Capstone Project
The Battle of Neighborhoods Report

Yu Han
07/08/2020

Introduction

Background & Problem Description

XYZ company is a famous Asian food high end restaurant chain. Customers normally spend $100 - $200 per person when they are dining at XYZ restaurants. This year, XYZ decided to expand their business to North America. They are trying to launch their first restaurant in San Francisco.San Francisco is located at the end of the peninsula, with the Pacific Ocean to the west, the San Francisco Bay to the east, and the Golden Gate Bridge to the north. Today, the city is regarded as one of the most popular international tourist destinations, and has many famous landmarks such as the Golden Gate Bridge, cable car, Alcatraz Island, Chinatown and Coit Tower.Moreover, according to the data from census department, the median household income of San Francisco is $104,552.00 which is almost double of the national median household income. However, as an international enterprise who just entered the San Francisco Market. XYZ's management team has the trouble to decide where they should put their first restaurant. They want to find a big venue in the city and the neighbor has to be safe. Moreover, They do not want that many asian restaurants nearby. In this report, we gonna use some public data to analyze the neighborhoods in San Francisco and help XYZ company make the final deci
sion.

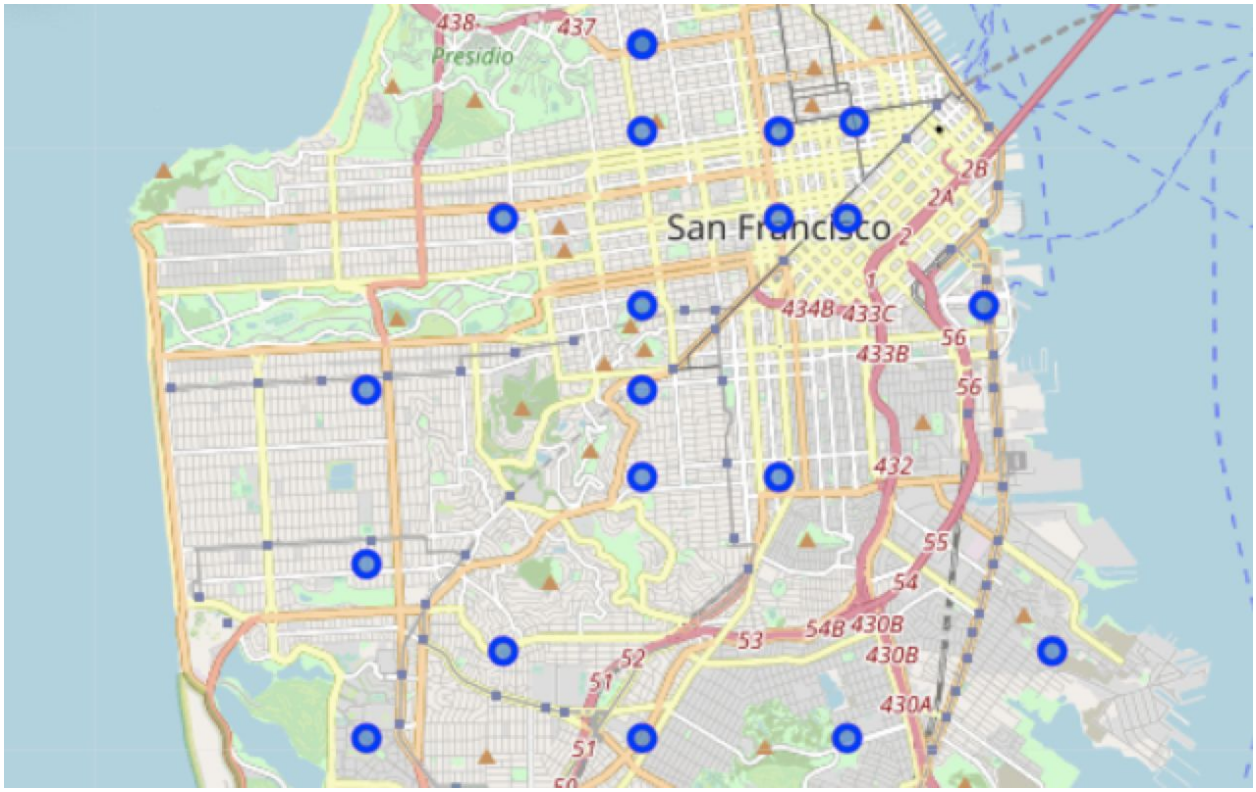
Methodology

I use BeautifulSoup to scrape boroughs from sf health org and organize a table containing Population,Zipcode, Neighborhoods information of San Francisco. Then I

