Best place to open new Gold's Gym

Oliveira, L. T.

June 29, 2020

1. Introduction

1.1. Background

The Gold's Gym is, maybe, the most famous chain gyms in the world, where many actors and famous people use to work out, for many persons say there is considered a temple of the worldwide bodybuilder. More especially, the unity which we want to study is located in Los Angeles (735 S Figueroa St Ste 100, Los Angeles, CA 90017, United States of America), where is very crowded because it has many users using the same place at the same time.

For spread and don't lose your clients, the owner had one idea: "I want to open a new gym around 500 meters the old gym!". But he knows in this neighborhood have a lot of competitors with other gyms, to decrease the change to lose clients the owner wants to know how he can open the new gym station in a location with fewer concurrent.

1.2. Problem description

There are a lot of other places in Los Angeles for service the gyms lovers, more specifically, in that neighborhood has many gyms struggling to expand their number of clients. Whether one stakeholder wants to improve his business-related gym in that neighborhood, he needs to know where is the best location for a new gym, or in other words, where is the localization has less concurrent in that neighborhood.

1.3. Interest

Obviously, gyms owners around the neighborhood, but in this case, we want to understand the Gold's Gym problem, relative its expansion and other agent interested in doing a market search about this mater in this region.

2. Data acquisition and cleaning

2.1. Data sources

The data source we'll use for solve this problem, it would acquire by using the Foursquare API developed for Python, because in that repository has a lot of information about the location of several types of location, including types of places and there are interesting data about gyms. But to acquire this amount of data, the API's user needs to create one account on the Foursquare and get access to the Development Portal (Web site URL: https://developer.foursquare.com/).

For one free account, the user has a limited amount of calls that we can access, that account allows a limit of 950 Regular API Calls per day and 50 Premium API Calls. Regular API calls only return basic firmographic data, the venue category, and a venue ID. Premium API calls return rich content such as photos, tips, menus, URLs, ratings etc [1]. For more detailed information, please come to Foursquare website, where is the documentation about that API.

2.2. Data cleaning

As mentioned above, the data source used to analyze this problem is the Foursquare API, which is received through a .json file. But, Json file is not useful directly for data analysis, for this is more suitable to use the relevant data into that file and transform in different types.

The indicated and more useable to data manipulation for Python language is Pandas package, which provides fast, flexible, and expressive data structures designed to make working with structured (tabular, multidimensional, potentially heterogeneous) and time-series data both easy and intuitive [2]. That library treats its data using a dataframe, which contains labeled axes (rows and columns), to transform Json file in Dataframe, there is function json_normalise. For example, the output data of Foursquare API, has been shown below in Figure 1.

{'meta': {'code': 200, 'requestId': '5efb947ad0415c07c6f20603'},
'response': {'venues': [{'id': '4b12d08ef964a520a78e23e3',
'name': "Gold's Gym",
'location': {'address': '735 S Figueroa St',
'crossStreet': 'at 7th St.',
'lat': 34.04866,
'lng': -118.26045,
'labeledLatLngs': [{'label': 'display',
'lat': 34.04866,
'lng': -118.26045}],
'distance': 0,
'postalCode': '90017',
'cc': 'US',
'city': 'Los Angeles',
'state': 'CA',
'country': 'United States',
'formattedAddress': ['735 S Figueroa St (at 7th St.)',
'Los Angeles, CA 90017',
'United States']},
'categories': [{'id': '4bf58dd8d48988d176941735',
'name': 'Gym',
'pluralName': 'Gyms',
'shortName': 'Gym',
'icon': {'prefix': 'https://ss3.4sqi.net/img/categories_v2/building/gym_',
'suffix': '.png'},
'primary': True}],
'venuePage': {'id': '554208029'},
'referralId': 'v-1593545835',
'hasPerk': False},

Figure 1 - Example of Json sends by Foursquare API.

Nevertheless, the useful part of this data or feature is provided by attribute "venues", whose are generated by Foursquare output. After those data had transformed to dataframe type, it allows the work with the data, to acquire the data in Foursquare API, it's necessary to put the latitude and longitude where it gets the data and in this case, the API output would get the data of the Gold's Gym location, but for this analyses, in a certain moment, this data had dropped in the dataframe.

