

IBM Data Science Capstone project: Battle of the neighbourhoods



Introduction

The aim of this project is to identify juice bars in the different neighborhoods of the Bay Area and to find out the best possible neighborhoods or areas where we could locate a new juice bar. The criteria we will use for this is proximity to gyms and wellness centers. Both have a very high probability of having a high percentage of customer base that might consume natural juices as part of a healthy life style.

Business case

The Juice business has been thriving over the past few years but there is still room for expansion and diversification. This element of a healthy lifestyle seems to be associated with other trends such as exercising and moving away from conventional beverages such as tea or coffee.

San Francisco has 16 gyms and other wellness businesses per square mile, according to the 2020 Wellness Index released by Mindbody, a wellness technology platform. It's indicative of millennials' priorities: They'll shell out for pricey gym memberships, boutique fitness classes, and wellness treatments.

Gyms and wellness centers have a very high probability of having a high percentage of customer base that might consume natural juices as part of a healthy lifestyle.

Therefore, this project would explore different areas of the Bay area of San Francisco where we can find these venues, current juice bars and where they are located and, after exploring the data, selecting the right neighborhood that would provide the new business venture with enough customers to create a profit.

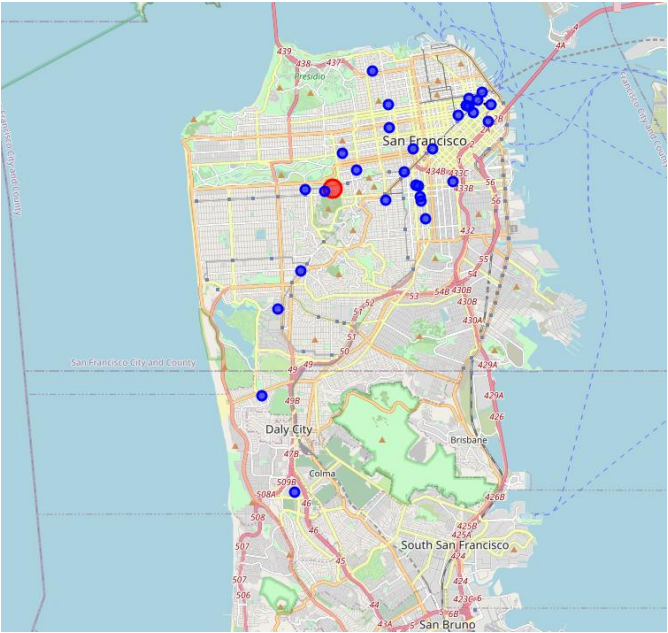
Data

The data used for the above analysis consist of data extracted from <http://www.healthysf.org/bdi/outcomes/zipmap.htm> in order to get neighborhoods and zip codes. Foursquare that contains the different venues that sell juice in the Bay area, gyms and wellness venues (see sample tables and maps below). These data was used to determine how many wellness venues are per neighborhood, comparing this

to number of juice bars and making recommendations of locations in order to set up a new juice bar business. Example of data used from Foursquare are shown below:

