Finding the best San Francisco neighborhood to open a bubble tea shop

Brandon Lee

April 3, 2020

1. Introduction

1.1. Background

Boba (aka bubble tea) is a popular Asian drink. Similar to coffee shops, boba shops are popular gathering places for people to do work and meet friends. Based on prior city clustering analysis in Toronto and New York, coffee shops are the most popular venue. In San Francisco (SF), coffee shops are similarly popular. However, boba shops are very similar to coffee shops and boba shop owners can take advantage of the large Asian population in the Bay Area.

1.2. Problem

There are already many boba shops in San Francisco. However, the rising Asian population in the Bay Area indicates that there is still a market opportunity for additional boba shops. If I were looking to open a boba shop in San Francisco (SF), which already has many boba shops, which neighborhood(s) should I open it?

1.3. Interest

This information is valuable to bubble tea corporations (e.g. Boba Guys, Gong Cha, Chatime, Quickly Tea House) or independent, small business owners who are looking to open a bubble tea shop location.

2. Data Acquisition and Cleaning

2.1. Data Sources

There are two data sources: (1) San Francisco zip codes with neighborhoods names (from here) and (2) Asian restaurants and Bubble Tea Shop venues (from Foursquare's Venue Explore API).

2.2. Data Cleaning

The SF zip codes and neighborhoods were placed in a Pandas dataframe. According to my data source, there are 21 total neighborhoods in San Francisco. I added latitude and longitude data from the SearchEngine library so that I could add markers for each neighborhood in a Folium map. I noticed that the SearchEngine library had difficulty with the Outer Richmond and Marina neighborhood locations as the markers were placed in the Pacific Ocean (Figure 1). As a result, I manually updated these locations by modifying the neighborhoods' coordinates based on data from simple Google searches (Figure 2) and remapped the neighborhood markers (Figure 3).

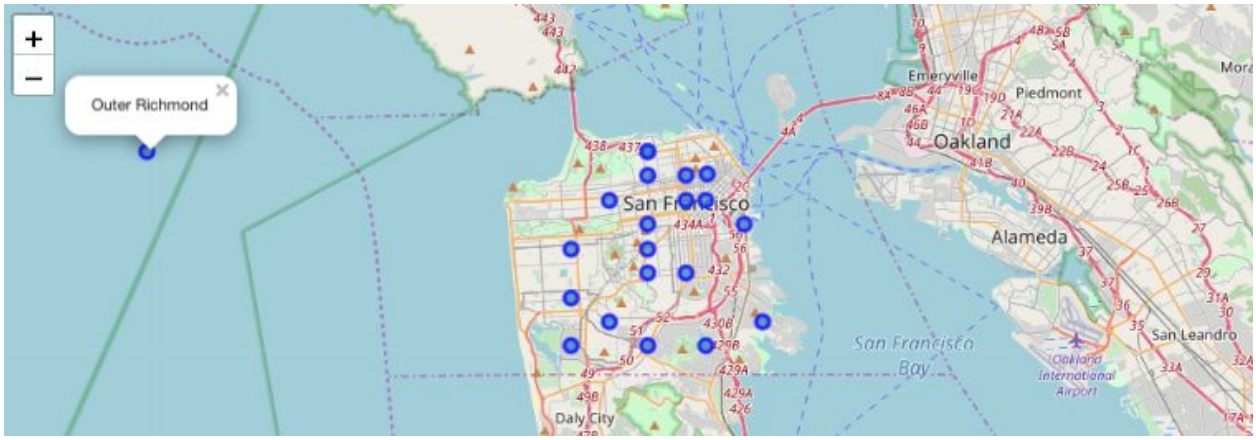

Figure 1. Initial map of San Francisco with misplaced Outer Richmond neighborhood marker.

