

Coursera Applied Data Science Capstone Project Report:

The Battle of Neighborhoods

by Anirban Neog

July 23, 2020

Introduction

Most professionals today care about their health and wellbeing just as much as their careers. New gym and fitness studio memberships have seen an increasing trend in the recent years. Due to having busy schedules, people tend to fit in their daily workouts in their work hours. As a result, many people seek access to a gym near their workplace. This project attempts to leverage Foursquare location data along with other publicly available data to find the most attractive locations in the Bay Area to open a new gym. This analysis will be useful for prospective entrepreneurs in the fitness space who wish to understand the competitive landscape in the Bay Area fitness industry.

Description of Data Used in the Project:

The data used in this analysis includes:

1. A list of tech companies in the Bay Area; this list is publicly available on Github and is maintained by connor11528 (connor11528/tech-companies-and-startups). I forked a version of this list to populate the leading tech companies in the Bay Area to understand where which areas of the Bay Area do tech professional work in.
2. Based on the list mentioned above, I will pull Foursquare location around the company office locations and populate a list of current gym/fitness studios. Using this data, I plan on finding out locations where setting up a new gym might be a lucrative prospect for a potential entrepreneur.

Methodology

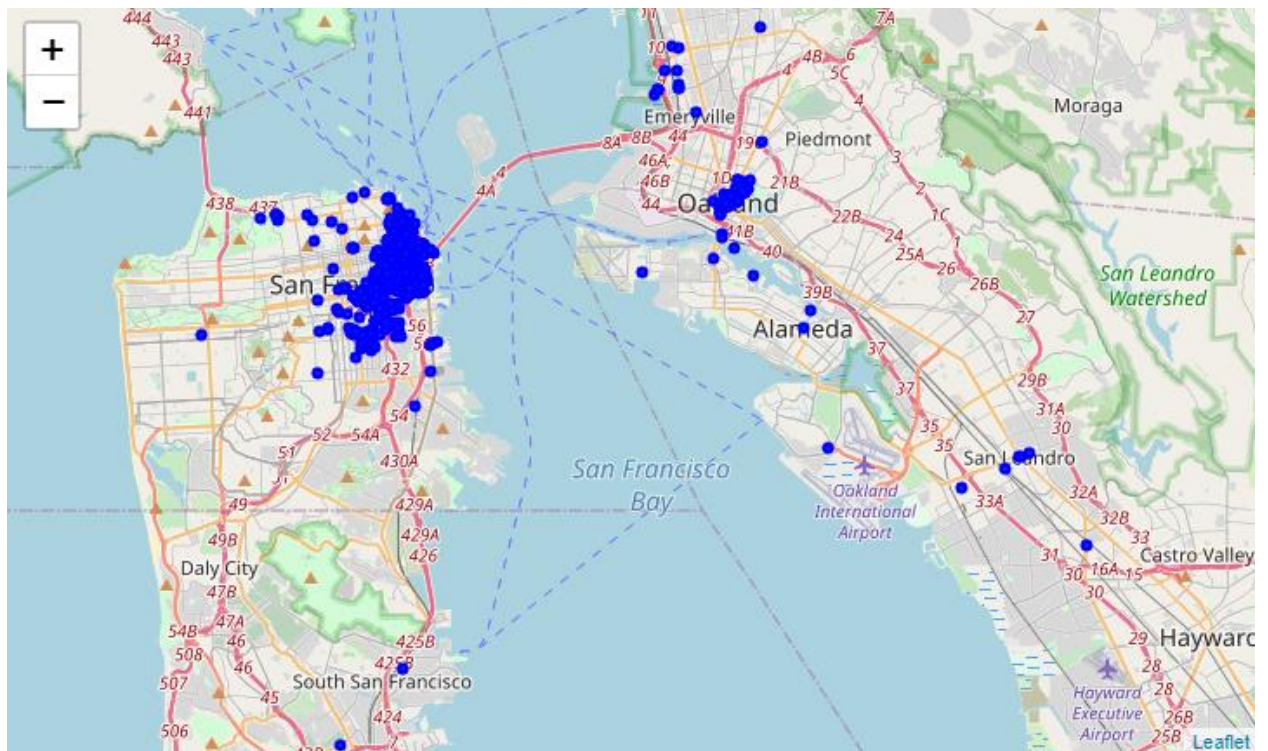
1. Extract company office location data from the Github (connor11528/tech-companies-and-startups). The following figure provides a snapshot of the data available.