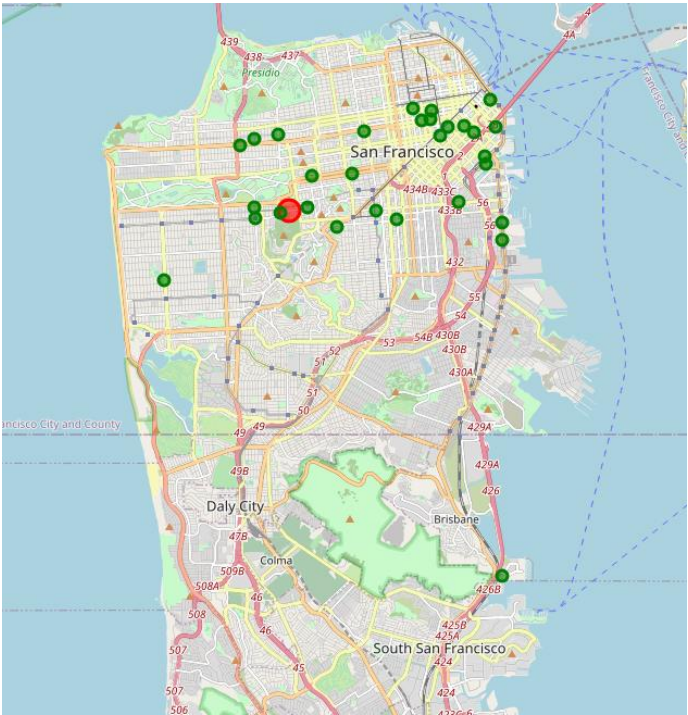
Juice selling venues

	name	categories	address	cc	city	country	crossStreet	di
0	Jamba Juice	Juice Bar	1300 9th Ave.	US	San Francisco	United States	at Irving	
1	Jamba Juice	Juice Bar	500 Parnassus Ave.	US	San Francisco	United States	UCSF Millberry Union	
2	Jamba Juice	Juice Bar	2014 Market Street	US	San Francisco	United States	at Reservoir St	
3	Project Juice	Juice Bar	790 Valencia St	US	San Francisco	United States	btwn 18th and 19th St	
4	Juice Shop	Juice Bar	432 Octavia St	US	San Francisco	United States	at Hayes St	
5	Jamba Juice	Juice Bar	2300 16th St. #245	US	San Francisco	United States	NaN	
6	JOE & THE JUICE	Juice Bar	301 Howard Street	US	San Francisco	United States	Beale St	
7	JOE & THE JUICE	Juice Bar	235 Montgomery Street	US	San Francisco	United States	Pine St	
8	Jamba Juice	Juice Bar	22 Battery Street #D	US	San Francisco	United States	at Bush & Market	
9	Jamba Juice	Juice Bar	74 New Montgomery St.	US	San Francisco	United States	near Mission	
10	Project Juice	Juice Bar	506 Castro St	US	San Francisco	United States	at 18th St.	
11	Planet Juice - Haight - ATM	Food	1467 Haight St	US	San Francisco	United States	NaN	



Gyms

	name	categories	address	cc	city	country	crossStreet
0	Mission Dolores Gym	Gym	3371 16th St	US	San Francisco	United States	Church
1	Gym	Gym	639 Geary St	US	San Francisco	United States	btwn Jones & Leavenworth St
2	Live Fit Gym	Gym / Fitness Center	953 Cole St	US	San Francisco	United States	NaN
3	Star Of The Sea Gym	College Basketball Court	345 8th Ave	US	San Francisco	United States	Geary
4	Sunset Gym	Gym	1247 9th Ave	US	San Francisco	United States	Lincoln Way
5	Gym 445	College Gym	445 Judah St	US	San Francisco	United States	NaN
6	Gym & Fitness Center	Gym / Fitness Center	1489 Webster St	US	San Francisco	United States	at Geary Blvd
7	McCullough Gym	Gym	NaN	US	San Francisco	United States	NaN
8	Live Fit Gym & Wellness Club	Gym / Fitness Center	403 Arguello Blvd	US	San Francisco	United States	Clement
9	Gym Beacon	Gym	250 King Street, 4th Floor Courtyard	US	San Francisco	United States	NaN
10	Gym	Gym / Fitness Center	NaN	US	San Francisco	United States	NaN
11	Avalon at Mission Bay Gym	Gym	383 King St	US	San Francisco	United States	at 5th St.



Methodology

1) Getting the neighborhoods in San Francisco

In order to have all neighborhoods I scraped the website

<http://www.healthysf.org/bdi/outcomes/zipmap.htm> using BeautifulSoup. The data I obtained after this was:

	Zip Code	Neighborhood
1	94102	Hayes Valley/Tenderloin/North of Market
2	94103	South of Market
3	94107	Potrero Hill
4	94108	Chinatown
5	94109	Polk/Russian Hill (Nob Hill)

