Research for a new gym in Recife, Brazil

Diego F. de Souza
October 08, 2019

Introduction

The SmartFit, the largest network of gyms in Latin America, specializing in the fitness gym, planning to open a new unity in Recife, Brazil. They want to do a market analysis to know where are the best neighborhoods to put the new unity of SmartFit

Background

Recife is the capital of the state of Pernambuco, Brazil, the city is known as a technology center for having many technology companies. The city has 94 neighborhoods, some neighborhoods are knowledge as a cultural pole and other is knowledge for the bars and restaurants.

Problem

Identify and analyze of venues so that we can decide the best neighborhoods Find neighborhoods that have gym demands but don't have 'gym fitness', type gym that the SmartFit is expert

Interest

Obviously, SmartFit would be very interested in the accurate prediction of best neighborhoods to install the new gym, for competitive advantage and business values.

Data acquisition and cleaning

Data sources

The data with name and geoposition of the neighborhoods from Recife can be found in dados.recife.pe.gov.br, a service of open data from the Recife city hall.
The venues of each neighborhood were got by the Foursquare API

Data cleaning

Only venues with the term "gym" in the category were used in the analysis at the first moment after a subgroup of venues was defined using the term “fitness” in the category as criteria.
Also, some columns were renamed to improve the match of values and the readable.

Exploratory Data Analysis

During the exploratory analysis of data, the following steps were done:

1. The geojson was download from http://dados.recife.pe.gov.br/dataset/area-urbana and a dataframe, with name “gdf”, was created using this data

```
# geojon data provided by http://dados.recife.pe.gov.br/dataset/area-urbana
!wget -q -O 'recife.json' http://dados.recife.pe.gov.br/dataset/c1f100f0-f56f-4dd4-9dcc-1aa4da28798a/resource/e43bee60-9448-4d3d
-92ff-2378bc3b5b00/download/bairros.geojson

with open('recife.json') as json_data:
    recife_data = json.load(json_data)

print("Loaded recife geojson")

gdf = gpd.GeoDataFrame.from_features(recife_data)
gdf['centroid_lon'] = gdf['geometry'].centroid.x
gdf['centroid_lat'] = gdf['geometry'].centroid.y
gdf.crs = {'init' :'epsg:4326'}

gdf = gdf.sort_values('bairro_nome')
del gdf['bairro_nome_ca']
del gdf['microrregiao']

gdf.rename(columns={'bairro_codigo': 'neighborhood_code', 'bairro_nome': 'neighborhood'}, inplace=True)

gdf.head()
```

2. Using the API from Foursquare, the venues of each neighborhood was get and converted in a dataframe, with name recife_venues, to be explored

```
recife_venues = getNearbyVenues(names=gdf['neighborhood'], latitudes=gdf['centroid_lat'], longitudes=gdf['centroid_lon'])

recife_venues.head()
```

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Aflitos	-8.04025	-34.896325	China 48 Express	-8.041123	-34.894546	Chinese Restaurant
1	Aflitos	-8.04025	-34.896325	Restaurante Ta San Yuen	-8.041105	-34.894456	Chinese Restaurant
2	Aflitos	-8.04025	-34.896325	Empório Baiano	-8.038079	-34.895796	Bar
3	Aflitos	-8.04025	-34.896325	Padaria Com.Pão	-8.041281	-34.898502	Bakery
4	Aflitos	-8.04025	-34.896325	Clube Náutico Capibaribe	-8.041406	-34.897634	Soccer Field

3. After this dataframe was filtered to get only venues with the term “gym” in the category. The new dataframe was name as recife_gym_venues.

```
recife_gym_venues = recife_venues[recife_venues['Venue Category'].str.contains('Gym')]
recife_gym_venues.head()
```