Employbl Company ID	Company Name	Website	Address 1	City	State	Zip	Latitude	Longitude	Company Description	
1080	1083	Switchboard Software	switchboard- software.com	115 Sansome Street FL 7	San Francisco	CA	94104.0	37.791540	-122.401103	Switchboard enables enterprise business teams ...
1081	1084	Tata Consultancy Services	tcs.com	5201 Great America Pkwy #522	Santa Clara	CA	95054.0	37.406306	-121.976493	Tata Consultancy Services is a global leader i...
1082	1085	Cognizant	cognizant.com	500 Frank W Burr Blvd	Teaneck	NJ	7666.0	40.873268	-74.006006	See how Cognizant advances digital growth with...
1083	1086	Infosys	infosys.com	555 Mission St. Suite 1950	San Francisco	CA	94105.0	37.788690	-122.398695	Infosys is a global leader in next- generation ...
1084	1087	IBM	ibm.com	555 Bailey Ave	San Jose	CA	95141.0	37.195907	-121.748586	For more than a century IBM has been dedicated...

2. Perform feature selection to extract necessary columns from the dataset as shown below.

	Company Name	City	Latitude	Longitude
0	3scale	San Francisco	37.774634	-122.398642
1	8tracks	San Francisco	37.765227	-122.429756
2	10 by 10	San Francisco	37.774929	-122.419415
3	15Five	San Francisco	37.781714	-122.403236
4	21Tech	Oakland	37.803900	-122.270794

- Use folium to plot company office locations on the map of Bay Area.



- Use Foursquare API venue explore to explore areas around office locations.

	Latitude	Longitude	Company	City	VenueName	VenueLatitude	VenueLongitude	VenueCategory
0	37.774634	-122.398642	3scale	San Francisco	SF Flower Mart	37.775435	-122.400913	Flower Shop
1	37.774634	-122.398642	3scale	San Francisco	Bay Club SF Tennis	37.776571	-122.397948	Tennis Court
2	37.774634	-122.398642	3scale	San Francisco	Local Brewing Co.	37.776550	-122.397171	Brewery
3	37.774634	-122.398642	3scale	San Francisco	Blick Art Materials - Campus Location	37.774133	-122.399310	Arts & Crafts Store
4	37.774634	-122.398642	3scale	San Francisco	Flower Market Cafe	37.775057	-122.401115	Coffee Shop
5	37.774634	-122.398642	3scale	San Francisco	Eviva Mission Bay Fitness Center	37.773056	-122.397292	Gym / Fitness Center
6	37.774634	-122.398642	3scale	San Francisco	Belancia's Flowers & Cargo	37.775457	-122.401074	Flower Shop
7	37.765227	-122.429756	8tracks	San Francisco	Thorough Bread and Pastry	37.766408	-122.429051	Bakery

5. Filter data for gyms and retrieve location and type.

	Latitude	Longitude	Company	City	VenueName	VenueLatitude	VenueLongitude	VenueCategory
5	37.774634	-122.398842	3scale	San Francisco	Eviva Mission Bay Fitness Center	37.773056	-122.397292	Gym / Fitness Center
17	37.765227	-122.429756	8tracks	San Francisco	Alex Fitness	37.764689	-122.432183	Gym
45	37.774929	-122.419415	10 by 10	San Francisco	NEMA Fitness Center	37.775700	-122.416942	Gym
55	37.774929	-122.419415	10 by 10	San Francisco	Orangetheory Fitness	37.775826	-122.417252	Gym / Fitness Center
61	37.781714	-122.403236	15Five	San Francisco	Core40	37.780562	-122.403194	Gym / Fitness Center
70	37.781714	-122.403236	15Five	San Francisco	SKYEFIT	37.780973	-122.402667	Gym
130	37.791722	-122.402020	80Twenty	San Francisco	Equinox Pine Street	37.792036	-122.401289	Gym

6. Use OneHotEncoding to create dummy variables for VenueCategory and use it as a categorical variable.

	Boxing Gym	Climbing Gym	Gym	Gym / Fitness Center	Gymnastics Gym
3768	1	0	0	0	0
7073	1	0	0	0	0
4793	1	0	0	0	0
7029	1	0	0	0	0
613	1	0	0	0	0

