New Chinese Restaurant in Bay Area – The Location Problem

Yang Yang

April, 2019

Introduction

I live in the bay area of California, United States. As the big technology companies in this area keep hiring new engineers and other professionals, more people from all over the country and abroad are moving into this area, a lot of them are Chinese or Asian. Intrigued by this trend, one of my friends is considering opening a new Chinese restaurant in this area.

One of the most important questions he needs to answer first is “Where should I open the new restaurant?” The location choice can be impacted by various factors such as population composition, median family income, competitors, neighborhood future business planning, etc.

As a first step, he wants me to help him find out how many Chinese restaurants already opened in each city of bay area. This is an important data for picking up the final location of the new restaurant as cities with more Chinese restaurants usually mean there are more potential customers for the new restaurant. At the same time, it also means it's more competitive in that area.

Data Acquisition and Preparation

To find out how many Chinese restaurants already opened up in each city of bay area, we first need the city names in the bay area. In the following website: https://data.sfgov.org/Geographic-Locations-and-Boundaries/Bay-Area-ZIP-Codes/u5j3-svi6, we can download the data of each zip code in the bay area along with its city name in either CSV or Json file format.

Out[2]:

	PO_NAME	the_geom	ZIP	STATE	Area__	Length__
0	NAPA	MULTIPOLYGON (((-122.10329200180091 38.5132829...	94558	CA	1.231326e+10	995176.225313
1	FAIRFIELD	MULTIPOLYGON (((-121.947475002335 38.301511000...	94533	CA	9.917861e+08	200772.556587
2	DIXON	MULTIPOLYGON (((-121.65335500334429 38.3133870...	95620	CA	7.236950e+09	441860.201400
3	SONOMA	MULTIPOLYGON (((-122.406843003057 38.155681999...	95476	CA	3.001414e+09	311318.546326
4	NAPA	MULTIPOLYGON (((-122.29368500225117 38.1552379...	94559	CA	1.194302e+09	359104.646602
5	PETALUMA	MULTIPOLYGON (((-122.45766900253919 38.1168949...	94954	CA	2.006544e+09	267474.490552
6	RIO VISTA	MULTIPOLYGON (((-121.8624620022998 38.06602999...	94571	CA	4.454446e+09	492056.752411
7	TRAVIS AFB	MULTIPOLYGON (((-121.89653900297888 38.2865679...	94535	CA	3.029397e+08	95232.008421
8	AMERICAN CANYON	MULTIPOLYGON (((-122.20418700285576 38.2096949...	94503	CA	6.931341e+08	136394.695137
9	NOVATO	MULTIPOLYGON (((-122.48655900081091 38.1005269...	94949	CA	4.316054e+08	119395.672078
10	NOVATO	MULTIPOLYGON (((-122.48655900081091 38.1005269...	94945	CA	7.537170e+08	159439.880452
11	BIRDS LANDING	MULTIPOLYGON (((-121.8861390034904 38.12559299...	94512	CA	2.559026e+08	94434.950401
12	VALLEJO	MULTIPOLYGON (((-122.22853700371564 38.1259699...	94591	CA	4.845067e+08	109660.438701
13	BENICIA	MULTIPOLYGON (((-122.064180002481 38.118817000...	94510	CA	8.567942e+08	180958.749249
14	VALLEJO	MULTIPOLYGON (((-122.27408000249419 38.1109719...	94592	CA	4.885450e+08	136524.530193
15	VALLEJO	MULTIPOLYGON (((-122.22853700371564 38.1259699...	94589	CA	1.713895e+08	68793.178537
16	NOVATO	MULTIPOLYGON (((-122.5407480037163 38.07354500...	94947	CA	5.262859e+08	142727.147200
17	VALLEJO	MULTIPOLYGON (((-122.22853700371564 38.1259699...	94590	CA	1.633788e+08	72484.570662
18	NICASIO	MULTIPOLYGON (((-122.70929400072164 38.0046029...	94946	CA	1.745904e+09	255832.255447
19	OAKLEY	MULTIPOLYGON (((-121.6626900026755 37.96880400...	94561	CA	1.765427e+09	254803.565820

Figure 1. Pandas data frame built from downloaded csv file

From the CSV file we can build a Pandas data frame for further data cleaning and preparation (as shown in Figure 1). Notice some cities have multiple zip codes, we can combine them together (as shown in Figure 2).

