San Francisco Neighborhoods and Gentrification

Priya Dhawka
February 2020

1 Introduction

1.1 Background

San Francisco is the "cultural, commercial and financial center of Northern California" (wikipedia.org). It has the largest concentration of high-tech companies in the US and is also the 13th most populous city in the US with 21 neighborhoods, which have experienced gentrification at different time periods due to a growing tech workforce.

1.2 Problem Description

We want to focus on Chinatown, San Francisco and try to understand the phenomenon of gentrification and how closely a slowly gentrifying Chinatown might resemble other more gentrified neighborhoods in San Francisco.

For the purposes of this report, we will define gentrification as the process of "changing the character of a neighborhood through the influx of more affluent residents and businesses".With a young, educated and affluent workforce comes fears of gentrification for less populous, historically poor neighborhoods. Historically poor "ethnic" neighborhoods or enclaves (geographic areas with high ethnic concentration, characteristic cultural identity and economic activity) in US cities are usually at a greater risk for gentrification.

1.3 Interest

Chinatown in San Francisco is the oldest Chinatown in North America. It is also the smallest neighborhood population-wise in San Francisco, with about 29% of its population living below the poverty line in 2013. This analysis is possibly relevant for stakeholders (urban city planners, policy makers) seeking to understand or assess the effects of gentrification on ethnic neighborhoods or for young tech professionals who might take an interest in the social effects of technological development.

2 Data Acquisition and Cleaning

2.1 Data Source

The data for San Francisco neighborhoods is taken from http://www.healthysf.org/bdi/outcomes/zipmap.htm, which contains a table with zip codes, neighborhoods and population size from a 2000 census.

2.2 Data Cleaning and Feature Selection

The data is read into a pandas dataframe using built-in pandas methods (read_html). We rename the columns of our dataframe and sort by population size to get a sense of which are the smallest neighborhoods and which are the largest neighborhoods. We make a decision about which neighborhoods to select at this stage based on our sorted dataframe. We drop the population column for which we have no use and later add the latitude and longitude information using uszipcode, an easily downloadable python zipcode database. We ignore any rows with null latitude and longitude values. We later use the Foursquare location data to get the most common venues in the selected neighborhoods.

Our resulting data frame consists of Zip Code, Neighborhood, Latitude and Longitude at this point.

3 Methodology

3.1 Preparing the data

We use the requests package to get the data from the link. We use the BeautifulSoup package to read the response text and situate the table that we need. We then use pandas read_html to read the data into a dataframe. We do further cleaning as mentioned previously to end up with the below dataframe.

	Zip Code	Neighborhood
3	94107	Potrero Hill
17	94127	St. Francis Wood/Miraloma/West Portal
15	94123	Marina
2	94103	South of Market
19	94132	Lake Merced

3.2 Visualizing the neighborhoods in San Francisco

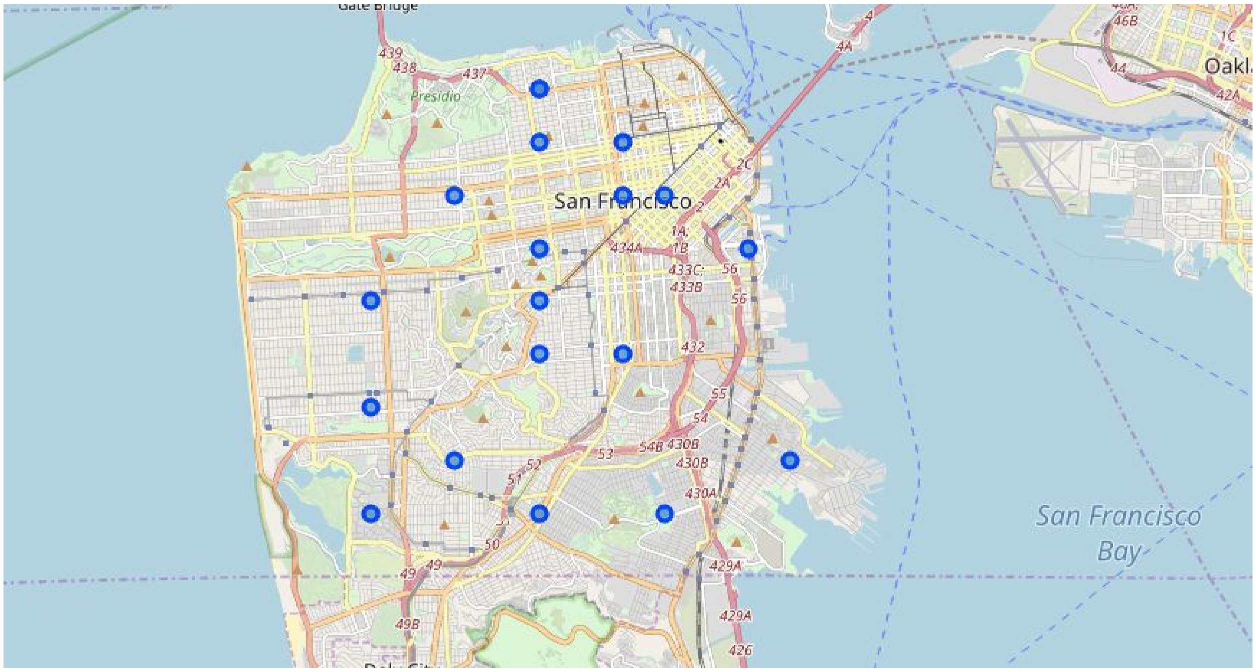
We use the uszipcode library to get the latitude and longitude information for the different neighborhoods. We drop any null values from the dataframe