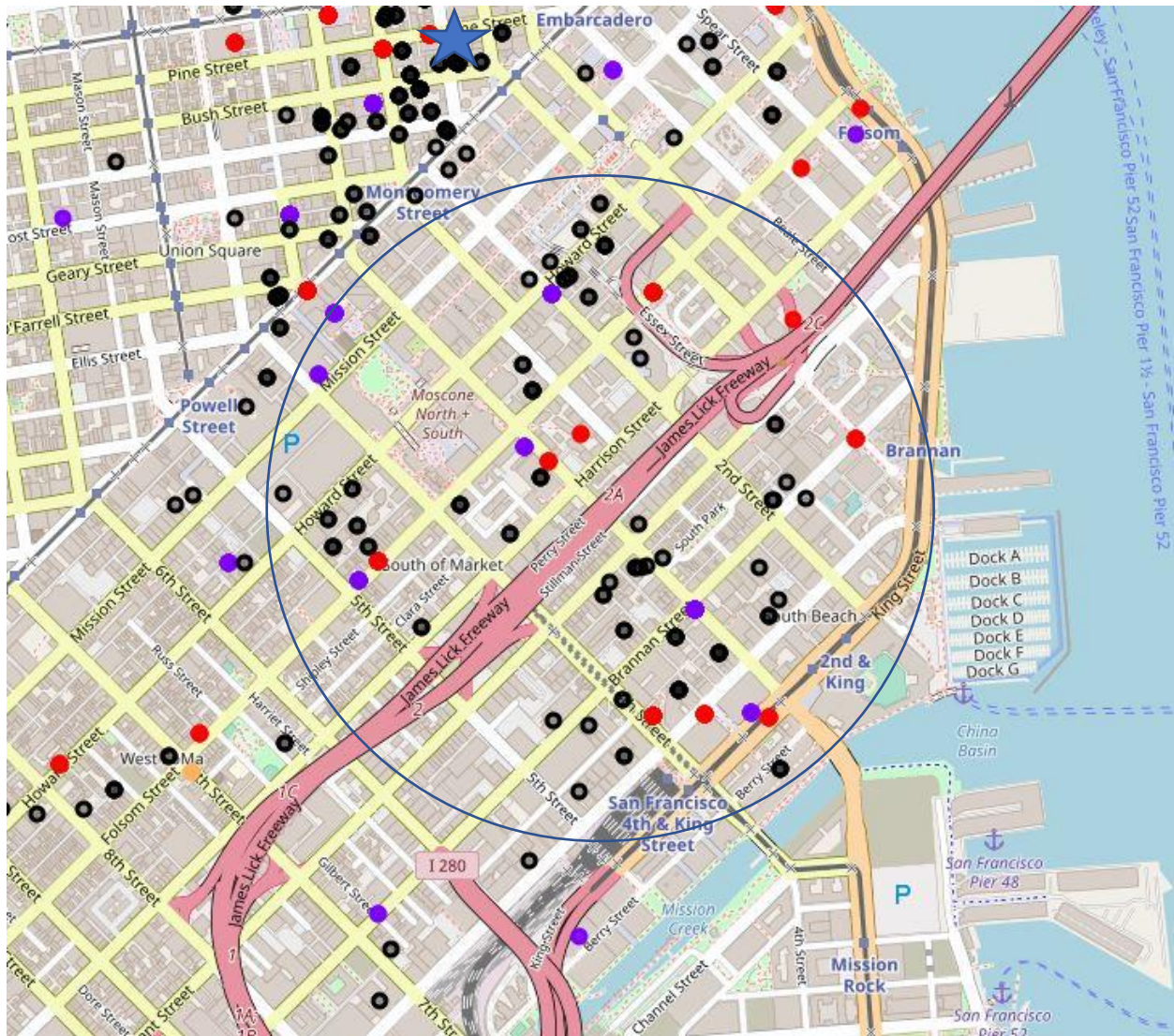
7. Use K Means machine learning algorithm to classify the different types of gyms.

	Latitude	Longitude	Company	City	VenueName	VenueLatitude	VenueLongitude	VenueCategory	Boxing Gym	Climbing Gym	Gym	Gym / Fitness Center	Gymnastics Gym
0	37.790029	-122.400840	Coinbase	San Francisco	Rumble Boxing - FIDi	37.791941	-122.400706	Boxing Gym	1	0	0	0	0
1	37.758335	-122.412322	Hoodline	San Francisco	Hit Fit SF	37.759795	-122.412651	Boxing Gym	1	0	0	0	0
2	37.791480	-122.400593	Duncan Channon	San Francisco	Rumble Boxing - FIDi	37.791941	-122.400706	Boxing Gym	1	0	0	0	0
3	37.792085	-122.399368	Holberton School	San Francisco	Rumble Boxing - FIDi	37.791941	-122.400706	Boxing Gym	1	0	0	0	0
4	37.792085	-122.399368	After College	San Francisco	Rumble Boxing - FIDi	37.791941	-122.400706	Boxing Gym	1	0	0	0	0

8. Overlay categorized gym locations on top of office locations on a Bay Area map to find pockets of opportunity.

Results and Discussion

It was observed that there are over 214 traditional gyms, 173 fitness centers, 22 boxing gyms, 2 climbing gyms, and 3 gymnastics gyms in the Bay Area in the vicinity of office locations of tech companies. One particular area of interest is shown in the map below. In the map, black circles denote office locations, red circles represent traditional gyms, blue circles represent fitness centers, purple circles denote boxing gyms, orange circles denote climbing gyms, and lime circles denote gymnastics gyms. In this area of interest, it was noticed that although there are several tech companies in the area, there are any boxing gyms yet. Hence, it might be worthwhile to evaluate the profitability potential of setting up a boxing gym in the area denoted by the circle. The nearest boxing gym to this area is denoted by a star.



Conclusion

It can be concluded from this project that Foursquare location data can be strongly leveraged to understand viability of starting new physical businesses. In this project, we looked at potential locations in the Bay Area to start a new gym location and we found an area in San Francisco where a new boxing gym might be an interesting opportunity to evaluate.