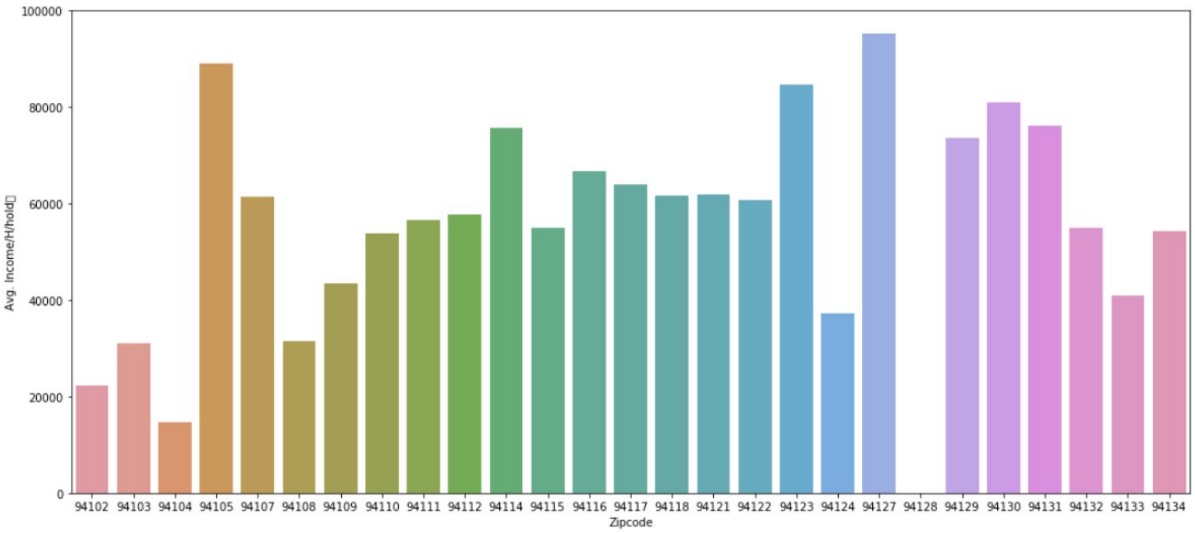
installed the uszipcode tool and use it to add the latitude and longitude to the San Francisco DataFrame.

	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

After the data frame looks good at first glance. I convert it to a map to do a brief overview of San Francisco Neighborhoods.



By utilize the household income data by zip code. I found out that the top 3 highest average household income are 94127（West Portal)， 94105(Mission)， 94123(Marina).


100000
80000
60000
40000
20000
0
Avg. Income/H/hold
94102 94103 94104 94105 94107 94108 94109 94110 94111 94112 94114 94115 94116 94117 94118 94121 94122 94123 94124 94127 94128 94129 94130 94131 94132 94133 94134
Zipcode

	Unnamed: 0	Zipcode	Latitude	Longitude	City	Population	Avg. Income/H/hold\t
0	1	94127	37.737	-122.457	San Francisco, California	20,624	95313.0
1	2	94105	37.789	-122.395	San Francisco, California	2,058	88976.0
2	3	94123	37.800	-122.437	San Francisco, California	22,903	84710.0
3	4	94130	37.821	-122.370	San Francisco, California	1,453	80959.0
4	5	94131	37.747	-122.443	San Francisco, California	27,897	76044.0
5	6	94114	37.758	-122.435	San Francisco, California	30,574	75727.0
6	7	94129	37.798	-122.465	San Francisco, California	2,228	73571.0
7	8	94116	37.744	-122.487	San Francisco, California	42,958	66627.0
8	9	94117	37.771	-122.445	San Francisco, California	38,738	63983.0
9	10	94121	37.777	-122.496	San Francisco, California	42,473	61776.0
10	11	94118	37.781	-122.462	San Francisco, California	38,939	61609.0
11	12	94107	37.769	-122.396	San Francisco, California	17,368	61362.0
12	13	94122	37.760	-122.485	San Francisco, California	55,492	60733.0
13	14	94112	37.720	-122.443	San Francisco, California	73,104	57629.0
14	15	94111	37.802	-122.402	San Francisco, California	3,335	56569.0
15	16	94132	37.722	-122.491	San Francisco, California	26,291	55000.0
16	17	94115	37.786	-122.437	San Francisco, California	33,115	54879.0
17	18	94134	37.721	-122.414	San Francisco, California	40,134	54342.0
18	19	94110	37.750	-122.415	San Francisco, California	74,633	53795.0
19	20	94109	37.794	-122.422	San Francisco, California	56,322	43444.0
20	21	94133	37.802	-122.411	San Francisco, California	26,827	40990.0
21	22	94124	37.732	-122.385	San Francisco, California	33,170	37146.0
22	23	94108	37.792	-122.409	San Francisco, California	13,716	31542.0