Our dataframe at this point looks as follows:

Out[5]:

	Zip Code	Neighborhood	Latitude	Longitude
3	94107	Potrero Hill	37.77	-122.39
17	94127	St. Francis Wood/Miraloma/West Portal	37.73	-122.46
15	94123	Marina	37.80	-122.44
2	94103	South of Market	37.78	-122.41
19	94132	Lake Merced	37.72	-122.48

Before we start manipulating the data further, we use the folium library in python to visualize a map of San Francisco with the neighborhoods. At this point, we have no clusters in our data.



3.3 Using the Foursquare API to get the most common venues per neighborhood

The Foursquare API is used to retrieve the most common venues in the previously selected neighborhoods to gain an understanding of whether a neighborhood can be categorized as having undergone gentrification simply based on the most common venues in that neighborhood. Since signs of gentrification include an "influx of small, specialized boutiques, hipster bars, quirky/higher-end restaurants and coffee shops", for example, if a neighborhood has "Coffee shop" as the most common venue, we can assume with some certainty that the neighborhood has experienced some gentrification.

This report takes a simple approach and compares Chinatown with other more affluent ethnic neighborhoods (Japantown, Mission District (Inner Mission/Bernal Heights) and with more gentrified neighborhoods (Potrero Hill, Nob Hill,South of Market).

We select the neighborhoods that we want to explore in one dataframe and get the top 50 venues within a radius of 600 meters of each selected neighborhood. We use one-hot encoding on our venue categories before we pass our data to the KMeans algorithm.

We can get the 10 most common venues in our selected neighborhoods and an initial glance at our sorted venues gives us a sense of the common venues across our selected neighborhoods. At this point, we can see that both North Beach/Chinatown and Parkside/Forest Hill neighborhoods have “French restaurant” as their 3rd most common venue while Inner Mission/Bernal Heights has “French restaurant” as its 1st most common venue.

Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	Co
Hayes Valley/Tenderloin /North of Market	Coffee Shop	Vietnamese Restaurant	Theater	Pizza Place	Food Truck	
Inner Mission/Bernal Heights	French Restaurant	Performing Arts Venue	Sushi Restaurant	Concert Hall	Wine Bar	Vege
North Beach/Chinatown	Wine Bar	Gym / Fitness Center	French Restaurant	Deli / Bodega	Salad Place	S
Parkside/Forest Hill	Wine Bar	Gym / Fitness Center	French Restaurant	Deli / Bodega	Salad Place	S
Polk/Russian Hill (Nob Hill)	Bookstore	American Restaurant	Gift Shop	Italian Restaurant	Park	

