

Finding an Optimal Location for a Fitness Center in Dallas

Fangwei Han 2020.05.25

1. Introduction

1.1 Purpose of analysis

Dallas is one of the major cities in Texas and stakeholders are often interested in all kinds of businesses in the City area. I used to live in Dallas, and I found there were not enough gyms or fitness centers in the area. Therefore, this would be a great city for stakeholders who want to open a fitness center. In the project, I will try to help stakeholder find a perfect location to open a fitness center in the city of Dallas.

2. Data acquisition and cleaning

2.1 Data sources

The data of Dallas neighborhoods is available at: <http://peopleorganizingplace.com/know/>. However, the data format is *geojson*, which needs to be parsed before using. Foursquare will be used to obtain venue information of all the neighborhoods in the city of Dallas. The address and coordinates data of Dallas can be obtained from *geopy* package.

2.2 Data cleaning

The downloaded *geojson* data of was loaded and normalized to a *pandas* data frame with 6 columns and 375 rows. The columns are type, geometry type, neighborhood ids, neighborhood names, neighborhood description and geometry coordinates of the neighborhood boundaries. For each row, the geometry coordinates contain an array of Polygon coordinate arrays for each neighborhood. They were converted to a single pair of coordinates by calculating the mean of latitudes and longitudes, which can be used for venue exploration later.

Then, the neighborhoods were plotted on the map on Dallas with the blue circles representing the mean of their latitudes and longitudes (Figure 1).

2.3 Foursquare

Foursquare API was used to get information of gyms in each neighborhood. Since we're interested in every kind of gyms, the root category was used.

I created four functions to obtain detailed information of gyms within 500 meters of the center of each neighborhood, including venue id, venue name, venue categories, venue coordinates, venue address and their distance from the center of the neighborhood.

In summary, a total of 355 gyms were found and the average number of gyms in each neighborhood is 1.13. The gyms are shown as red circles on the map (Figure 2).