The resultant dataframe, which has content with information of the gyms located around in radius 800 meters of the gold's gym installation, has a shape with 49 rows and 15 columns. The image with all the 49 gyms, which has searched on the API call, is shown in Figure 2.
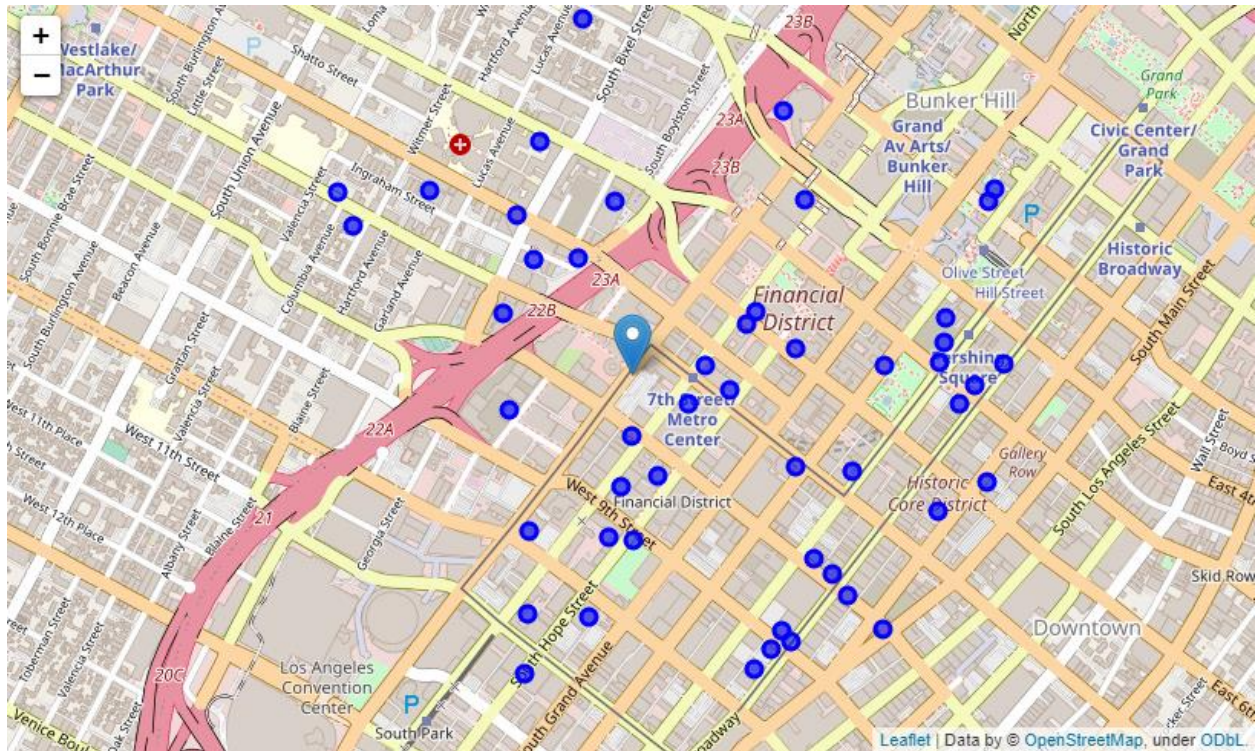
Figure 2 - Gyms around of Gold's Gym installation at 735 S Figueroa St Ste 100, Los Angeles.

3. Methodology

As mentioned above, this works is looking for the best location for a new installation of Gold's Gym, but for that, the start is clustering the location in groups using the location point. The expected result is groups of gyms with more quantity of concurrent for that new undertaking.

To continue the study, the principal objective to achieve with clustering the points in the map is to maximize the distance between groups and minimize the distance between a member of the same group. To do this, the cluster methodology chosen was K-Means, because that is an unsupervised algorithm aims to partition into k clusters in which observation belongs to the cluster to the cluster with the nearest mean, whereof has cluster centroid which serves as a prototype of the cluster.

The K-Means is iterative mythology to cluster a dataset because the initial point is important for an appropriate result, this means whether the different beginning point may cause different clusters. For more logical reality, the initial point has chosen manually based in referential cardinal points got near the localization the Gold's Gym installation which has been studied. Hence, the quantity of clusters has thought, in other words, the coefficient k was 4, it was to be consonant with the aforementioned cardinal points.

