

SEARCHING FOR THE BEST NEIGHBORHOODS

Vinicius Lago

August 2, 2021

Business Problem

We work in a real estate development company in the city of São Caetano do Sul, a small city near São Paulo – Brazil. We work in the area of prospecting for new land and our main goal is to find the best potential neighborhoods in São Caetano to launch our new product.

Our new product is designed for younger people and we want the that the neighborhood where the product will be launch has the profile of this audience. In our previous surveys, we could identify the following most important characteristics for our potential customers:

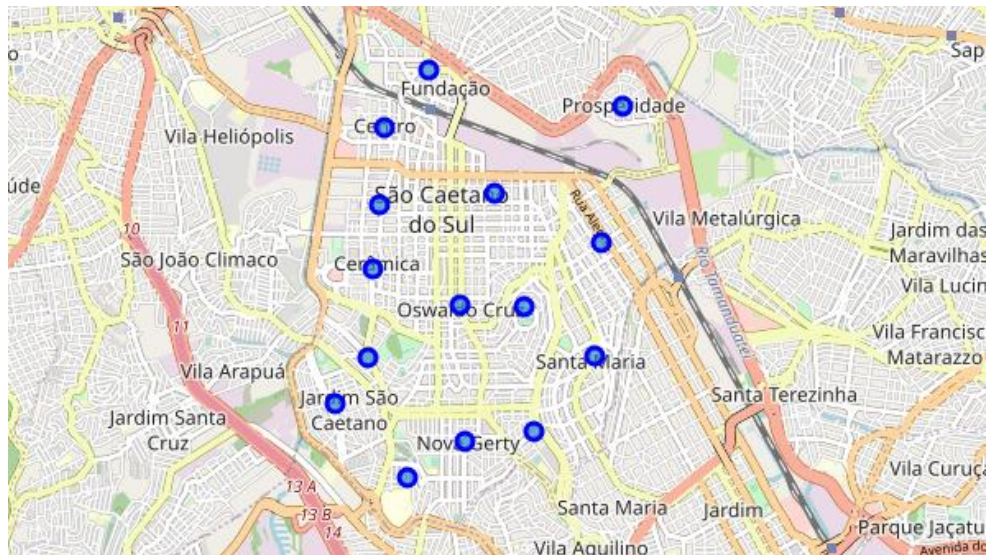
1. Restaurants;
2. Gyms.

So, we will use the geolocation of the neighborhoods of São Caetano do Sul and the data of API Foursquare to identify which 3 neighborhoods have the greatest number of these kinds of establishments.

Data

To achieve our goal, we will use 2 different datasets:

1. **Geolocation of the neighborhoods of São Caetano do Sul** - This dataset has the information of name, latitude and longitude for the 15 neighborhoods of São Caetano do Sul. The next map shows the neighborhoods centers (blue circles):



2. **API Foursquare** – This API allows you to search the information of establishments around a specific geographic position and we will use to get the information for a specific radius of neighborhoods of São Caetano do Sul. In detail, we will search for establishments within a radius of up to 500 meters from the center of each neighborhood and the API will respond to the names and categories of the establishments found. We will filter the categories of interest, like “Restaurant” and “Gym”. Lastly, we will summarize the occurrences of categories and compare the neighborhoods to identify the 3 best options. A example of the dataset that API will return is in following:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Barcelona	-23.622787	-46.552142	Feira Livre	-23.623861	-46.553367	Farmers Market
1	Barcelona	-23.622787	-46.552142	Droga Raia	-23.623749	-46.550823	Pharmacy
2	Barcelona	-23.622787	-46.552142	Old Man Sandwich Shop	-23.624427	-46.554718	Burger Joint
3	Barcelona	-23.622787	-46.552142	Sodié Doces	-23.626227	-46.553213	Dessert Shop
4	Barcelona	-23.622787	-46.552142	Meet Café	-23.620706	-46.553122	Coffee Shop

Methodology

To achieve our goal, we are going to analyze the number of establishments in the Restaurants and Gyms categories within 500 meters for each neighborhood of São Caetano do Sul.