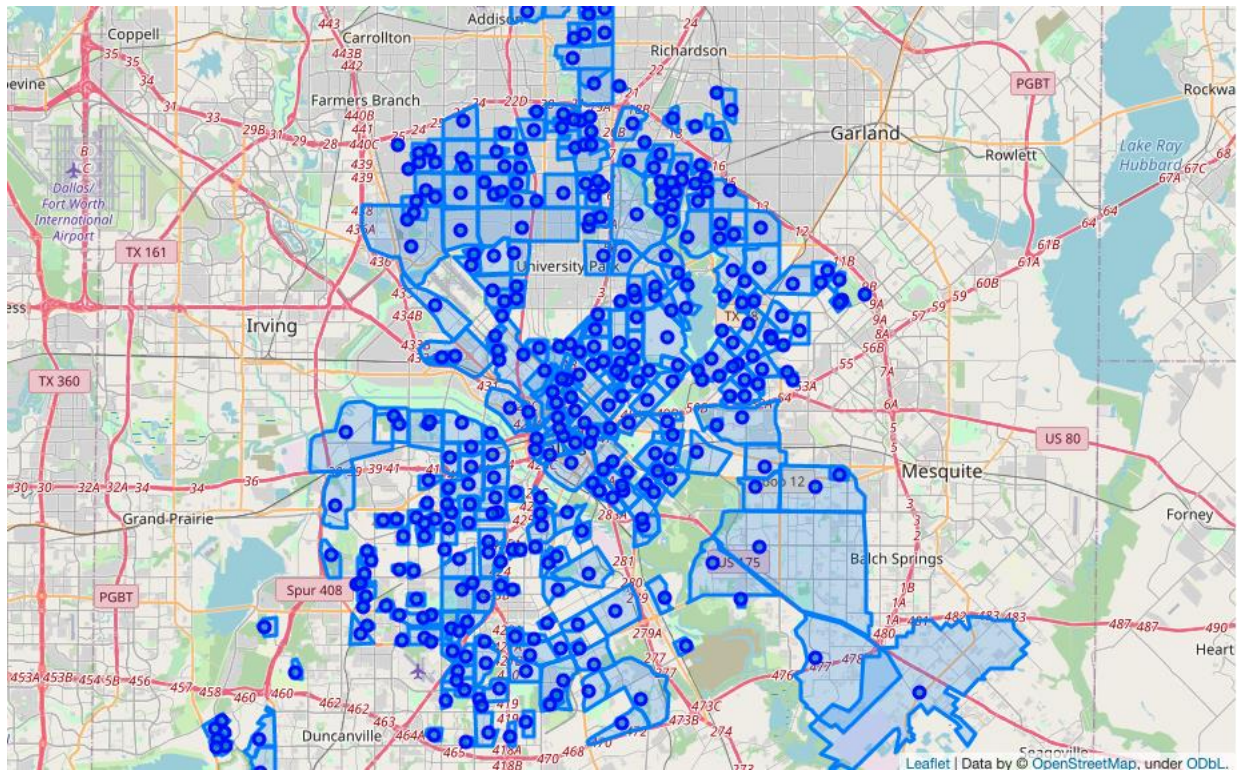


Figure 1. Dallas neighborhoods

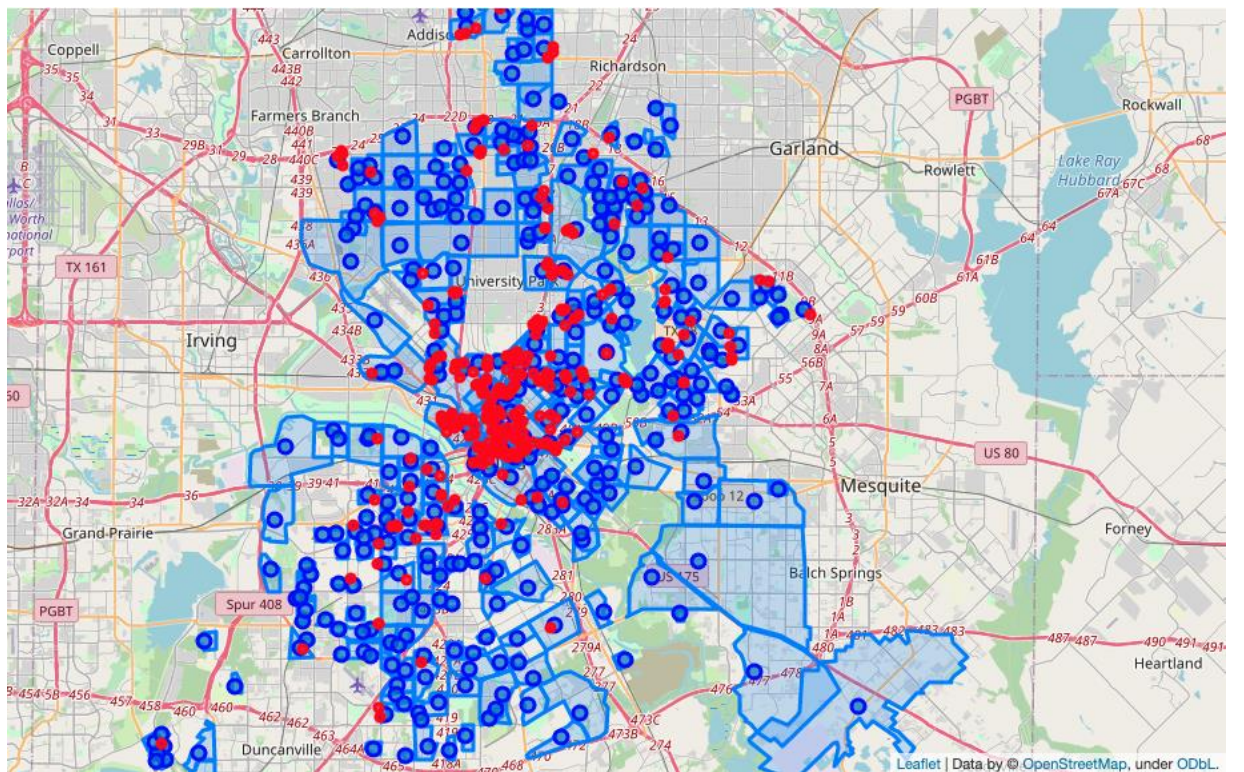


Figure 2. Gyms and neighborhoods in Dallas

3. Methodology

3.1 Heatmap

First, the number of nearby gyms was calculated for each neighborhood within a radius of 500 meters. The data frame was then filtered by eliminating those neighborhoods with more than 1 nearby gyms. The coordinates of the remaining neighborhoods were treated as “good latitudes” and “good longitudes”.