	Zip Code	Neighborhood	Population (Census 2000)	Latitude	Longitude
1	94102	Hayes Valley/Tenderloin/North of Market	28991	37.780	-122.420
2	94103	South of Market	23016	37.780	-122.410
3	94107	Potrero Hill	17368	37.770	-122.390
4	94108	Chinatown	13716	37.791	-122.409
5	94109	Polk/Russian Hill (Nob Hill)	56322	37.790	-122.420
6	94110	Inner Mission/Bernal Heights	74633	37.750	-122.420
7	94112	Ingelside-Excelsior/Crocker-Amazon	73104	37.720	-122.440
8	94114	Castro/Noe Valley	30574	37.760	-122.440
9	94115	Western Addition/Japantown	33115	37.790	-122.440
10	94116	Parkside/Forest Hill	42958	37.740	-122.480
11	94117	Haight-Ashbury	38738	37.770	-122.440
12	94118	Inner Richmond	38939	37.780	-122.460
13	94121	Outer Richmond	42473	37.781	-122.498
14	94122	Sunset	55492	37.760	-122.480
15	94123	Marina	22903	37.802	-122.438
16	94124	Bayview-Hunters Point	33170	37.730	-122.380
17	94127	St. Francis Wood/Miraloma/West Portal	20624	37.730	-122.460
18	94131	Twin Peaks-Glen Park	27897	37.750	-122.440
19	94132	Lake Merced	26291	37.720	-122.480
20	94133	North Beach/Chinatown	26827	37.800	-122.440
21	94134	Visitacion Valley/Sunnydale	40134	37.720	-122.410

Figure 2: highlighted rows had latitude and longitude values manually updated.

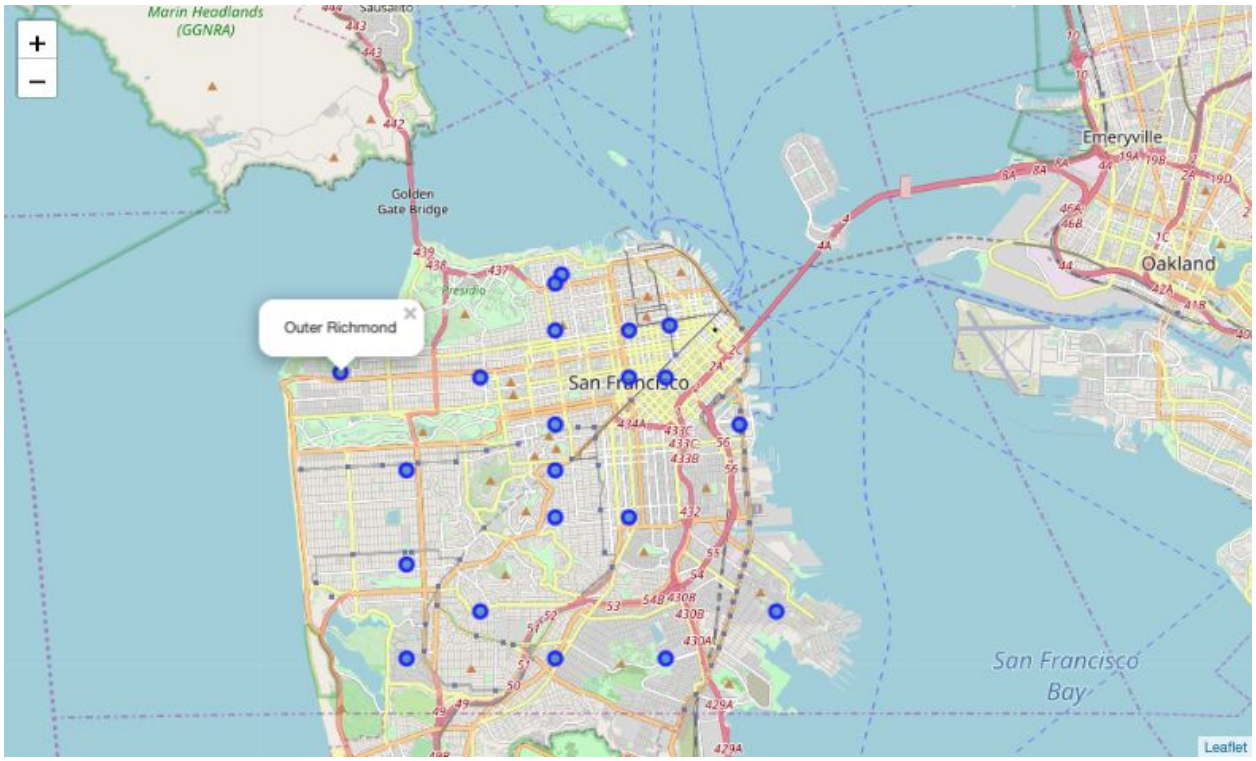

Figure 3: Map of San Francisco with correctly placed neighborhood markers.

