

SF Bay Area: Battle of Neighborhoods

Hui Zeng

July 1, 2020

1. Introduction

San Francisco Bay Area is a populous region in Northern California with nearly 7.8 million people within a nine-county region. It is a major job hub for high tech workers, the population has grown by over 600k since 2010 according to a report by KQED news [1].

There are lots of challenges for thousands of newcomers when they first arrive, and one of the most important and frequent questions is where should I live? Within a commute of 1-2 hours to either major job centers like San Francisco or Silicon Valley, there are lots of cities one can choose from.

In this project, I want to use publicly available social and geographic data to help answer this question utilizing the Data Science tools I have learned during this course.

2. Data source

2.1 Target City Selection

There are around ~100 cities in Bay Area, so to limit the scope for this project, I filter out 25 most populated cities as candidate cities based on area size and population per square mile. First, I get a list of cities in Bay Area from the wiki page [2] which include both the area in square miles and population. I filter out 50% of smaller cities in terms of area and sort the remaining cities by population per square miles. I pick the top 25 cities from this list as target cities in this study.

Then, I will select a list of somewhat arbitrary criteria for classifying each city: schools ranking, housing costs, neighborhoods, crime rates. One can argue there might be better indicators such as commute time should also be included, but above criteria are chosen mainly because of its availability, also my personal preference or interest.

Following discuss the data source for each of them.

2.2 School Ranking

It is difficult to scrape online data directly with limited methods I know of, and to not invest too much time on this part of data collection, I looked up all 25 cities' school rankings by browsing through the website www.niche.com [5]. This data is the only one collected manually by average the top 3 highest ranked high school ratings of each city according to the website. I saved the ranking into a csv file and uploaded it to the jupyter notebook for further analysis.

2.3 Housing Cost

For housing cost, I found the average home price for each city in 2018 from Vitalsigns's website [3]. The data is in csv format, it is loaded into a jupyter notebook and extracted for median home as housing cost for this project.

2.4 Neighborhood

I use the venues data from Foursquare to get information about the neighborhood around each city [6]. Most cities have similar venues like coffee shops, restaurants as most popular businesses. There are hundreds of different venues reported, to avoid diluting other criteria such as housing and school, I classified the neighborhood into 5 different labels and only use this summary label in the overall classification at the end.

2.5 Crime rate

The crime rate data is coming from Wikipedia's California locations by crime rate page [4]. It might be the most important indicator for lots of people with options on where to live.

3. Methodology

The project is focusing on classification of cities around the Bay Area for the purpose of providing extra information in helping decide where to live. There are four metrics used as features as discussed above.

For the classification, I use the K-nearest neighbors (KNN) algorithm to group cities with similar metrics together and display them on a map. A final table with all the data and classification is included in the report.

4. Data Analysis and Results

From the list of target cities, San Jose as biggest in terms of area, 176 Square miles, and San Francisco as most densely populated, 17k people per square mile. It also includes some notable cities in Silicon Valley, such as Mountain View and Santa Clara.

	city	area	population per sq mi
0	San Francisco	46.87	17180.1792
1	Berkeley	10.47	10752.6266
2	San Mateo	12.13	8013.76752
3	Oakland	55.79	7003.47733
4	South San Francisco	9.14	6961.9256
5	Alameda	10.61	6956.83318
6	Sunnyvale	21.99	6370.21373
7	San Leandro	13.34	6368.06597
8	Santa Clara	18.41	6326.34438
9	Mountain View	12	6172.16667
10	San Jose	176.53	5358.53396
11	Cupertino	11.26	5177.79751
12	Milpitas	13.59	4914.64312
13	Napa	17.84	4311.37892
14	Santa Rosa	41.5	4279.18072
15	Petaluma	14.38	4029.27677
16	Concord	30.55	3995.64648
17	San Ramon	18.06	3994.90587
18	Redwood City	19.42	3955.45829
19	Vallejo	30.67	3780.30649
20	Pittsburg	17.22	3673.8676
21	Antioch	28.35	3611.00529
22	Union City	19.47	3570.41603
23	San Rafael	16.47	3504.12872
24	Brentwood	14.79	3480.79784

4.1 Housing

From the home price data, I find the costliest areas are close to Silicon Valley or SF, like Cupertino which is the home base of Apple. The price drops with distance increase from these centers.