The next step was to create the heatmap with those “good latitudes” and “good longitudes” (Figure 3). The “hot” areas represent the high density of “good locations” with few gyms and the “cold” areas represent the “bad locations”. We can easily identify neighborhoods with low number of gyms by observing the heatmap.

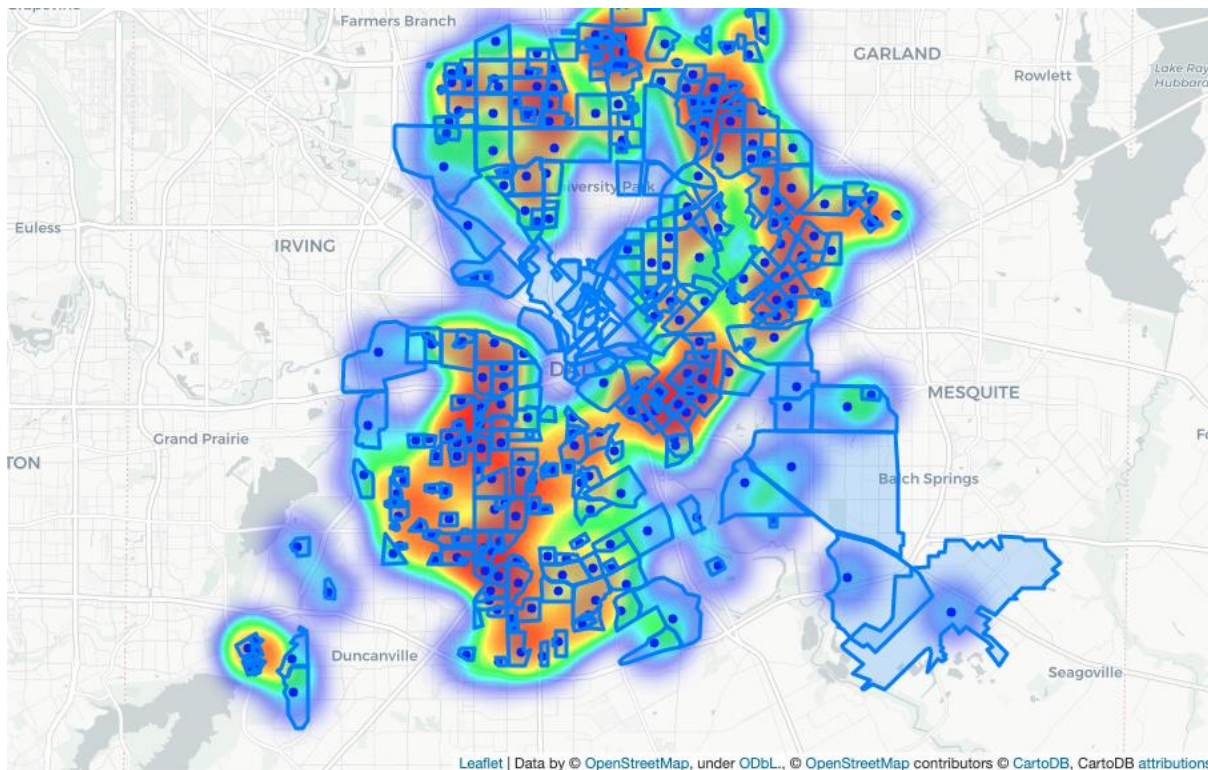


Figure 3. Heatmap of good locations

3.2 Clustering

The neighborhoods which met the requirement (nearby gyms ≤ 1) was clustered using k-means clustering. The number of clusters was set to 15 after several attempts. The coordinates of the center of the clusters were obtained and plotted on the map of Dallas. Also, a heatmap of gyms, which could be used to represent “bad locations”, was created on the same map (Figure 4).

As a result, the centers of the clusters should be considered the optimal locations for a new gym.