In first step, we will filter the establishments with "Restaurant" and "Gym" categories in dataset of neighborhoods of são caetano do sul.

In second step, we will create a list for each category and order the neighborhoods according to the number of establishments for each (rank 1 = greater number of establishments / rank 15 = smaller number of establishments).

In third step, we will calculate the average of the position of each of the neighborhoods in the two lists and order them from lowest to highest average (rank 1 = lowest average position / rank 15 = higher average position).

The three neighborhoods with the best ranking will be chosen as our potential neighborhoods for the product. In case of a tie within the categories, we will use the name of the neighborhoods as key. In case of a tie the final rank, the neighborhood with the highest number of restaurants will be chosen.

Analysis

After get the Geolocation of the neighborhoods of São Caetano do Sul dataset, we use it in the API of Foursquare and we get the information about the establishments around the neighborhoods.

First, we filter the dataset using interest categories (categories that contain “Restaurant” or “Gym”).

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
6	Barcelona	-23.622787	-46.552142	O Pirata	-23.622540	-46.556518	Seafood Restaurant
7	Barcelona	-23.622787	-46.552142	Flor de Romã Restaurante	-23.619769	-46.552782	Brazilian Restaurant
8	Barcelona	-23.622787	-46.552142	Cantinho da Síria	-23.625611	-46.553610	Middle Eastern Restaurant
10	Barcelona	-23.622787	-46.552142	Academia	-23.620523	-46.548989	Gym / Fitness Center
11	Barcelona	-23.622787	-46.552142	Fit&Co Coaching Results	-23.622729	-46.552751	Gym

After this, we normalize the names of categories to just “Restaurant” or “Gym”.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
6	Barcelona	-23.622787	-46.552142	O Pirata	-23.622540	-46.556518	Restaurant
7	Barcelona	-23.622787	-46.552142	Flor de Romã Restaurante	-23.619769	-46.552782	Restaurant
8	Barcelona	-23.622787	-46.552142	Cantinho da Síria	-23.625611	-46.553610	Restaurant
10	Barcelona	-23.622787	-46.552142	Academia	-23.620523	-46.548989	Gym
11	Barcelona	-23.622787	-46.552142	Fit&Co Coaching Results	-23.622729	-46.552751	Gym

The next step is to group the neighborhoods, summarize as occurrences of the categories and order the lists according to the rules. First for the restaurants:

	Neighborhood	Number of restaurants	Rank Restaurant
0	Santa Paula	26.0	1
1	Centro	14.0	2
2	Barcelona	7.0	3
3	Cerâmica	6.0	4
4	Santo Antônio	4.0	5
5	São José	3.0	6
6	Mauá	3.0	7
7	Boa Vista	3.0	8
8	Osvaldo Cruz	2.0	9
9	Fundação	2.0	10
10	Prosperidade	1.0	11
11	Olímpico	1.0	12
12	Nova Gerty	1.0	13
13	Jardim São Caetano	1.0	14
14	Santa Maria	0.0	15

Second, for the gyms:

	Neighborhood	Number of gyms	Rank Gym
0	Santa Paula	6.0	1
1	Barcelona	4.0	2
2	Santo Antônio	2.0	3
3	Santa Maria	2.0	4
4	Olímpico	2.0	5
5	Nova Gerty	2.0	6
6	Fundação	2.0	7
7	São José	1.0	8
8	Osvaldo Cruz	1.0	9
9	Cerâmica	1.0	10
10	Centro	1.0	11
11	Prosperidade	0.0	12
12	Mauá	0.0	13
13	Jardim São Caetano	0.0	14
14	Boa Vista	0.0	15