Moreover, We need to take a look at neighborhood safety. The data of the police department incident report could help us to better understand the crime situation of the neighborhoods in San Francisco. After analyzing the data of Police Department Incident Reports 2018 to Present. We found out that neighborhoods have a relatively low number of police reports.

Out[15]:

McLaren Park	354
Seacliff	395
Lincoln Park	424
Presidio	790
Treasure Island	1125
Glen Park	1762
Twin Peaks	1783
Presidio Heights	2075
Noe Valley	3311
Oceanview/Merced/Ingleside	3524
Japantown	3579
Golden Gate Park	3658
Portola	3841
Visitacion Valley	3895
Inner Sunset	4097
Lakeshore	4113
Lone Mountain/USF	4208
Inner Richmond	4565
Outer Mission	4776
Excelsior	5599
Potrero Hill	5608
Haight Ashbury	5793
Mission Bay	5881
Pacific Heights	5888
Bernal Heights	6064
Chinatown	6822
West of Twin Peaks	6877
Outer Richmond	7601
Russian Hill	7830
Marina	8405
Hayes Valley	9496
Nob Hill	10032
Sunset/Parkside	10059
Castro/Upper Market	10498
North Beach	10527
Western Addition	11396
Bayview Hunters Point	20358
South of Market	29674

Then I managed to use the FourSquare API to analyze each neighborhood. Surprisingly. There are 243 unique categories. And we figured the top 5 most common venues for each neighbor. Let's check the neighbors with the higher income.

Higher income：

West Portal -

```
----St. Francis Wood/Miraloma/West Portal----
                 venue  freq
0  Chinese Restaurant  0.12
1  Light Rail Station  0.09
2               Café  0.06
3        Pizza Place  0.06
4        Yoga Studio  0.06
```

Mission -

----Inner Mission/Bernal Heights----

	venue	freq
0	Mexican Restaurant	0.09
1	Coffee Shop	0.06
2	Pizza Place	0.04
3	Grocery Store	0.04
4	Italian Restaurant	0.04

Marina -

----Marina----

	venue	freq
0	French Restaurant	0.05
1	Gym / Fitness Center	0.05
2	Cosmetics Shop	0.05
3	Italian Restaurant	0.05
4	Wine Bar	0.04