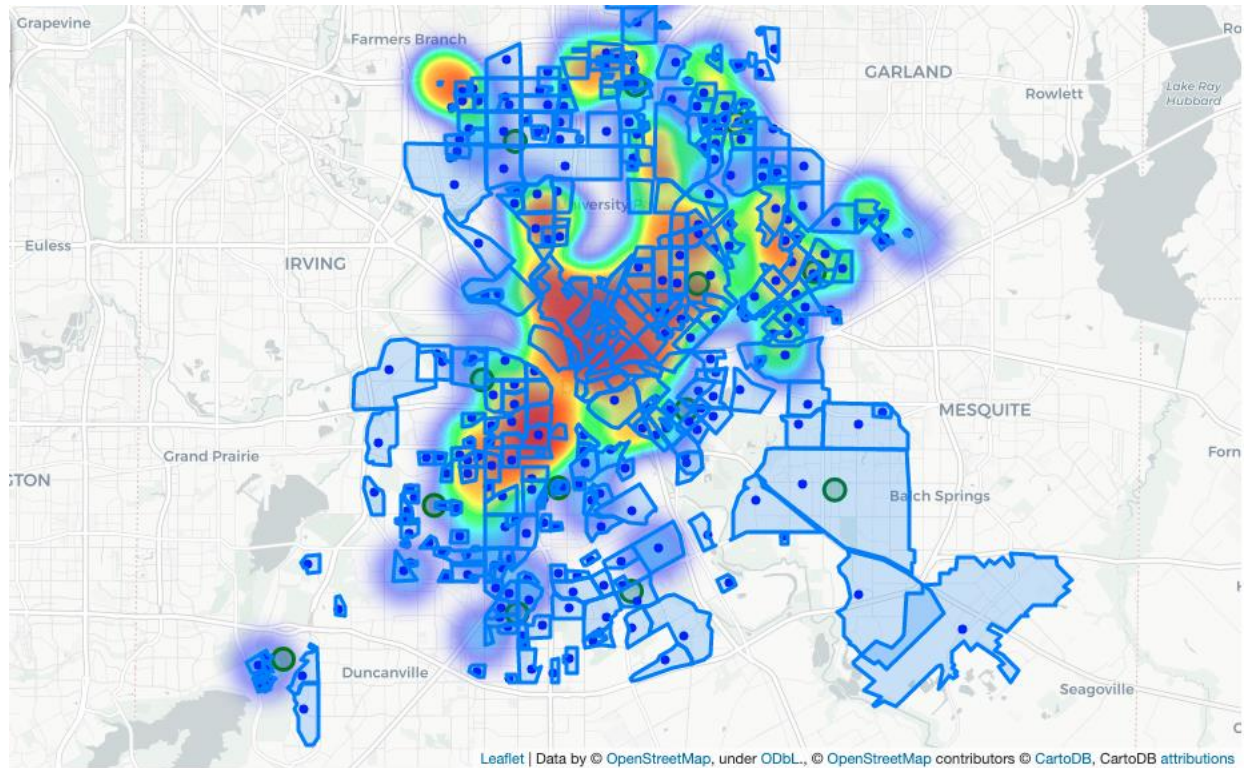


Figure 4. Clusters of good Locations and the heatmap of gyms

4. Results

The cluster centers were plotted on the map without the heatmap to provide a better view of them (Figure 5). The addresses of the cluster centers are our results and they can be obtained using the *geopy* package. The result list is shown as below:

2424, 56th Street, Fruitdale, Dallas, Dallas County, Texas, 75241, United States of America
 11606, Saint Michael's Drive, Gifford, Dallas, Dallas County, Texas, 75230, United States of America
 6396, Eagle Ford Drive, Ledbetter Hills, Dallas, Dallas County, Texas, 75249, United States of America
 9778, Twin Creek Drive, Reinhardt, Dallas, Dallas County, Texas, 75228, United States of America
 2111, Singleton Boulevard, Eagle Ford, Dallas, Dallas County, Texas, 75212, United States of America
 1331, Cedar Oaks Boulevard, Bishop Arts District, Oak Cliff, Dallas, Dallas County, Texas, 75216, United States of America
 8342, Nisqually Street, Dallas, Dallas County, Texas, 75217, United States of America
 4520, Frankford Road, Dallas, Collin County, Texas, 75287, United States of America
 3822, Poinsettia Drive, Oak Cliff, Kenwood, Dallas, Dallas County, Texas, 75211, United States of America
 3311, Rutledge Street, Fair Park, Sargent, Dallas, Dallas County, Texas, 75215, United States of America
 10125, Fieldfare Court, Meaders, Dallas, Dallas County, Texas, 75229, United States of America
 White Rock Trail, Lake Highlands, Dallas, Dallas County, Texas, 75238, United States of America
 5717, South Polk Street, Westwood Park, Dallas, Dallas County, Texas, 75232, United States of America
 6910, Echo Bluff Drive, Renner, Dallas, Dallas County, Texas, 75248, United States of America
 6651, Lakeshore Drive, Lakewood, Rawlins, Dallas, Dallas County, Texas, 75214, United States of America