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
29	Aflitos	-8.04025	-34.896325	Academia do The British Country Club	-8.038952	-34.898006	Gym
31	Aflitos	-8.04025	-34.896325	Ciafit Academia	-8.043493	-34.895280	Gym
37	Aflitos	-8.04025	-34.896325	Fit Club	-8.040329	-34.899703	Gym
46	Aflitos	-8.04025	-34.896325	Academia One Fitness	-8.043238	-34.899501	Gym
63	Aflitos	-8.04025	-34.896325	Corpore Sano Fitness	-8.038899	-34.896104	Gym / Fitness Center

4. Using the recife_gym_venues, a filter was applied to get only the register with term “fitness” in the category. The new dataframe was name as recife_fitness_venues.

```
recife_fitness_venues = recife_gym_venues[recife_gym_venues['Venue Category'].str.contains('Fitness')]
recife_fitness_venues.head()
```

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
63	Aflitos	-8.040250	-34.896325	Corpore Sano Fitness	-8.038899	-34.896104	Gym / Fitness Center
66	Aflitos	-8.040250	-34.896325	Corpo e Movimento Studio Pilates	-8.040174	-34.892880	Gym / Fitness Center
106	Afogados	-8.077604	-34.911484	Academia Beleza e Saúde	-8.080426	-34.909182	Gym / Fitness Center
116	Afogados	-8.077604	-34.911484	Academia Master Fitness	-8.077572	-34.909122	Gym / Fitness Center
201	Areias	-8.095114	-34.931702	Fisiforma Academia	-8.098017	-34.931076	Gym / Fitness Center

5. Applying a group by neighborhood and merge between dataframe, the dataframe "datagrame_merge_gym" was created with the information of neighborhoods by number of gyms

Merge recife_gym_venues with 'gdf'

```
serie = recife_gym_venues.groupby('Neighborhood')['Venue Category'].count()
dataframe_groupby = pd.DataFrame({'neighborhood':serie.index, 'count':serie.values})

datagrame_merge_gym = gdf.merge(dataframe_groupby, on='neighborhood')
datagrame_merge_gym.head()
```

	neighborhood_code	neighborhood	geometry	rpa	centroid_lon	centroid_lat	count
0	132	Aflitos	POLYGON ((-34.8925588423082 -8.039902088598563...	3	-34.896325	-8.040250	6
1	779	Afogados	(POLYGON ((-34.91208929135482 -8.0640594534133...	5	-34.911484	-8.077604	3
2	523	Alto José do Pinho	POLYGON ((-34.91034752450803 -8.01863165508249...	3	-34.908407	-8.022161	1
3	949	Alto Santa Terezinha	POLYGON ((-34.90508640139721 -8.01277658019323...	2	-34.908097	-8.015759	1
4	493	Alto do Mandu	POLYGON ((-34.92782185721514 -8.01876640643194...	3	-34.927936	-8.021916	1

6. The same process was applying to create the "datagrame_merge_fitness"

Merge recife_fitness_venues with 'gdf'

```
serie = recife_fitness_venues.groupby('Neighborhood')['Venue Category'].count()
dataframe_groupby = pd.DataFrame({'neighborhood':serie.index, 'count':serie.values})

datagrame_merge_fitness = gdf.merge(dataframe_groupby, on='neighborhood')
datagrame_merge_fitness.head()
```

