'Gym' resulted in an infinite value. These neighborhoods were removed from the analysis assuming that the areas with at least one gym had already proven its viability for this type of business. In the areas with no gym, a deeper market analysis would be needed in order to assure the feasibility of starting a new venue. Considering this condition, the neighborhood Jd. Santa Lucia/V. Uniao/Jd. do Lago is clearly the best place to place the new branch due its high population that is served by only one fitness center. All the other neighborhoods have less than half of the population per gym ratio, which represents a much lower potential for the new business.

## Conclusion

The present work showed how geospatial data and Foursquare API can be used to as tool to create a business plan. All information is free and publicly available, making it a powerful resource for business owners, investors and entrepreneurs. One remark worth mentioning is that the total number of venues in the 'Gym' or 'Gym / Fitness Center' appears to be low considering the size of the city. A deeper analysis would be recommended as a second step trying to combine other data sources as class associations or market reports in order to expand the current analysis.

## References

- [1] <https://pt.wikipedia.org/wiki/Campinas>
- [2] <https://informacao-didc.campinas.sp.gov.br/metadados.php>
- [3] <https://ogre.adc4gis.com/>