The common venue of Marina：

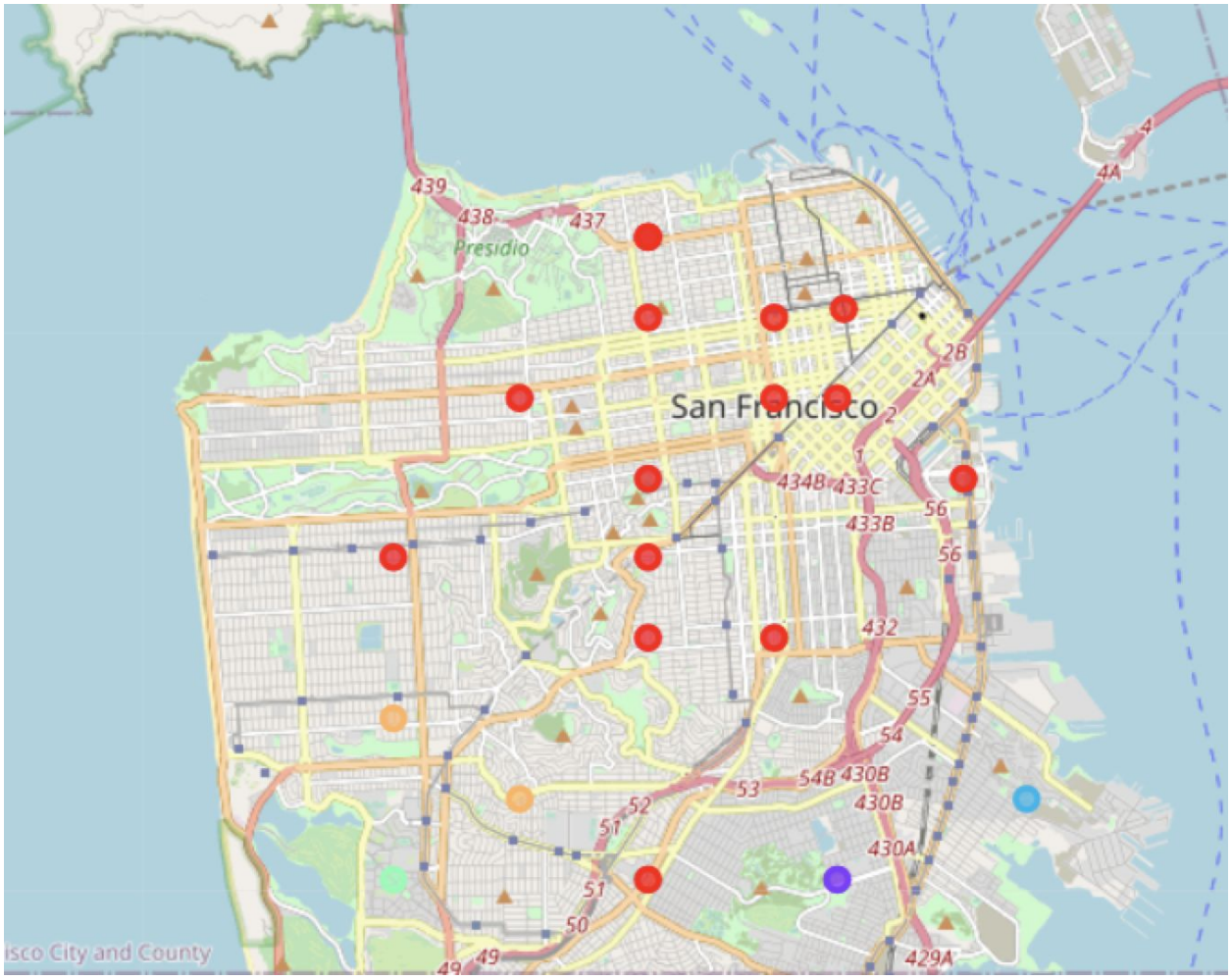
9	Marina	Gym / Fitness Center	Cosmetics Shop	French Restaurant	Italian Restaurant	Wine Bar	Playground	Juice Bar	Sandwich Place	Salad Place	Coffee Shop

From the result above, we can tell that the marina district has low frequency of every category. The most important. There is no Asian Restaurant here. Let's take a look at the crime rate of marina district.

```
Marina                                    8405
```

Compared to other districts, the crime reports of Marina district is not that low but still lower than the other dangerous areas.

Now lets run K-means to cluster the neighborhood into 5 clusters.


San Francisco
Presidio
439
438
437
4
4A
2B
2A
2
1
434B
433C
433B
56
56
432
55
54
54B
430B
430B
430A
53
52
51
51
50
49
429A
isco City and County

Conclusion

We can see that most dots are in red. The neighborhood marina is also part of it. By combining all the analysis we made before. XYZ should set up their first restaurant in Marina district to fulfill all the requirements. People who live around the marina area have relatively higher income which can afford the restaurant. Moreover, there is not that much asian cuisine there so the competitors are less. Last but not least, the Marina area is not that dangerous like other area in San Francisco.

Reference

zipcode:http://www.healthysf.org/bdi/outcomes/zipmap.htm

SF crime:https://sfgov.org/services/sf-crime-statistics

SF House income:

http://zipatlas.com/us/ca/san-francisco/zip-code-comparison/median-household-income.htm