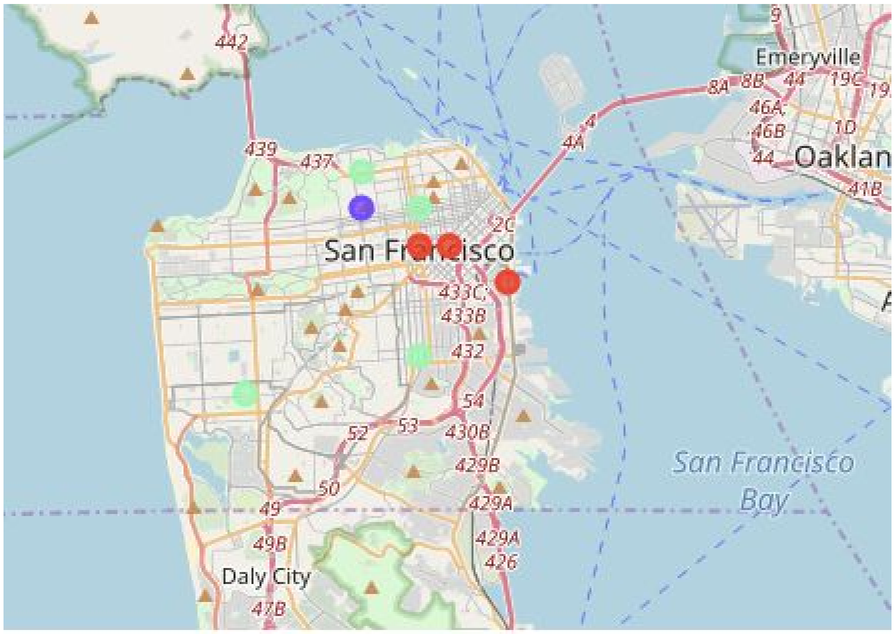
3.4 KMeans Clustering Algorithm

Given that there are common venues in our selected neighborhoods, we can choose to run our KMeans clustering algorithm. We set our number of cluster to 3, given that we are already working with a small number of neighborhoods (8 neighborhoods)

4 Results

The result of the KMeans algorithm is as follows:

We can visualize the clusters on a map of San Francisco and we can further examine our clustered neighborhoods in detail.



5 Discussion

We can analyze our clusters individually. We see that Japantown is in a cluster by itself:

Cluster 1:

Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
Western Addition/Japantown	Pizza Place	Café	Mexican Restaurant	Sandwich Place	Juice Bar

We see that the remainder of our neighborhoods are in 2 main clusters:

Cluster 0: The “gentrified” cluster

Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
Potrero Hill	Food Truck	Coffee Shop	Pharmacy	Park	Café
South of Market	Chinese Restaurant	Yoga Studio	Café	Pizza Place	Shipping Store
Hayes Valley/Tenderloin /North of Market	Coffee Shop	Vietnamese Restaurant	Theater	Pizza Place	Food Truck

Cluster 2: “The gentrifying” cluster

Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	
North Beach/Chinatown	Wine Bar	Gym / Fitness Center	French Restaurant	Deli / Bodega	Salad Place	
Parkside/Forest Hill	Wine Bar	Gym / Fitness Center	French Restaurant	Deli / Bodega	Salad Place	
Polk/Russian Hill (Nob Hill)	Bookstore	American Restaurant	Gift Shop	Italian Restaurant	Park	
Inner Mission/Bernal Heights	French Restaurant	Performing Arts Venue	Sushi Restaurant	Concert Hall	Wine Bar	V

6 Conclusion

We explored the neighborhoods in San Francisco based on population size to gain an understanding of how gentrification might affect the composition of an “ethnic” neighborhood and how the said neighborhood might then compare with other already highly gentrified neighborhoods. We see that Chinatown is clustered with our “gentrifying” cluster and the other ethnic neighborhood we compared, Japantown is in a cluster of its own.

This report can influence new immigrants or workers to San Francisco as they choose which neighborhoods they want to settle in.

One noticeable criticism of this project is that the approach to clustering neighborhoods in this project is a simple approach and needs further thought. Also, we notice that the number of neighborhoods in San Francisco is small and we can further explore the phenomenon of gentrification in cities with a larger number of neighborhoods, where we can better contrast gentrified and ethnic neighborhoods. We can include other markers of gentrification such as housing costs, demographic breakdown/changes to increase the accuracy of our model.

7 References

- San Francisco Zip codes by neighborhood: http://www.healthysf.org/bdi/outcomes/zipmap.htm

- Wikipedia page for Chinatown, San Francisco: https://en.wikipedia.org/wiki/Chinatown,_San_Francisco
- Wikipedia page for Gentrification: https://en.wikipedia.org/wiki/Gentrification
- Foursquare API