Next, I used the Foursquare Venue Explore API endpoint to pull venues for each neighborhood within a 1 mile radius. I pulled the top 200 venues for each neighborhood. This included venues that are categorized as Wine Shops, Art Galleries, Construction & Landscaping, etc. (Figure 4). I realized that I would need to filter these results to the relevant venue categories in order to provide an accurate recommendation.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Hayes Valley/Tenderloin/North of Market	37.78	-122.42	Herbst Theater	37.779548	-122.420953	Concert Hall
1	Hayes Valley/Tenderloin/North of Market	37.78	-122.42	War Memorial Opera House	37.778601	-122.420816	Opera House
2	Hayes Valley/Tenderloin/North of Market	37.78	-122.42	San Francisco Ballet	37.778580	-122.420798	Dance Studio
3	Hayes Valley/Tenderloin/North of Market	37.78	-122.42	Louise M. Davies Symphony Hall	37.777976	-122.420157	Concert Hall
4	Hayes Valley/Tenderloin/North of Market	37.78	-122.42	Asian Art Museum	37.780178	-122.416505	Art Museum

Figure 4: All venue category types included in initial Foursquare search

2.3. Feature Selection

The Foursquare Venue API luckily includes a search parameter for venue category. Even more helpful, Foursquare's venue categories are hierarchical. While some venue categories are very specific (e.g. Tapiocaria within Brazilian Restaurant within South American Restaurant within Latin American Restaurant within Food), the nested nature of the categories allowed me to pull all Asian Restaurants (60+ nested categories under "Asian Restaurant" category). Additionally, "Bubble Tea Shop" is a separate venue category that is *not* nested under Asian Restaurant and therefore does not conflate the analysis. I passed these two venue category IDs into the explore API to pull the top 200 venues within a 1 mile radius for each of the 21 San Francisco neighborhoods.

An additional note, I initially tried using Foursquare's Venue Trending API endpoint instead of the Venue Explore API. The Trending endpoint seemed like it would provide a better representation of the "best" venues for Asian and Bubble Tea locations. Alas, the Trending feature is based on the number of Foursquare user check-ins at a given venue. Since I performed this data pull during the COVID-19 pandemic, many restaurant venues are currently closed under "shelter in place" orders and therefore no venue results were returned from the Trending API (Figure 5).

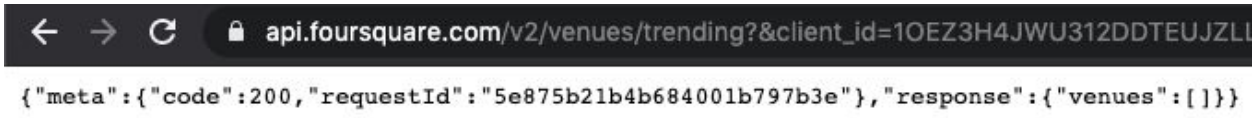

Figure 5: Empty venues search result on Trending API endpoint.

3. Exploratory Data Analysis / Methodology

3.1. Comparing All Venues to Asian Restaurant and Bubble Tea Shop Venues

I performed two searches using the Foursquare API pulling the top 200 venues within 1 mile radius for each neighborhood. The first search was unfiltered (no venue category parameter) and contained 13,993 total venues. The second search was filtered on Asian Restaurant and Bubble Tea Shop venue categories and contained 12,320 total venues (Figure 6). This shows that SF neighborhoods have many Asian and Bubble Tea locations and that the analysis has a healthy sample size for neighborhood comparison.

Compare size of SF venues and SF asian venues

```
sf_venues.size
```

```
13993
```

```
sf_venues_asian.size
```

```
12320
```

Figure 6: number of venues

3.2. Comparing the number of Asian & Bubble Tea Venues by Neighborhood

Next, we want to understand which neighborhoods have the most Asian and Bubble Tea venues to give us an idea of which neighborhoods might be best for our new boba shop location. We can see that Nob Hill has the most venues, followed by Chinatown and South of Market neighborhoods (Figure 7).

Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Polk/Russian Hill (Nob Hill)	144	144	144	144	144	144
Chinatown	139	139	139	139	139	139
South of Market	131	131	131	131	131	131
Hayes Valley/Tenderloin/North of Market	122	122	122	122	122	122
Sunset	119	119	119	119	119	119
Parkside/Forest Hill	116	116	116	116	116	116
Inner Richmond	109	109	109	109	109	109
Western Addition/Japantown	106	106	106	106	106	106
Inner Mission/Bernal Heights	98	98	98	98	98	98
Haight-Ashbury	93	93	93	93	93	93
Castro/Noe Valley	73	73	73	73	73	73
Outer Richmond	72	72	72	72	72	72
St. Francis Wood/Miraloma/West Portal	71	71	71	71	71	71
Marina	65	65	65	65	65	65
North Beach/Chinatown	64	64	64	64	64	64
Potrero Hill	61	61	61	61	61	61
Ingelside-Excelsior/Crocker-Amazon	56	56	56	56	56	56
Lake Merced	42	42	42	42	42	42
Twin Peaks-Glen Park	34	34	34	34	34	34
Visitacion Valley/Sunnydale	34	34	34	34	34	34
Bayview-Hunters Point	11	11	11	11	11	11

Figure 7: Asian and Bubble Tea venue counts by neighborhood.

4. Clustering / Results

4.1. One Hot Encoding for Most Common Asian Venues

Digging deeper into the neighborhood analysis, I started with one hot encoding to get the top 5 most common Asian venues within each neighborhood. This will be the foundation of my analysis to group neighborhoods based on similar density of common Asian restaurant venues. The Foursquare venue category data has unique category values for each type of Asian restaurant (e.g. Chinese Restaurant, Thai Restaurant, Sushi Restaurant, etc.) which helps provide more specificity to the types of Asian restaurants in a given neighborhood. As a result, each neighborhood has the top 5 most common venues (Figure 8).

```
----Bayview-Hunters Point----
                  venue  freq
0     Chinese Restaurant  0.36
1  Vietnamese Restaurant  0.18
2       Asian Restaurant  0.09
3    Japanese Restaurant  0.09
4    Hawaiian Restaurant  0.09


----Castro/Noe Valley----
                  venue  freq
0       Thai Restaurant  0.26
1      Sushi Restaurant  0.22
2  Vietnamese Restaurant  0.08
3       Asian Restaurant  0.08
4     Chinese Restaurant  0.05
```

Figure 8: Each neighborhood has the top 5 most common venues

4.2. K-means Cluster Analysis

After trial and error, I concluded that 4 clusters was the ideal number to segment the original 21 San Francisco neighborhoods. Having too few clusters meant that some neighborhoods were forced a cluster that didn't necessarily fit. On the other hand, having too many clusters added too much specificity and it was difficult to determine the distinguishing characteristics between clusters. At 4 clusters, each cluster is easily identifiable.

Each of the 21 neighborhoods were assigned a cluster label and color. These markers were then added to the San Francisco map to visually show which neighborhoods are related based on common Asian venues (Figure 9).