	Name	Median Price
--	------	--------------

0	Cupertino	2289033
1	Mountain View	1875867
2	Sunnyvale	1857625
3	Redwood City	1636433
4	San Mateo	1437700
5	Santa Clara	1385692
6	San Francisco	1337250
7	Berkeley	1258050
8	Milpitas	1094092
9	San Jose	1067425
10	San Ramon	1040308
11	South San Francisco	1014283
12	San Rafael	1008283
13	Alameda	977150
14	Union City	903742
15	Oakland	739217
16	Petaluma	705008
17	San Leandro	662850
18	Napa	644142
19	Concord	592467
20	Santa Rosa	579275
21	Brentwood	575833
22	Antioch	435133
23	Pittsburg	423358
24	Vallejo	406692

4.2 School Ranking

The school ranking in the following table represents its standing among all high schools in California. Cupertino has the best schools and not surprisingly it has the most expensive housing cost from the above section, around 2.3M.

	Name	Avg ranking of top 3 high schools
0	Cupertino	22
1	Mountain View	26
2	Sunnyvale	30

3	San Mateo	45
4	San Ramon	70
5	San Jose	70
6	Berkeley	80
7	Redwood City	90
8	San Francisco	120
9	Alameda	130
10	Santa Clara	140
11	Oakland	210
12	Santa Rosa	230
13	Antioch	330
14	Concord	330
15	Petaluma	370
16	Pittsburg	400
17	Milpitas	400
18	Union City	430
19	South San Francisco	450
20	Brentwood	500
21	Napa	550
22	San Rafael	600
23	Vallejo	600
24	San Leandro	600

4.3 Neighborhood

I collected the venues information for each city and classified all cities into 5 categories based on the frequency of each venue. Labels 0 and 1 are similar to urban neighborhoods with lots of coffee shops and restaurants. Label 2 has more parks. Label 3 has historical sites and Label 4 has a lake!

Name	Label	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
Oakland	0	Coffee Shop	Brewery	Beer Garden	Bar	Mexican Restaurant
Santa Clara	0	Golf Course	Coffee Shop	Trail	Pizza Place	Park
Napa	0	Hotel	American Restaurant	Grocery Store	Italian Restaurant	Coffee Shop

Concord	0	Donut Shop	American Restaurant	Pizza Place	Sandwich Place	Convenience Store
Pittsburg	0	Ice Cream Shop	American Restaurant	Taco Place	Bar	Park
South San Francisco	1	Mexican Restaurant	Park	Filipino Restaurant	Grocery Store	Sandwich Place
San Leandro	1	Burger Joint	Mexican Restaurant	Park	Ice Cream Shop	Vietnamese Restaurant
San Jose	1	Mexican Restaurant	Sandwich Place	Grocery Store	Pizza Place	Bar
Milpitas	1	Pizza Place	Mexican Restaurant	Bakery	Grocery Store	Trail
Petaluma	1	Coffee Shop	Pizza Place	Park	Grocery Store	Ice Cream Shop
Vallejo	1	Coffee Shop	Mexican Restaurant	Breakfast Spot	Grocery Store	Theme Park Ride / Attraction
Antioch	1	Coffee Shop	Fast Food Restaurant	Mexican Restaurant	Pizza Place	Ice Cream Shop
Union City	1	Mexican Restaurant	Park	Chinese Restaurant	Bakery	Coffee Shop
San Rafael	1	Mexican Restaurant	Grocery Store	Pizza Place	Sandwich Place	American Restaurant
Brentwood	1	Coffee Shop	Grocery Store	Pharmacy	Pizza Place	Mexican Restaurant
San Francisco	2	Coffee Shop	Park	Bakery	Yoga Studio	Grocery Store
Berkeley	2	Pizza Place	Park	Coffee Shop	Grocery Store	Brewery
San Mateo	2	Park	Grocery Store	Japanese Restaurant	Sandwich Place	Trail
Alameda	2	Coffee Shop	Park	Grocery Store	Trail	Ice Cream Shop
Sunnyvale	2	Park	Grocery Store	Sandwich Place	Pizza Place	Coffee Shop
Mountain View	2	Park	Grocery Store	Sandwich Place	Indian Restaurant	Farmers Market
Cupertino	2	Park	Grocery Store	Sandwich Place	Coffee Shop	Bakery
Redwood City	2	Sandwich Place	Grocery Store	Park	Coffee Shop	Gym

San Ramon	3	Restaurant	Historic Site	Mediterranean Restaurant	Food	Italian Restaurant
Santa Rosa	4	Lake	Mountain	Restaurant	Hotel Pool	Outdoors & Recreation

4.4 Crime Rate

The data from Wikipedia is the number of Violent crimes per 1000 people in 2014, and the average in California is 3.96. Again, Cupertino is one of the safest cities in the list.