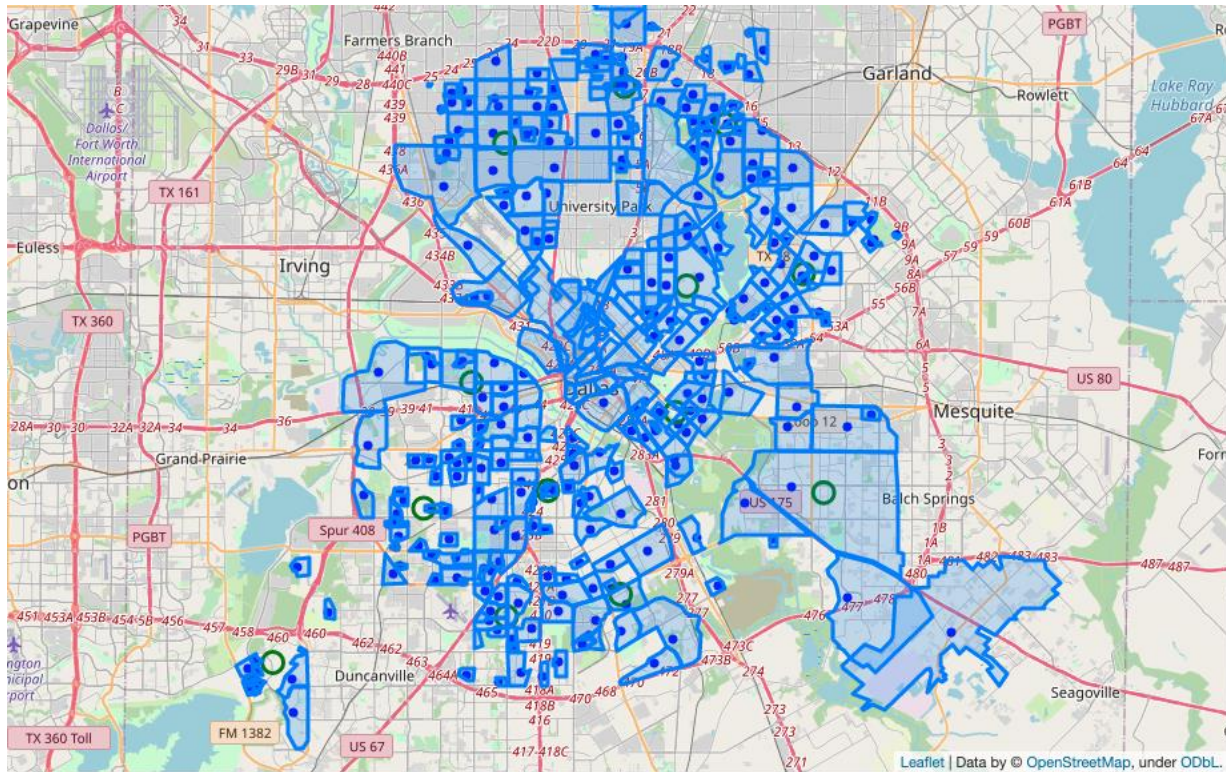


Figure 5. Clusters centers

5. Discussion

In the data cleaning step, the geometry coordinates arrays were converted to a single pair of coordinates by calculating the mean of latitudes and longitudes. However, since each neighborhood has its unique shape and area, it is possible that the mean of the coordinates cannot represent the corresponding neighborhood very well. This method was used for the simplicity of this project.

All kinds of gyms were taken into account in the Foursquare data preparation step and this included the fitness center with specific functions such as boxing centers, yoga centers, etc. To help a stakeholder who wants to open a multifunctional fitness center, these types of fitness centers should not be included in the analysis. This should be addressed in future analyses.

6. Conclusion

In conclusion, I will suggest a stakeholder to open a fitness center at one of the optimal locations in the previous list.