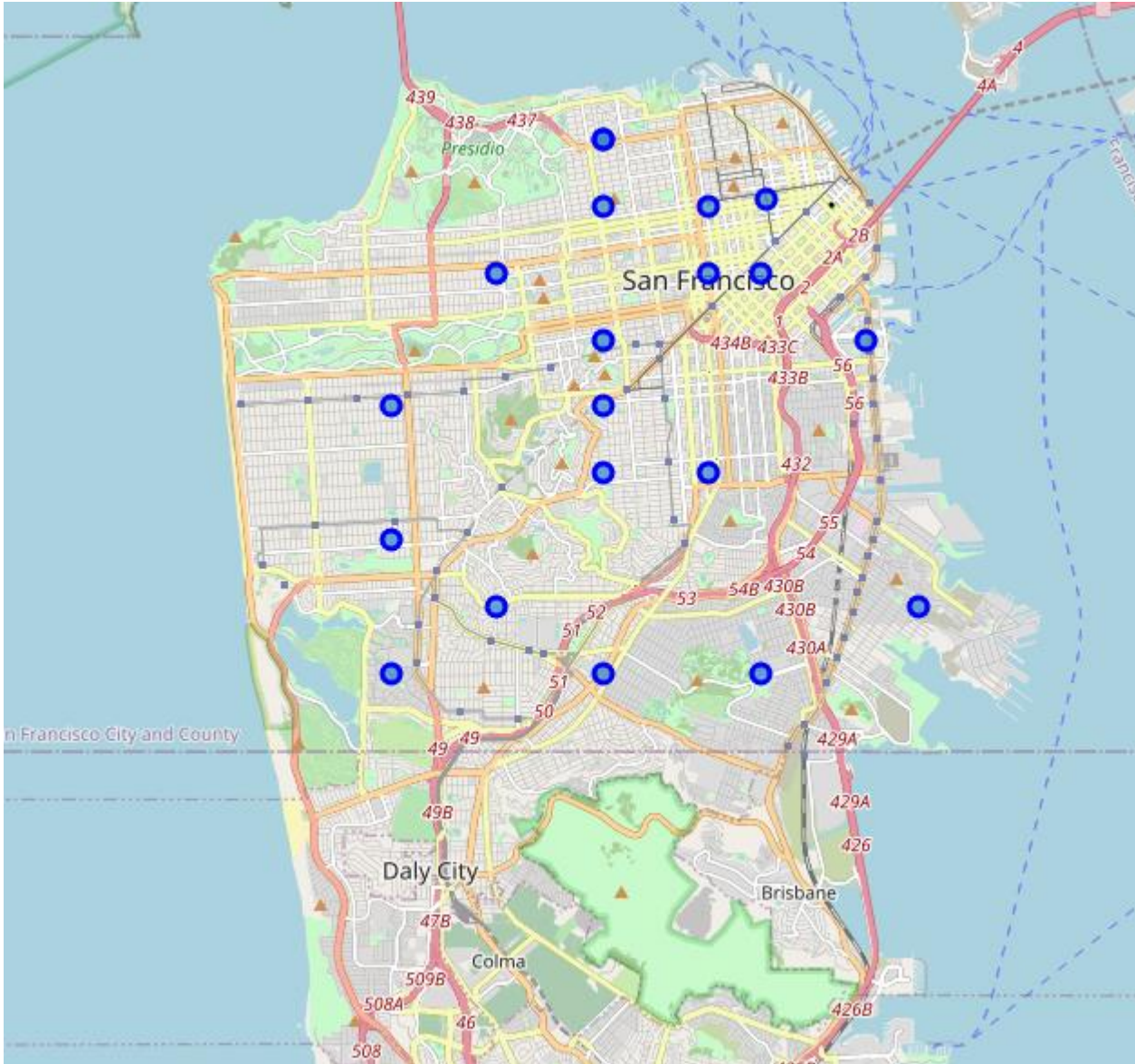
2) Assigning Latitude and Longitude values to the neighborhoods

Installing uszipcode and mapping the different zip codes to latitudes and longitudes gave me the following table:

	Zip Code	Neighborhood	Latitude	Longitude
1	94102	Hayes Valley/Tenderloin/North of Market	37.780	-122.420
2	94103	South of Market	37.780	-122.410
3	94107	Potrero Hill	37.770	-122.390
4	94108	Chinatown	37.791	-122.409
5	94109	Polk/Russian Hill (Nob Hill)	37.790	-122.420

3) Mapping in Folium

In order to have a visual of the location of different neighborhoods I mapped them in Folium with the following result:



4) Acquiring data from Foursquare

The first I did was to get all the venues nearby the lat long of neighborhoods within a radius of 500 meters.