	Name	Crime Rate
0	San Ramon	0.31
1	Cupertino	0.66
2	Sunnyvale	1.12
3	Santa Clara	1.34
4	Milpitas	1.59
5	Brentwood	1.83
6	Alameda	1.88
7	Mountain View	1.98
8	San Mateo	2.25
9	South San Francisco	2.34
10	Redwood City	2.37
11	Pittsburg	2.59
12	Union City	2.83
13	Napa	3.13
14	San Jose	3.21
15	San Rafael	3.26
16	Petaluma	3.34
17	Berkeley	3.66
18	Concord	3.67
19	Santa Rosa	3.68
20	San Leandro	4.16
21	Antioch	7.84
22	San Francisco	7.95
23	Vallejo	8.65
24	Oakland	16.85

5. Discussion

From the above data we can see a strong correlation between house price and crime rate and school. The most expensive cities usually have good schools and low crime rates. I use RNN to further classify all cities into 5 categories.

From the numeric labels generated by RNN, I associate each label with the most significant patterns among the group as follows.

5.1 Best (Label 4) - 1 city

The most safe and expensive category has only one member: Cupertino. It has the best school and lowest crime rate. If one has resources, this is definitely a good choice.

5.2 Second (Label 1) - 3 cities

The second-best place to live with a slightly less expensive area includes 3 cities: Sunnyvale, Redwood City and Mountain View. It has excellent schools, low crime rate and city life, might be a good choice for both young families with/without kids.

5.3 Urban (Label 3) - 4 cities

The next section includes Santa Clara, San Mateo, Berkeley and San Francisco. It has more affordable houses with good schools. It is also the most densely populated category.

5.4 Middle (label 0) - 7 cities

The Middle section represents cities scored average in evaluated metrics. It fits for middle class families looking for reasonable housing costs and good schools. For example, Milpitas has a low crime rate and low population density, it is the home of many young immigrant families.

5.5 Affordable (Label 2) - 10 cities

The last section has most cities and represents a low housing cost section with relatively less performing schools and a higher crime rate. Oakland has a 4X violent crime rate than the California average.

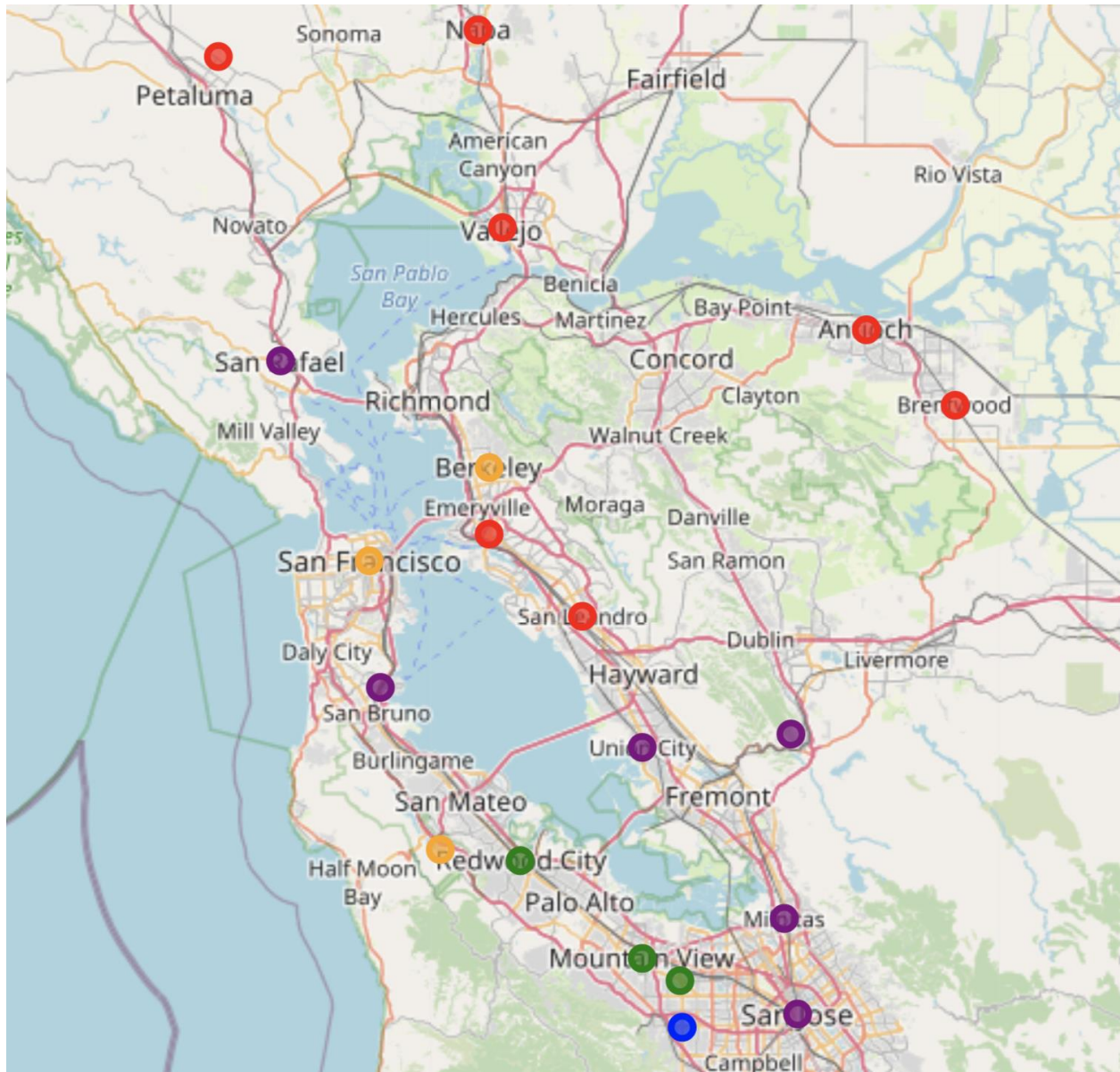
Label	Name	sq mi	population_per_sq_mi	MedPrice	Crime Rate	Cluster Labels
0	Milpitas	13.59	4914.64312	1094092	1.59	1