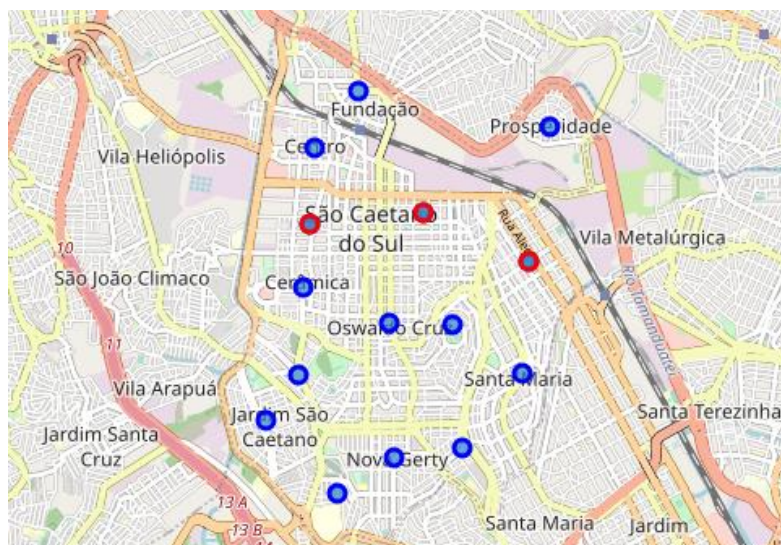
In the next step, we will join the two datasets, `scs_restaurant` and `scs_gym`, calculate the average rank and create our Final Rank.

	Neighborhood	Number of restaurants	Rank Restaurant	Number of gyms	Rank Gym	Final Rank
0	Santa Paula	26.0	1	6.0	1	1.0
1	Barcelona	7.0	3	4.0	2	2.5
2	Santo Antônio	4.0	5	2.0	3	4.0
3	Centro	14.0	2	1.0	11	6.5
4	Cerâmica	6.0	4	1.0	10	7.0
5	São José	3.0	6	1.0	8	7.0
6	Fundação	2.0	10	2.0	7	8.5
7	Olímpico	1.0	12	2.0	5	8.5
8	Oswaldo Cruz	2.0	9	1.0	9	9.0
9	Nova Gerty	1.0	13	2.0	6	9.5
10	Santa Maria	0.0	15	2.0	4	9.5
11	Mauá	3.0	7	0.0	13	10.0
12	Boa Vista	3.0	8	0.0	15	11.5
13	Prosperidade	1.0	11	0.0	12	11.5
14	Jardim São Caetano	1.0	14	0.0	14	14.0

Lastly, the table contain the information for each neighborhood and the Final Rank. As we can see, the three top neighborhoods in our study are Santa Paula, Barcelona e Santo Antônio.

The average number of restaurant per neighborhood was 4.9 and the average number of gyms was 1.6.

The next map shows the neighborhoods of São Caetano do Sul with the best 3 in highlight (red circle).



Results and Discussion

In the previous section, through the definitions made previously, we arrived at the list of potential neighborhoods and their respective positions in our rank.

The neighborhood with the greatest positive highlight was Santa Paula, which was placed first in both rankings. Santa Paula has 26 restaurants, against the average of 4.9, and 6 gyms, against the average of 1.6. As we can see on the map above (red dots), it's interesting to note that the other two top 3 neighborhoods are relatively in the same region of the city (center-north). This may reflect a concentration of points of interest in the vicinity of the city's most famous avenue, Góias Avenue, which is close to the three neighborhoods.

In the other hand, Jardim São Caetano was the worst performing neighborhood in our rank. This result was predictable for those who know the city. Jardim São Caetano is a neighborhood of big and expensive houses and, consequently, a very residential neighborhood. So, Jardim São Caetano was expected to have fewest establishments.

Conclusion

The main objective of this project was to use geographic and establishment information to identify which neighborhoods have the desired characteristics for our new product.

Using the dataset of geographic position of neighborhoods and the API of Foursquare, we were able to analyze the neighborhoods of São Caetano do Sul and rank them according our rules.

According to results showed in the previous section, we can define the Santa Paula, Barcelona and Santo Antônio as our best neighborhoods to launch our new product.