From this general table I extracted those venues that were gyms, juice bars and yoga studios.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
126	South of Market	37.78	-122.41	LaLanne Fitness CrossFit	37.780912	-122.406661	Gym / Fitness Center
139	South of Market	37.78	-122.41	Ralph Gracie Jiu-Jitsu	37.776686	-122.411155	Gym
150	South of Market	37.78	-122.41	Raven Fitness	37.776526	-122.407658	Gymnastics Gym
182	Potrero Hill	37.77	-122.39	UCSF Bakar Fitness & Rec Center	37.768146	-122.393290	Gym
224	Potrero Hill	37.77	-122.39	Madrone Gym	37.771969	-122.389431	Gym

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
177	Potrero Hill	37.770	-122.390	CorePower Yoga	37.772415	-122.391539	Yoga Studio
293	Chinatown	37.791	-122.409	Yoga On The Labyrinth	37.791608	-122.413318	Yoga Studio
331	Polk/Russian Hill (Nob Hill)	37.790	-122.420	Bikram Yoga	37.789354	-122.420442	Yoga Studio
391	Polk/Russian Hill (Nob Hill)	37.790	-122.420	California Yoga Company	37.793768	-122.421353	Yoga Studio
437	Inner Mission/Bernal Heights	37.750	-122.420	Yoga Tree Valencia	37.753093	-122.420960	Yoga Studio

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
45	Hayes Valley/Tenderloin/North of Market	37.78	-122.42	Project Juice	37.777035	-122.422375	Juice Bar
217	Potrero Hill	37.77	-122.39	JOE & THE JUICE	37.773257	-122.391227	Juice Bar
540	Castro/Noe Valley	37.76	-122.44	Project Juice	37.760716	-122.435023	Juice Bar
832	Marina	37.80	-122.44	Project Juice	37.800367	-122.440067	Juice Bar
901	Lake Merced	37.72	-122.48	Natural Sensations	37.722021	-122.478612	Juice Bar

These are already natural clusters so there was no need to perform a k-means clustering for this analysis.

5) Analyzing Data

I grouped by neighborhood and counted the number of venues and counted the number per neighborhood:

Gyms:

	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Neighborhood						
Castro/Noe Valley	1	1	1	1	1	1
Chinatown	5	5	5	5	5	5
Haight-Ashbury	1	1	1	1	1	1
Inner Mission/Bernal Heights	2	2	2	2	2	2
Lake Merced	2	2	2	2	2	2
Marina	5	5	5	5	5	5
North Beach/Chinatown	5	5	5	5	5	5
Polk/Russian Hill (Nob Hill)	4	4	4	4	4	4
Potrero Hill	6	6	6	6	6	6
South of Market	3	3	3	3	3	3
Western Addition/Japantown	2	2	2	2	2	2

