

# Exploring Zip Code 22030 and Its Nearby Venues to Determine Location to Open A Gym

Son Nguyen

January 15, 2020

## I. Introduction

### 1. Background

The fitness industry is growing fast since young adults nowadays are more aware of their health. According to the IHRSA, the industry has been growing by 3 – 4% annually for the last decades and it will continue to grow. Knowing the trends, fitness centers is popping up. However, many centers failed to attract customers to maintain their business because they chose bad locations to open their gyms. Therefore, it is vital to be able to choose the most favorable location to open a gym so that it will bring back profit to the investors.

### 2. Problem

A client would like to open a gym in or the Zip Code 22030 of Fairfax city, Virginia. He wants me to initially determine which area is the most favorable to open a gym. That means in depth research can be done later. He knows that there are gyms already in the area and would like to compete with them. He also wants to have an out-of-the-box location that helps him compete with other gyms.

## II. Data

**Foursquare API:** The client is rushing to have the result as soon as possible. There are no available ready-to-go data for neighborhoods in Fairfax city. Therefore, Foursquare API will be used to retrieve 100 venues of the area with the search box of 6km, their latitudes, and their longitudes. These venues then will be clustered based on their locations (latitudes and longitudes) to determine crowded locations that are popular in the city. These clusters are possible locations for the gym to be opened. After that, the top 10 common venues for each cluster will be computed to see which one is the most favorable for my client to open a gym.

**Fairfax County Zip Codes geojson:** Also, to visualize the zip code 22030, I modified Fairfax County Zip Codes geojson file using geojson.io to only includes 22030. The geojson file will be named map22030.geojson.

## III. Analysis

First, let's create a map of zip code 22030 using its latitude and longitude. The geographical coordinates of Fairfax 22030 are 38.84, -77.34 based on the website [www.unitedstateszipcodes.org/22030](http://www.unitedstateszipcodes.org/22030). Let's use folium with map22030.geojson file to map out the location. The Zip Code will be filled with the green color for easy reading.

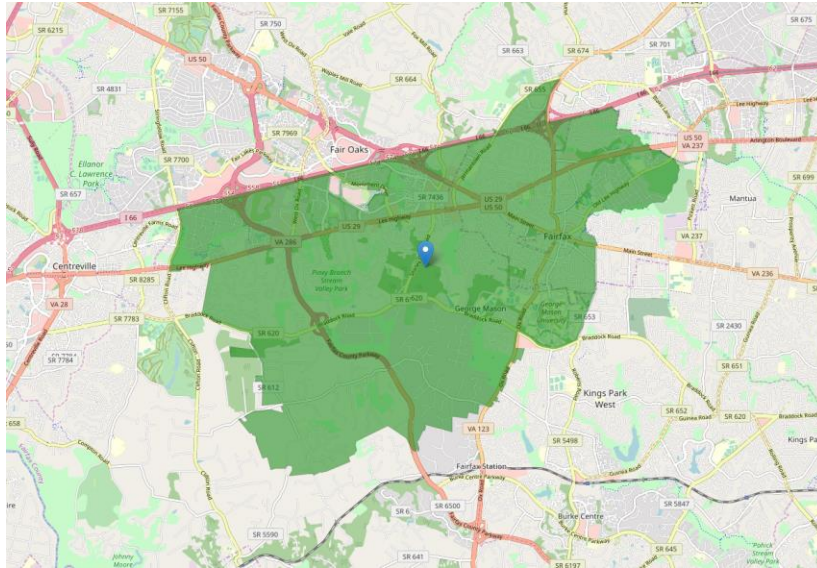


Figure 1. Fairfax 22030 Zip Code

As stated above, there are no available data for the neighborhoods with zip code 22030 of Fairfax City, Virginia. Therefore, I will use an unsupervised learning algorithm called K-means clustering to divide venues that retrieve through Foursquare API into neighborhoods of the areas. I will cluster the venues based on their locations. After retrieving and cleaning the data from Foursquare API, the result table will consist of venue names, categories, their latitudes, longitudes, and postal codes.