Out[4]:

	PO_NAME	ZIP
0	ALAMEDA	94502, 94501
1	ALAMO	94507
2	ALBANY	94706
3	ALVISO	95002
4	AMERICAN CANYON	94503
5	ANTIOCH	94509, 94531
6	ATHERTON	94027
7	BELMONT	94002
8	BELVEDERE TIBURON	94920
9	BENICIA	94510
10	BERKELEY	94708, 94720, 94707, 94710, 94709, 94703, 9470...
11	BIRDS LANDING	94512
12	BRENTWOOD	94513
13	BRISBANE	94005
14	BURLINGAME	94010
15	BYRON	94514
16	CASTRO VALLEY	94546, 94552
17	CLAYTON	94517
18	CONCORD	94520, 94521, 94519, 94518
19	CORTE MADERA	94925
20	CROCKETT	94525
21	DALY CITY	94015, 94014
22	DANVILLE	94526, 94506
23	DIABLO	94528

Figure 2. New data frame after combing multiple zip codes of the same city into one row

After getting the city names, next step we need to get the latitude and longitude value of each city, this can be accomplished by using geocoders in the geopy library.

Out[6]:

	PO_NAME	ZIP	Latitude	Longitude
0	ALAMEDA	94502, 94501	37.609029	-121.899142
1	ALAMO	94507	37.850203	-122.032184
2	ALBANY	94706	37.886870	-122.297747
3	ALVISO	95002	37.426051	-121.975237
4	AMERICAN CANYON	94503	38.174918	-122.260804
5	ANTIOCH	94509, 94531	38.004921	-121.805789
6	ATHERTON	94027	37.461327	-122.197743
7	BELMONT	94002	37.520215	-122.275801
8	BELVEDERE TIBURON	94920	37.878023	-122.469145
9	BENICIA	94510	38.049365	-122.158578
10	BERKELEY	94708, 94720, 94707, 94710, 94709, 94703, 9470...	37.870839	-122.272864
11	BIRDS LANDING	94512	38.132695	-121.870793
12	BRENTWOOD	94513	37.931777	-121.696027
13	BRISBANE	94005	37.680766	-122.399972
14	BURLINGAME	94010	37.584103	-122.366083
15	BYRON	94514	37.874063	-121.634657
16	CASTRO VALLEY	94546, 94552	37.715818	-122.090288
17	CLAYTON	94517	37.941034	-121.935793
18	CONCORD	94520, 94521, 94519, 94518	37.976852	-122.033562
19	CORTE MADERA	94925	37.925481	-122.527475

Figure 3. New data frame with latitude and longitude of each city

Next step, we will use Foursquare API to get the venues information of neighborhoods of each city.

Methodology

We use Foursquare API to find the top 500 venues that are within a radius of 2000 meters to all the cities. The resulting data frame is shown in Figure 4.

Out[11]:

	City	City Latitude	City Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	ALAMEDA	37.609029	-121.899142	Pleasanton Ridge Regional Park	37.614761	-121.881874	Trail
1	ALAMEDA	37.609029	-121.899142	Elliston Vineyards	37.601171	-121.890326	Winery
2	ALAMEDA	37.609029	-121.899142	Bosco's Bones & Brew	37.593791	-121.888016	American Restaurant
3	ALAMEDA	37.609029	-121.899142	Niles Canyon Farwell Trestle	37.594598	-121.905016	Trail
4	ALAMEDA	37.609029	-121.899142	Sunol Station Niles Canyon Railway	37.594190	-121.888893	Train Station
5	ALAMEDA	37.609029	-121.899142	Sunol Food & Liquor	37.594004	-121.887918	Grocery Store
6	ALAMEDA	37.609029	-121.899142	Casa Bella Bistro	37.593703	-121.888451	Bistro
7	ALAMO	37.850203	-122.032184	Alamo Cafe	37.849932	-122.032031	American Restaurant
8	ALAMO	37.850203	-122.032184	Extreme Pizza	37.848096	-122.031570	Pizza Place
9	ALAMO	37.850203	-122.032184	Safeway	37.850845	-122.035224	Grocery Store
10	ALAMO	37.850203	-122.032184	Panera Bread	37.851191	-122.033797	Sandwich Place
11	ALAMO	37.850203	-122.032184	cherubini coffee house	37.850007	-122.031462	Café
12	ALAMO	37.850203	-122.032184	CVS pharmacy	37.851902	-122.032449	Pharmacy
13	ALAMO	37.850203	-122.032184	Alamo Palace	37.852612	-122.035127	Chinese Restaurant
14	ALAMO	37.850203	-122.032184	Rite Aid	37.852412	-122.035486	Pharmacy
15	ALAMO	37.850203	-122.032184	Five Guys	37.851432	-122.035319	Burger Joint
16	ALAMO	37.850203	-122.032184	Peet's Coffee & Tea	37.851114	-122.034124	Coffee Shop
17	ALAMO	37.850203	-122.032184	The Peasant's Courtyard	37.849670	-122.032752	New American Restaurant
18	ALAMO	37.850203	-122.032184	Orangetheory Fitness	37.851279	-122.034997	Gym / Fitness Center
19	ALAMO	37.850203	-122.032184	Round Table Pizza	37.852124	-122.036282	Pizza Place
20	ALAMO	37.850203	-122.032184	Brass Bear Delicatessen	37.851238	-122.035644	Deli / Bodega
21	ALAMO	37.850203	-122.032184	The UPS Store	37.852625	-122.033553	Shipping Store

Figure 4. Top 500 venues in each city

As we can see in the column “Venue Category”, it gives us information what kind of venue it is. We’ll use the filter to get all the venues in each bay area city that have the value “Chinese Restaurant”. Figure 5 shows the filtered data frame.

Out[12]:

	City	City Latitude	City Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
13	ALAMO	37.850203	-122.032184	Alamo Palace	37.852612	-122.035127	Chinese Restaurant
40	ALAMO	37.850203	-122.032184	Yan's China Bistro	37.851542	-122.031984	Chinese Restaurant
64	ALBANY	37.886870	-122.297747	China Village Restaurant	37.890831	-122.291234	Chinese Restaurant
185	AMERICAN CANYON	38.174918	-122.260804	Panda Express	38.180866	-122.254158	Chinese Restaurant
236	ANTIOCH	38.004921	-121.805789	China City Restaurant	37.994895	-121.806587	Chinese Restaurant
262	ANTIOCH	38.004921	-121.805789	Canton City	38.016345	-121.813927	Chinese Restaurant
358	BELMONT	37.520215	-122.275801	Panda Express	37.519659	-122.275432	Chinese Restaurant
372	BELMONT	37.520215	-122.275801	Blue Sky Cafe	37.514525	-122.268431	Chinese Restaurant
408	BELMONT	37.520215	-122.275801	Peking Alley	37.529998	-122.290073	Chinese Restaurant
431	BELMONT	37.520215	-122.275801	King Chuan	37.506726	-122.261280	Chinese Restaurant
468	BELVEDERE TIBURON	37.878023	-122.469145	Lily Kai	37.873820	-122.455203	Chinese Restaurant
524	BENICIA	38.049365	-122.158578	Panda Express	38.066390	-122.164756	Chinese Restaurant
575	BERKELEY	37.870839	-122.272864	Long Life Vegi House	37.871144	-122.276611	Chinese Restaurant
594	BERKELEY	37.870839	-122.272864	Great China	37.867636	-122.266123	Chinese Restaurant
647	BERKELEY	37.870839	-122.272864	King Dong	37.865095	-122.267370	Chinese Restaurant
716	BRENTWOOD	37.931777	-121.696027	Dragon City Restaurant	37.947260	-121.697257	Chinese Restaurant
723	BRENTWOOD	37.931777	-121.696027	Canton Garden	37.942104	-121.694956	Chinese Restaurant
737	BRISBANE	37.680766	-122.399972	Na Na's Kitchen	37.682355	-122.403598	Chinese Restaurant
758	BRISBANE	37.680766	-122.399972	Lucky House	37.683393	-122.403187	Chinese Restaurant
968	CONCORD	37.976852	-122.033562	Shan Shan Low	37.974222	-122.040852	Chinese Restaurant
1208	DALY CITY	37.705767	-122.461921	Chopstix	37.698214	-122.464380	Chinese Restaurant
1212	DALY CITY	37.705767	-122.461921	Egg Roll King	37.699229	-122.463903	Chinese Restaurant