4. Results

As mentioned above, the cluster with 4 partitions and plotted in the map using the folium library in python. Those clusters are separated by color, how has been shown in Figure 3.
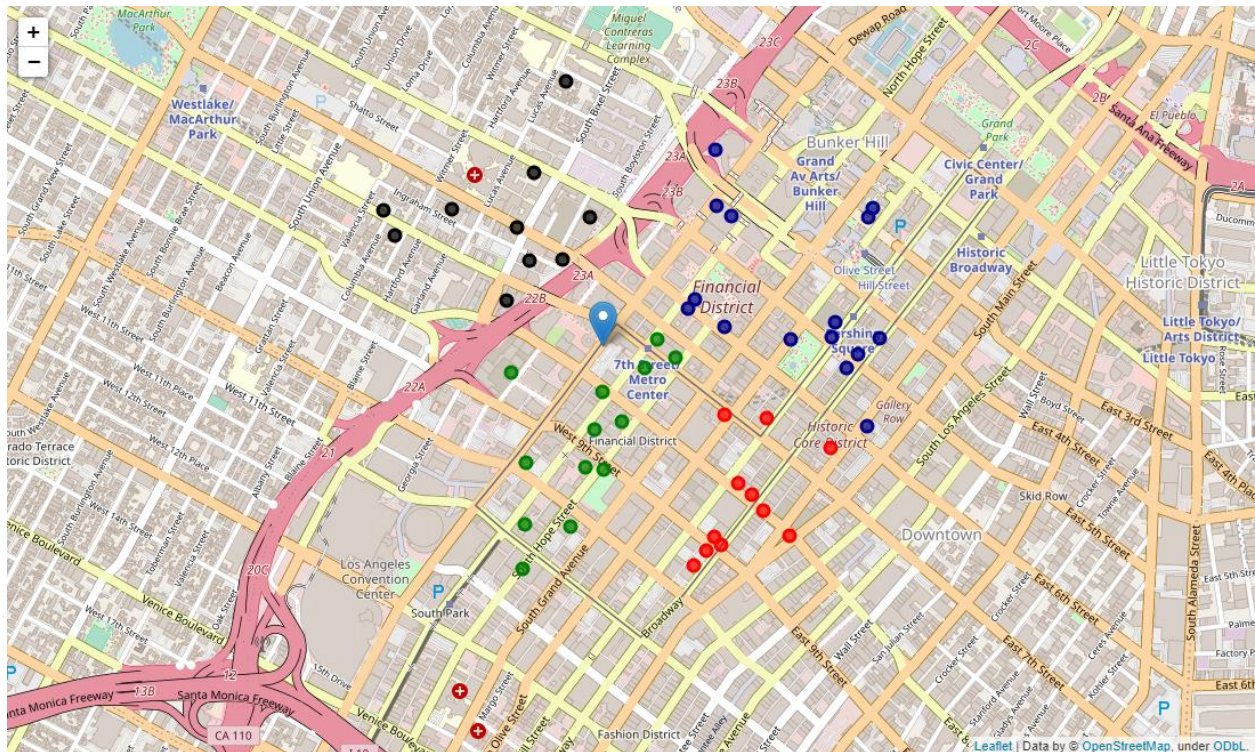

Figure 3 - Map with the 4 clusters, separated by color.

With the results in hand, we can look for the quantity in each cluster to take the decision where is the best place for new Gold's Gym installation, and that gives good

clustering the points in the map above. The dataframe with the results grouped by clusters based on each initial point can be seen in Table 1.

Table 1 - The number of gyms per cluster.

	Cluster Labels	Number of gyms
0	West	10
1	South	13
2	East	11
3	North	15

For the best understanding of the number of gyms per cluster, Figure 4 shows the same results in the radar chart.
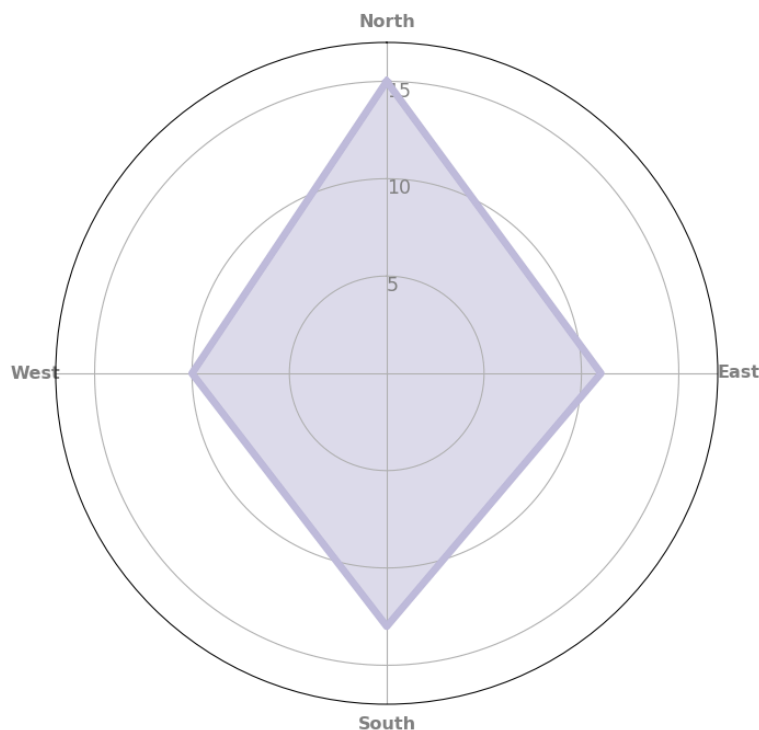

Figure 4 - The radar plot with the number of gyms per cluster.