	name	categories	lat	lng	postalCode
0	Wegmans	Supermarket	38.851318	-77.347595	22030
1	Pad Thai	Thai Restaurant	38.851300	-77.332860	22030
2	Arigato Sushi Japanese Restaurant	Sushi Restaurant	38.851485	-77.332683	22030
3	Simply Social Coffee	Coffee Shop	38.854951	-77.334554	22030
4	Chuy's Fairfax	Mexican Restaurant	38.850999	-77.334739	22030
...	...	...	...	...	...
95	The Cheese Grille	Bar	38.857911	-77.377458	22033
96	The Auld Shebeen	Pub	38.846982	-77.306184	22030
97	Divine Nail Spa	Cosmetics Shop	38.873067	-77.373369	22033
98	Chick-fil-A	Fast Food Restaurant	38.862358	-77.387586	22033
99	Captain Pell's Fairfax Crabhouse	Seafood Restaurant	38.861090	-77.296428	22030

100 rows × 5 columns

Figure 2. Relevant Venues of Fairfax 22030

Foursquare API use the coordinates of 38.84, -77.34 we provided to search with the radius of 6km. The radius of 6km is estimated to cover all areas of the zip code. Therefore, there will be venues which are not from postal code 22030.

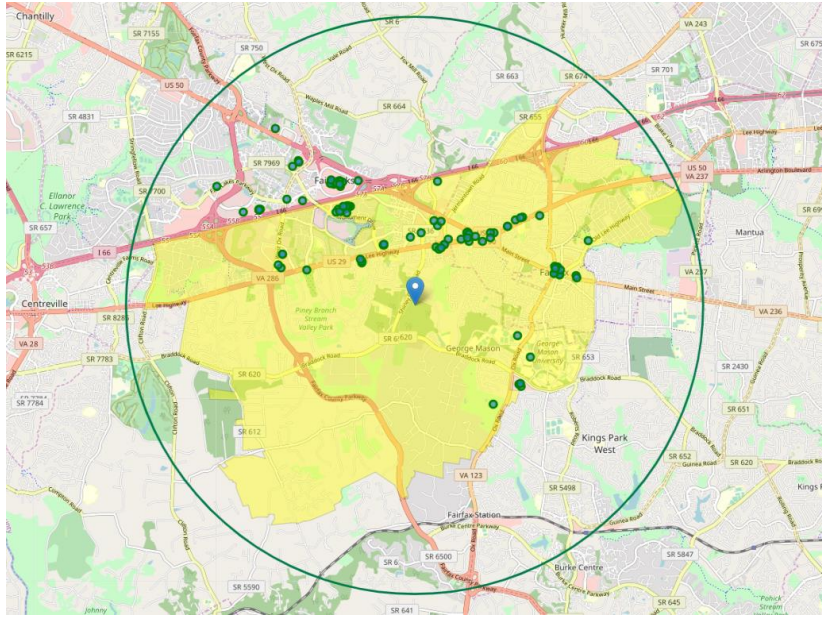


Figure 3. Fairfax 22030 With Relevant Venues And Search Circle

For this visualization, I change the color of the area of the zip code 22030 to light yellow. The blue marker is the location of zip code 22030 (38.84, -77.34) provided by [www.unitedstateszipcodes.org/22030](http://www.unitedstateszipcodes.org/22030). The green circle is the area that we use Foursquare API to search. The green dots are the venues that I retrieved.

Evaluating the map by naked eye, I definitely see there are clusters of the venues. To identify how many clusters are appropriate, I will use a method called the Elbow method.

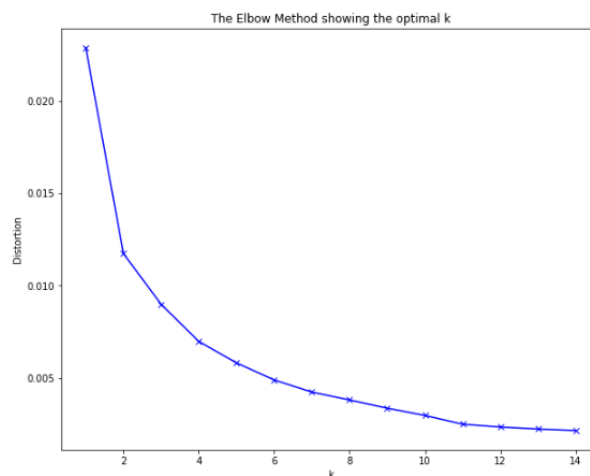


Figure 4. The Elbow Method to Select k

Based on the plot generated, I will choose  $k = 4$ . After performing clustering to the venues, here are the 4 clusters:

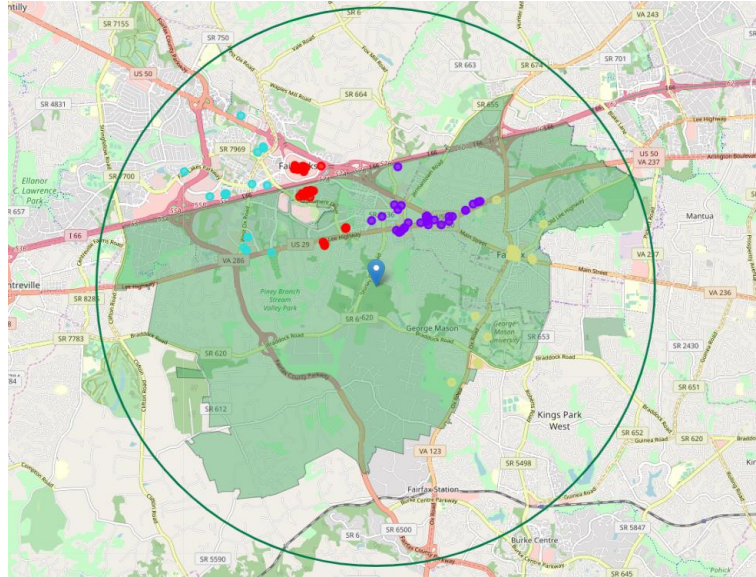


Figure 5. 4 Clusters of Venues Based on Locations.

Finally, to initially determine which cluster is appropriate for the gym to be opened, I will identify the 10 most common locations of each cluster. To do this, I encode the categories of each venue and group all venues by cluster. The values of categories will also be replaced with the frequency of their occurrences. Finally, I created a new data frame and populated clusters and its most common venues as you can see below.

cluster_geo		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	9th Most Common Venue	10th Most Common Venue
0	0	Comic Shop	Indian Restaurant	Record Shop	Beer Bar	Café	College Gym	Deli / Bodega	Donut Shop	Golf Course	Greek Restaurant
1	1	Ice Cream Shop	Mexican Restaurant	Italian Restaurant	Coffee Shop	Cosmetics Shop	Shopping Mall	Indian Restaurant	Movie Theater	Mediterranean Restaurant	Kitchen Supply Store
2	2	Mexican Restaurant	Coffee Shop	Supermarket	Café	Gym	Mediterranean Restaurant	Middle Eastern Restaurant	Music Store	Organic Grocery	Italian Restaurant
3	3	Deli / Bodega	Coffee Shop	Burger Joint	Gym	Grocery Store	Garden Center	Food Court	Fast Food Restaurant	Diner	Cosmetics Shop

Figure 6. Top 10 Common Venues of Each Cluster

Based on the table above, cluster 1 will be chosen as the initial favorable location to open a gym

#### IV. Results and Discussion

Based on the generated table, area of cluster 1 is the best place for my client to open a gym. Cluster 1 is the area where there are 2 shopping malls. Thus, we can see that the common venues of cluster 1 are mainly restaurants, shops, and theaters. People naturally got attracted by those venues because they are the places for them to relax. They are likely to go there with happy mood and no precautions. After enjoying their meals and other activities, they see the gym and they may feel guilty because of the unhealthy meals and decide to go to the gym. The gym is right next to the shopping malls so we hope that people will go there right after having meals and shopping.

For clusters 2 and 3, we can see that those are the living areas where coffee shop, supermarket, and deli/bodega are popular. People in those areas will cook meals at home. They will more likely to eat

healthier meals then when they go out. Therefore, they will not feel motivated to go out to the gym as much as the ones whose feel guilty because of having unhealthy meals.

Cluster 0 has George Mason University which already has their own gyms. it may have promotions for its students, so it is not a good idea to open a gym and compete there.

## **V. Conclusion**

I am able to determine a possible location to open a gym in Fairfax City, Virginia. However, more research should be performed to test the hypothesis that we stated in the discussion section to determine the best specific location to open the gym.