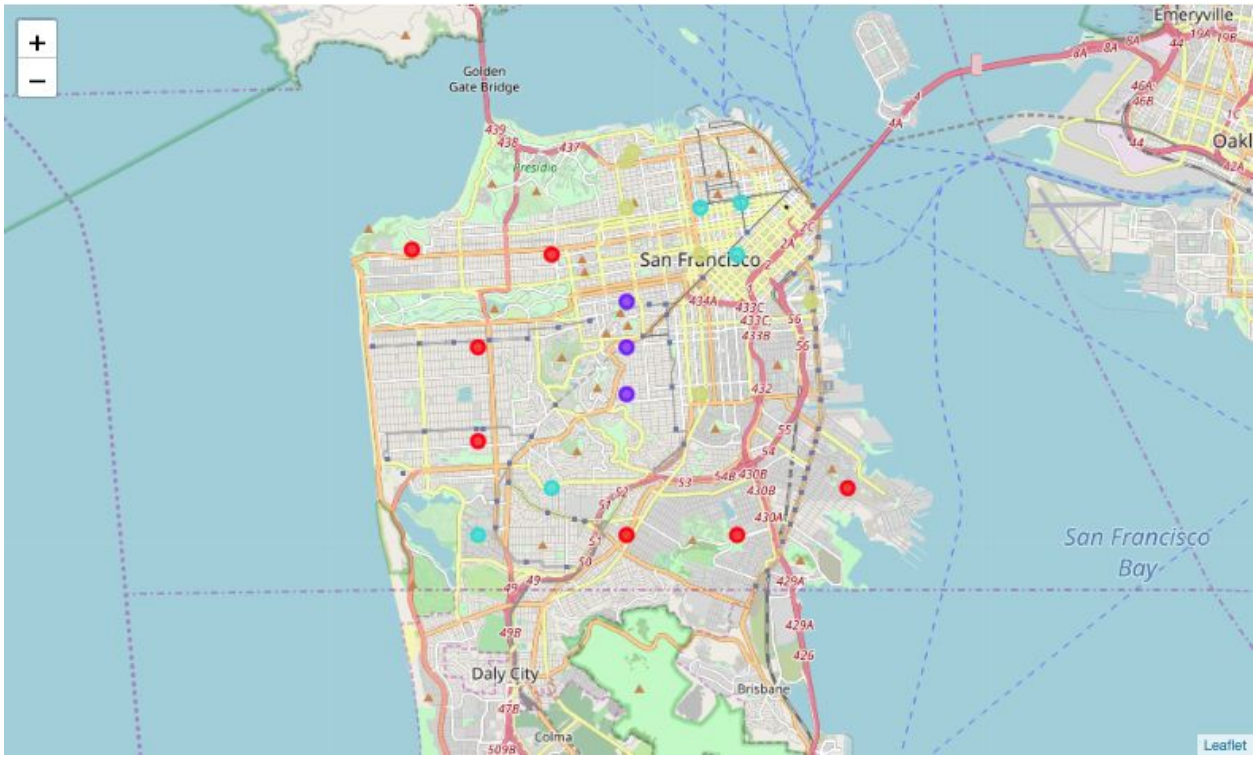

Figure 9: 4 clusters of color coded neighborhoods

5. Conclusions

Outer Richmond and Bayview-Hunters Point neighborhoods are found to be the best neighborhoods to open a boba shop based on this analysis. These neighborhoods have a high density of Chinese Restaurants (boba is specifically a Taiwanese dessert) and low density of Bubble Tea Shops.

Both of these neighborhoods are part of Cluster #1 (colored in Red in Figure 9). Cluster #1 neighborhoods have a high density of Chinese Restaurants (Figure 10). However, Outer Richmond and Bayview-Hunters Point neighborhoods are ideal because neither have a high density of Bubble Tea Shop venues. All of the other neighborhoods in Cluster #1 have Bubble Tea Shops in the top 5 most common venues.

Neighborhood	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
Ingelside-Excelsior/Crocker-Amazon	0	Chinese Restaurant	Vietnamese Restaurant	Bubble Tea Shop	Filipino Restaurant	Asian Restaurant
Parkside/Forest Hill	0	Chinese Restaurant	Bubble Tea Shop	Sushi Restaurant	Asian Restaurant	Japanese Restaurant
Inner Richmond	0	Chinese Restaurant	Asian Restaurant	Sushi Restaurant	Thai Restaurant	Bubble Tea Shop
Outer Richmond	0	Chinese Restaurant	Sushi Restaurant	Vietnamese Restaurant	Japanese Restaurant	Asian Restaurant
Sunset	0	Chinese Restaurant	Bubble Tea Shop	Japanese Restaurant	Asian Restaurant	Vietnamese Restaurant
Bayview-Hunters Point	0	Chinese Restaurant	Vietnamese Restaurant	Asian Restaurant	Sushi Restaurant	Japanese Restaurant
Visitacion Valley/Sunnydale	0	Vietnamese Restaurant	Chinese Restaurant	Asian Restaurant	Bubble Tea Shop	Dim Sum Restaurant

Figure 10: Cluster #1 neighborhoods

Cluster #2 are neighborhoods with a high density of Thai and Sushi Restaurants. Cluster #3 are neighborhoods with high density of bubble tea shops. Cluster #4 are neighborhoods with

high density of Japanese and Sushi Restaurants. Cluster #3 neighborhoods should be eliminated from our search since we want to minimize competition for our own boba shop. Cluster #2 and Cluster #4 neighborhoods should be considered, but with less priority than Cluster #1 neighborhoods because we don't know if non-Chinese Asian food translates to higher boba tea sales.

6. Future Directions & Further Discussion

This analysis is solely based on Asian restaurant density, but does not provide any detail about population demographics, neighborhood crime, or commercial rent prices. When opening a boba shop, the population demographics are important to know the type of customer that would demand bubble tea. Do Outer Richmond and Bayview-Hunters Point neighborhoods have high density of Asian populations that might be more open to buying boba? Similarly, a boba shop owner would want to know the neighborhood crime rates to avoid any issues, as well as commercial rent prices to know whether the business will be sustainable in the long run.