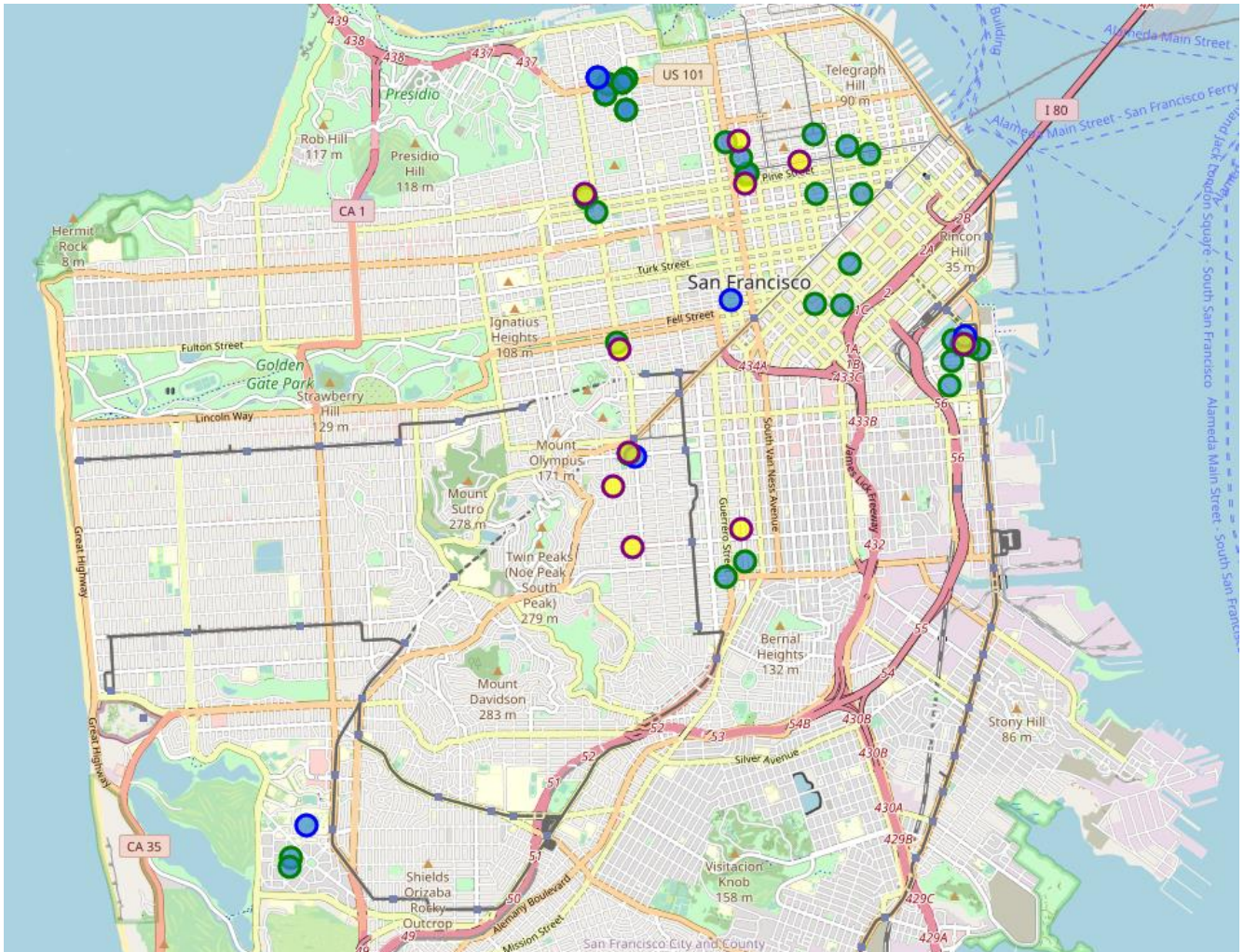
Yoga studios

	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Neighborhood						
Castro/Noe Valley	2	2	2	2	2	2
Chinatown	1	1	1	1	1	1
Haight-Ashbury	1	1	1	1	1	1
Inner Mission/Bernal Heights	1	1	1	1	1	1
Polk/Russian Hill (Nob Hill)	2	2	2	2	2	2
Potrero Hill	1	1	1	1	1	1
Twin Peaks-Glen Park	1	1	1	1	1	1
Western Addition/Japantown	1	1	1	1	1	1

Juice Bars:

	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Neighborhood						
Castro/Noe Valley	1	1	1	1	1	1
Hayes Valley/Tenderloin/North of Market	1	1	1	1	1	1
Lake Merced	1	1	1	1	1	1
Marina	1	1	1	1	1	1
North Beach/Chinatown	1	1	1	1	1	1
Potrero Hill	1	1	1	1	1	1

And mapped the data:



Yellow= Yoga studios

Green = Gyms

Blue = Juice Bars

6) Discussion

Within the constrain of the search with was venues within 500 meters of the lat long coordinates of the neighborhoods, there seem to be a genuine lack of Juice Bar compare to the number of wellness centers. In particular, Chinatown and Polk/Russian Hill have a high number of wellness centers but there are no Juice Bars nearby.

7) Recommendation

As a recommendation for a future business entrepreneur who wants to open up a Juice Bar in San Francisco Bay, the customer base seems to be big enough and the offer not yet saturated. Any are of the city seem to have potential but as pioneers in an area the data shows that ChinaTown or Polk/Russian Hill are a good place to start as the number of wellness venues is 6 in each of them with no nearby Juice Bar.