

SF Bay Area: Battle of Neighborhoods

IBM data science course - Capstone Project

Hui Zeng

July 4, 2020

Introduction

- San Francisco Bay Area is a populous region in Northern California.
- It is a major job hub for high tech workers, the population has grown by over 600k since 2010.
- Out of 100+ cities, where should I call home?
- In this project, I want to use publicly available social and geographic data to help answer this question utilizing the Data Science tools learned during this course.

Target City Selection

- 100+ cities in SF Bay Area
- Pre-selected 25 cities based on:
 - Larger than average city size (among upper half of cities in terms of area)
 - More densely populated (based on population per square mile)

Data Source

- School Ranking from niche.com (collected manually)
- Housing cost from Vitalsign's website (in CSV file)
- Neighborhood information from Foursquare.com (Venue API)
- City list and Crime Rate are both from Wikipedia (Webpage scrape)

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:
 - Housing
 - School
 - Neighborhood
 - Crime
- I use the K-nearest neighbors (KNN) algorithm to group cities with similar metrics together to analyze common patterns.

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.

Here are the top 5 most expensive cities:

	Name	Median Price
0	Cupertino	2289033
1	Mountain View	1875867
2	Sunnyvale	1857625
3	Redwood City	1636433
4	San Mateo	1437700

School

- 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

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 (eg., Santa Clara, San Jose).
- Label 2 has more parks (Sunnyvale, Redwood City).
- Label 3 has historical sites (San Ramon).
- Label 4 has a lake! (Santa Rosa)

Crime

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

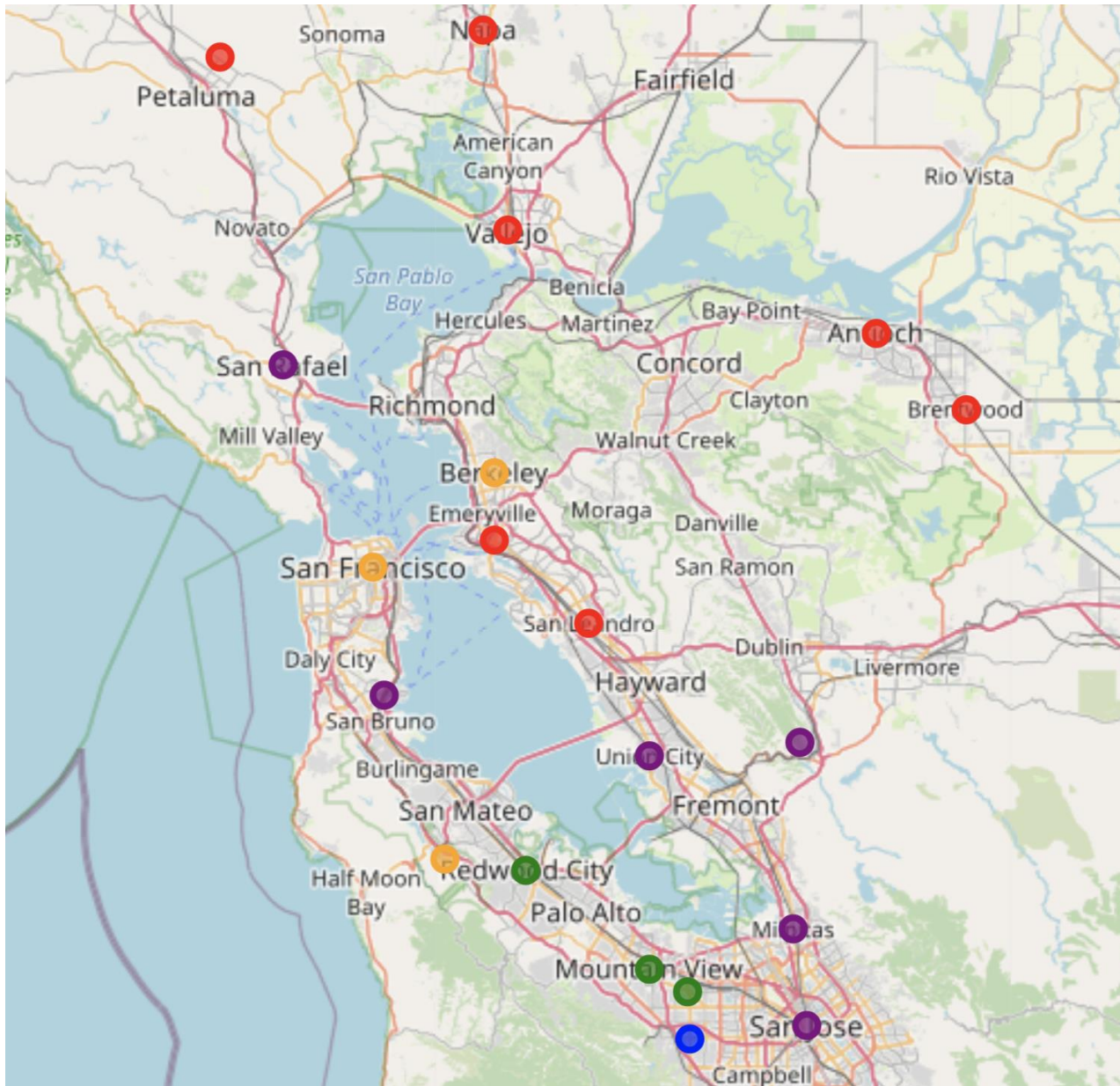
	Name	Crime Rate
0	San Ramon	0.31
1	Cupertino	0.66
2	Sunnyvale	1.12
3	Santa Clara	1.34
4	Milpitas	1.59

RNN Classification

- Use RNN to further classify all cities into 5 categories.
- 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.

Resulting Categories

- Best - 1 city:
 - only one member: Cupertino. It has the best school and lowest crime rate.
- Second - 3 cities:
 - Sunnyvale, Redwood City and Mountain View. It has excellent schools, low crime rate and city life
- Urban – 4 cities:
 - Santa Clara, San Mateo, Berkeley and San Francisco. It has more affordable houses with good schools. It is also the most densely populated category.
- Middle – 7 cities:
 - It fits for middle class families looking for reasonable housing costs and good schools.
- Affordable – 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.



Blue - Best, Green - Second, Orange - Urban, Purple - Middle, Red - Affordable.

Conclusion

- Classified the top 25 most populous cities in the SF Bay Area in terms of housing cost, crime rate, schools, neighborhood
- People who are interested in choosing where to live can get better insights into the options this area presents.
 - For a young family 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.

Thanks for Read!