Figure 5. Filtered data frame of “Chinese Restaurant”

In order to visualize Chinese restaurants locations in different cities with different colors in the map, we need to add an integer label representing individual city for every row, we call the new column 'City Label'. Figure 6 shows the data frame after adding the new column.

Out[13]:

	City	City Latitude	City Label	City Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
13	ALAMO	37.850203	0	-122.032184	Alamo Palace	37.852612	-122.035127	Chinese Restaurant
40	ALAMO	37.850203	0	-122.032184	Yan's China Bistro	37.851542	-122.031984	Chinese Restaurant
64	ALBANY	37.886870	1	-122.297747	China Village Restaurant	37.890831	-122.291234	Chinese Restaurant
185	AMERICAN CANYON	38.174918	2	-122.260804	Panda Express	38.180866	-122.254158	Chinese Restaurant
236	ANTIOCH	38.004921	3	-121.805789	China City Restaurant	37.994895	-121.806587	Chinese Restaurant
262	ANTIOCH	38.004921	3	-121.805789	Canton City	38.016345	-121.813927	Chinese Restaurant
358	BELMONT	37.520215	4	-122.275801	Panda Express	37.519659	-122.275432	Chinese Restaurant

Figure 6. Data frame adding new column “City Label”

Results

After sorting the data frame, we can get the descending rank of the cities with Chinese restaurants, as shown in Figure 7. Top 5 cities are Milpitas, Millbrae, Pinole, Fremont and Daly City.

Out[15]:

City	City Latitude	City Label	City Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
MILPITAS	6	6	6	6	6	6	6
MILLBRAE	5	5	5	5	5	5	5
PINOLE	5	5	5	5	5	5	5
FREMONT	4	4	4	4	4	4	4
DALY CITY	4	4	4	4	4	4	4
BELMONT	4	4	4	4	4	4	4
PITTSBURG	4	4	4	4	4	4	4
HAYWARD	4	4	4	4	4	4	4
FAIRFIELD	3	3	3	3	3	3	3
EL SOBRANTE	3	3	3	3	3	3	3
PLEASANT HILL	3	3	3	3	3	3	3
MOUNTAIN VIEW	3	3	3	3	3	3	3
SAN CARLOS	3	3	3	3	3	3	3
OAKLAND	3	3	3	3	3	3	3
BERKELEY	3	3	3	3	3	3	3
SAN RAMON	3	3	3	3	3	3	3
SUNNYVALE	3	3	3	3	3	3	3
PALO ALTO	2	2	2	2	2	2	2
REDWOOD CITY	2	2	2	2	2	2	2
PLEASANTON	2	2	2	2	2	2	2
NEWARK	2	2	2	2	2	2	2
RICHMOND	2	2	2	2	2	2	2

Figure 7. Sorted data frame with descending order

Discussion

To better analyze the result, let's visualize it. Using folium library and the data frame, we can generate the map as shown in Figure 8.