0	Union City	19.47	3570.416025	903742	2.83	1
0	South San Francisco	9.14	6961.925602	1014283	2.34	1
0	Alameda	10.61	6956.833176	977150	1.88	2
0	San Jose	176.53	5358.53396	1067425	3.21	1
0	San Ramon	18.06	3994.905869	1040308	0.31	3
0	San Rafael	16.47	3504.128719	1008283	3.26	1
1	Sunnyvale	21.99	6370.213734	1857625	1.12	2
1	Redwood City	19.42	3955.45829	1636433	2.37	2
1	Mountain View	12	6172.166667	1875867	1.98	2
2	Antioch	28.35	3611.005291	435133	7.84	1
2	Pittsburg	17.22	3673.867596	423358	2.59	0
2	Vallejo	30.67	3780.306488	406692	8.65	1
2	Concord	30.55	3995.646481	592467	3.67	0
2	Petaluma	14.38	4029.276773	705008	3.34	1
2	Brentwood	14.79	3480.797836	575833	1.83	1
2	Napa	17.84	4311.378924	644142	3.13	0
2	San Leandro	13.34	6368.065967	662850	4.16	1
2	Oakland	55.79	7003.477326	739217	16.85	0
2	Santa Rosa	41.5	4279.180723	579275	3.68	4
3	Santa Clara	18.41	6326.344378	1385692	1.34	0
3	San Mateo	12.13	8013.767519	1437700	2.25	2
3	Berkeley	10.47	10752.62655	1258050	3.66	2
3	San Francisco	46.87	17180.17922	1337250	7.95	2
4	Cupertino	11.26	5177.797513	2289033	0.66	2

6. Conclusion

In this project, I classified the top 25 most populous cities in the SF Bay Area in terms of housing cost, crime rate, schools, neighborhood so that people who are interested in choosing where to live can get better insights into the options this area presents.

Below is an area map with each target city color coded with its categories: Blue - Best, Green - Second, Orange - Urban, Purple - Middle, Red - Affordable.



For example, for a young family with both parents working in a high tech industry with small children, they might find good choices in Blue and Green cities. For singles, they might value a neighborhood with more restaurants, bars, they can look into more dense areas in Orange

cities. In general, one should be able to find some good choices in each of 5 groups I classified here.

What is your first choice?

Reference:

1. "MAP: The Bay Area Leads California in Population Growth": <https://www.kqed.org/news/11741275/map-the-bay-area-leads-california-in-population-growth>
2. "List of cities and towns in the San Francisco Bay Area": https://en.wikipedia.org/wiki/List_of_cities_and_towns_in_the_San_Francisco_Bay_Area
3. "Home Price": <https://www.vitalsigns.mtc.ca.gov/home-prices>
4. "California locations by crime rate": https://en.wikipedia.org/wiki/California_locations_by_crime_rate
5. "Best public high schools": <https://www.niche.com/k12/search/best-public-high-schools>
6. "Venue Search": <https://developer.foursquare.com/docs/api-reference/venues/search/>