In accordance with introduced above, it's possible to observe the number of gyms in the cluster more northerly point is the most crowded with other places with the same

approach than Gold's Gym. On the other hand, the region with has less concurrent with the establishment of new business-related with the gym is the cluster located more westerly point in the map.

5. Conclusions

Willing with results framework is more propriety to make any decision or recommendation to an expansion or possibly of changes the location of the installation for the gym that has been studied during this paper. The results shown are the best place or around the actual installation in that neighborhood is located more west, considering just the number of concurrent, this cluster can be seen in the Figure 5 above.
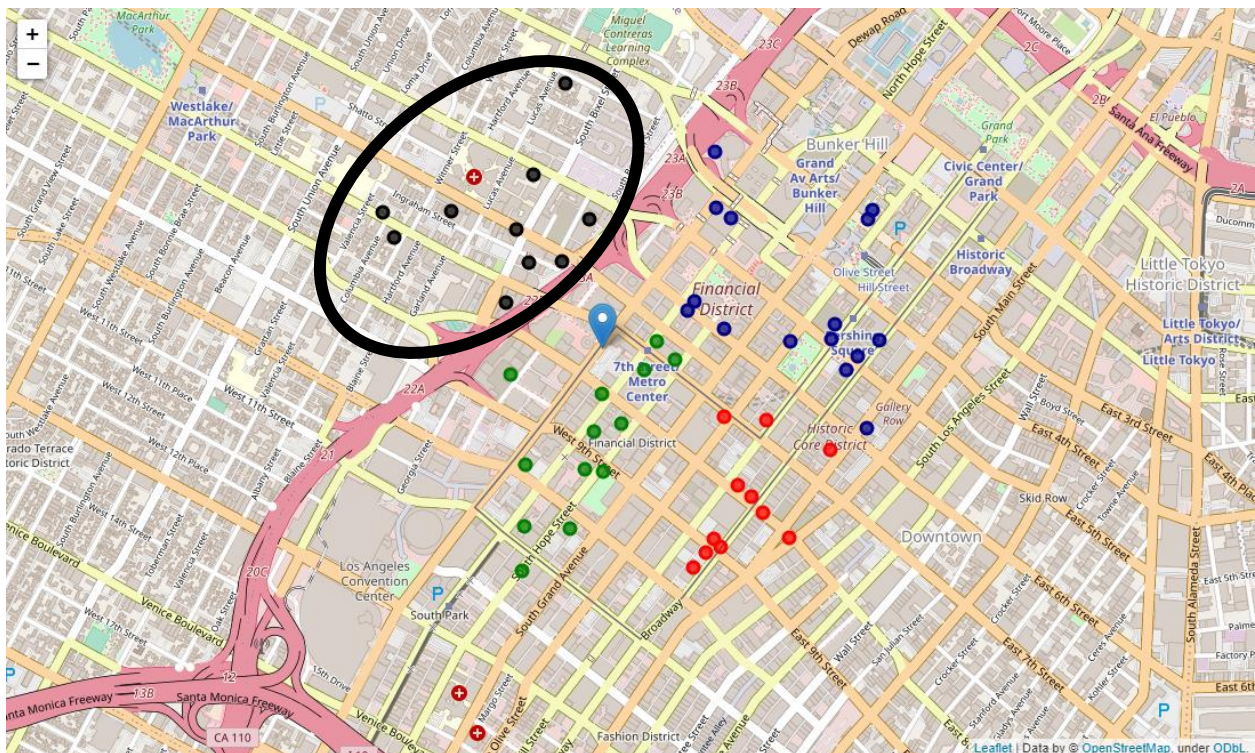

Figure 5 - The cluster with less gyms.

However, it's just an initial study to guide decisions makes because it is a quantitative recommendation based on the Foursquare search. The next step is making an interview with likely consumers for this service and looking for the viability to implement

this idea. But it's important to enhance the productivity for that future market search, thus the search marker can focus would be in the location with less concurrent.

6. Bibliography

[1] Foursquare, "Fousquare - Rate Limits," Foursquare, [Online]. Available: https://developer.foursquare.com/docs/places-api/rate-limits/. [Accessed 29 July 2020].

[2] Pandas, "https://pypi.org/project/pandas/," the pandas development team, [Online]. Available: https://pandas.pydata.org/pandas-docs/stable/. [Accessed 30 June 2020].