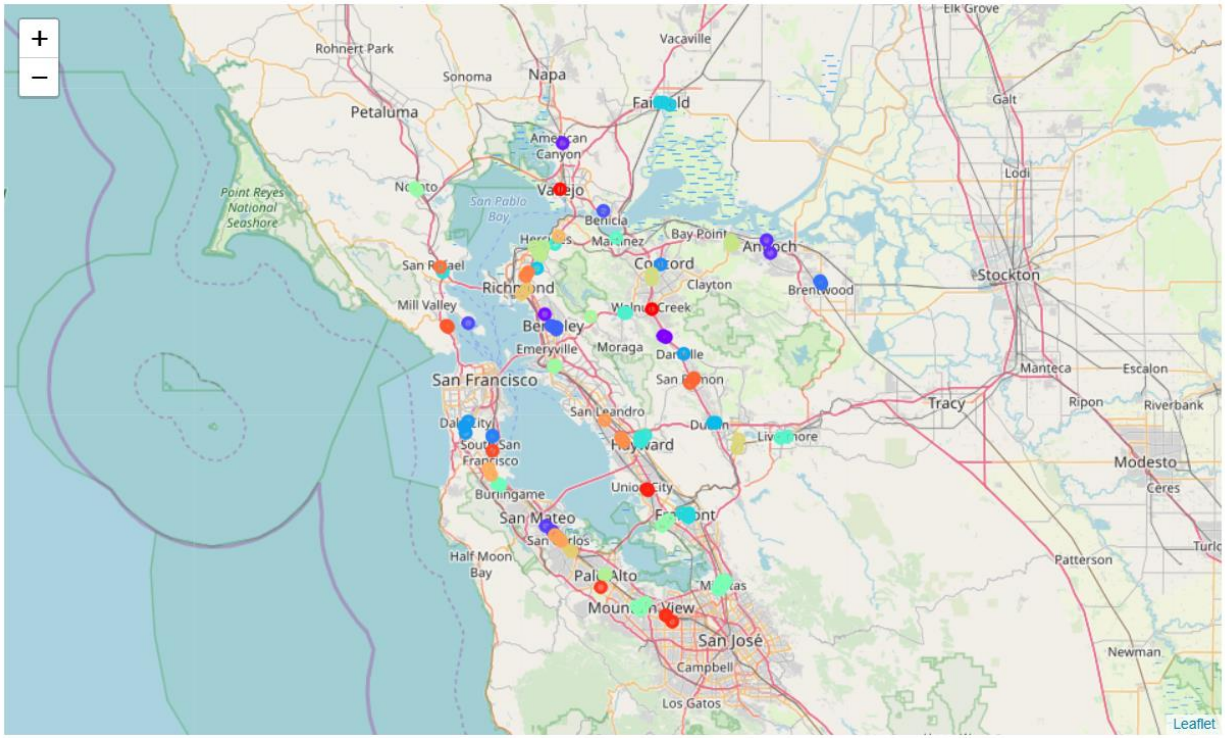

Figure 8. Chinese restaurants locations in bay area cities

As we can see from the map, most restaurants locations are along the San Francisco Bay. In fact, 6 out of the top 8 cities are located alongside the Bay (Milpitas, Millbrae, Fremont, Daly City, Belmont, Hayward). This result agrees with my impression as the cities alongside the San Francisco Bay are the most populated cities in the whole area and most technology companies are located here which brings a lot of business to the restaurants.

Conclusion

Through this work, we've answered the question of how many Chinese restaurants already opened in each city of bay area and find the cities alongside the San Francisco Bay are the area where most Chinese restaurants are located. However, this is only the first step of the whole process to pick the location of my friend's proposed new restaurant. Aforementioned other factors such as competitors, neighborhood future business planning are also very important ones to consider. I will work with my friend to collect available data regarding these

factors and use data science methodology and tools to come up with the final decision. Stay tuned.