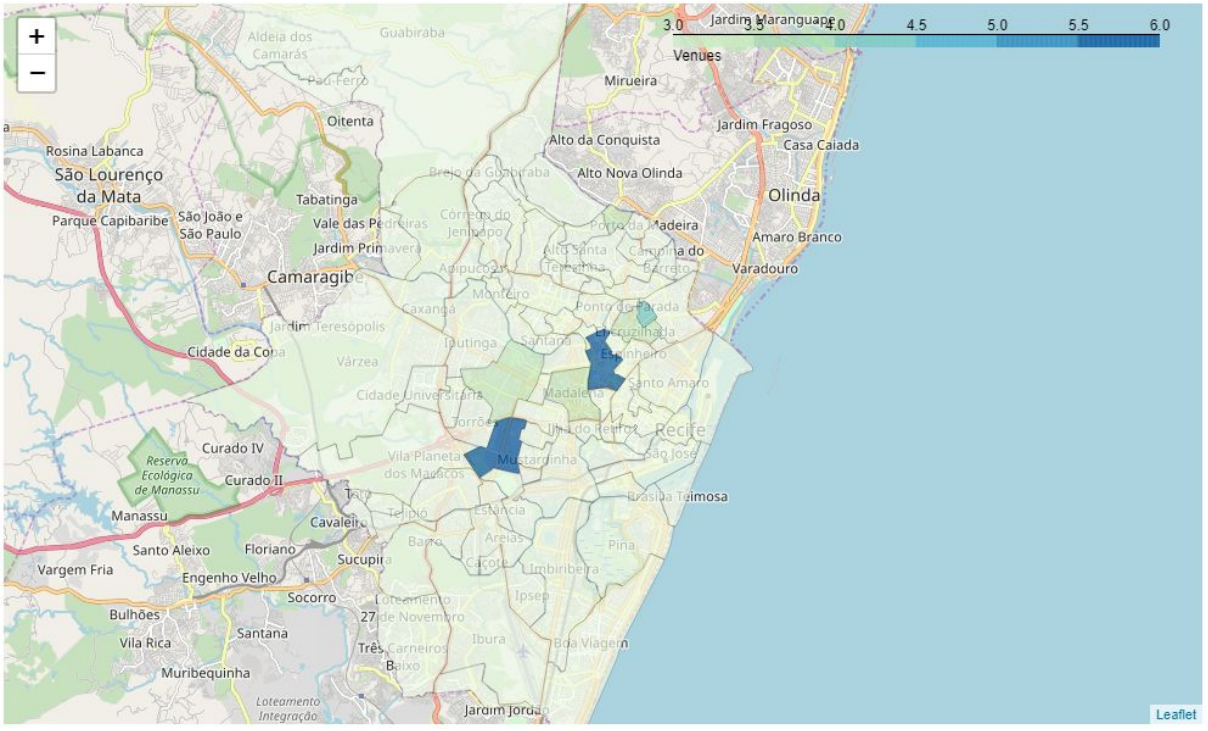
	neighborhood_code	neighborhood	geometry	rpa	centroid_lon	centroid_lat	count
0	132	Aflitos	POLYGON ((-34.8925588423082 -8.039902088598563...	3	-34.896325	-8.040250	2
1	779	Afogados	(POLYGON ((-34.91208929135482 -8.0640594534133...	5	-34.911484	-8.077604	2
2	850	Areias	POLYGON ((-34.92534411302587 -8.09271014943841...	5	-34.931702	-8.095114	3
3	205	Boa Viagem	POLYGON ((-34.90789748496555 -8.15534022816226...	6	-34.902890	-8.128004	2
4	574	Brejo de Beberibe	POLYGON ((-34.92952514601965 -7.98825822997313...	3	-34.923294	-7.998455	1

7. A final merge between the data about “gym” and “fitness” was applying with a filter of neighborhoods with 3 or more gyms

Neighborhoods with 3 or more gym but without fitness

```
dataframe_result = datagrame_merge_gym.merge(datagrame_merge_fitness,on=['neighborhood_code'],how='left')
# Select where not contains fitness
dataframe_result = dataframe_result[dataframe_result['count_y'].isnull()]
dataframe_result = dataframe_result[dataframe_result['count_x'] >= 3]
dataframe_result.head(10)
```

8. In the end, a map was created to display the neighborhoods with a highlight the inter target areas



Results Section

The map "Gyms by neighborhoods" show that the gyms are grouped in the center of the city. In the map “Fitness by neighborhoods” show that the 'gym fitness' are distributed.

Using more data like, per capita income and the number of people by region is possible to refine the research.
Other types of information are important are know if other company tried to install a gym in the neighborhoods and the profile of the residents of each region

Conclusion

The grouped gyms can be interpreted as an opportunity to offer a new type of Gym and conquer the people. Another opportunity is to place the new gym in the south of the city where the large area will avoid concurrence.