



BATTLE OF NEIGHBORHOODS

Finding the best office location in San Francisco

Introduction

San Francisco is the one of the world's most prosperous financial centers, and many business owners and entrepreneurs dream of setting open their own office in the city of San Francisco. For this project, I will use data analysis to assess the neighborhoods and try to provide some useful recommendation and insights for those who want to find the optimal place to set up their office in San Francisco.

A good office can make or break your company. It has a direct impact on your finances and a significant influence on your employees' morale and productivity. This is especially true for startup companies who, more often than not, have limited cash and need to retain their best talent. Location can play a major role in your organization's success. Let's say you locate your business in a major tech hub like San Francisco, for instance. You'll be competing for talent in an incredibly fierce employment ecosystem.

Based on the research I did, the first thing is transportation. A good place with convenient public transportations will definitely be helpful for your employees to commute and clients to visit. Also, it would be great to have some nice restaurants nearby that can provide quality meals for your employees. Having some nice restaurants can somewhat motivate employees to work harder. Moreover, it is important to consider the crime rate of the neighborhood because a safe neighborhood will reassure your employees and prevent the company from potential loss. Lastly, we surely need to consider the price of the place and see if make some tradeoff between the price and the quality of neighborhood. Foursquare API enables me to collect the

nearby restaurants and their ratings determine the whether the transportation is good based on number of public transportations nearby.

Data

Since I plan to use choropleth to show how crime rates vary across analysis neighborhoods, the first data I will use is the GeoJson file containing the borders of San Francisco neighborhoods. Also, to show how crime rates vary across neighborhoods, I am going to use the crime rate data which is in csv format, so it is easy to be converted to pandas DataFrame to create the choropleth. The above two datasets are available on San Francisco open data website (<https://datasf.org/opendata/>). Foursquare API enables me to explore some venues in each neighborhood and compare them. Some questions can be answered using Foursquare API such as: does the neighborhood have adequate restaurants? How is the quality? Does the neighborhood have convenient transportation? These questions will all be answered by looking at the data Foursquare provides.

Take Bueno Vista for example, we might want to see how many restaurants there are in the neighborhood. Then by processing the query response we can get a data frame like the following:

| | name | categories | address | lat | lng | labeledLatLngs | distance | postalCode | cc | city | state | country | formattedAddress | cr |
|---|---|------------------|----------------|-----------|-------------|---|----------|------------|----|---------------|-------|---------------|--|----|
| 0 | Kezar Bar & Restaurant | Food | 900 Cole St | 37.769476 | -122.432570 | [{"label": "display", "lat": 37.769476, "lng": ...} | 950 | 94117 | US | San Francisco | CA | United States | [900 Cole St, San Francisco, CA 94117, United ... | |
| 1 | Kezar Bar & Restaurant | Restaurant | 900 Cole St | 37.765751 | -122.449909 | [{"label": "display", "lat": 37.765751, "lng": ...} | 678 | 94117 | US | San Francisco | CA | United States | [900 Cole St (at Carl St), San Francisco, CA 94... | |
| 2 | The Restaurant at the End of the Universe | Spiritual Center | 722 Duboce Ave | 37.769290 | -122.435622 | [{"label": "display", "lat": 37.769290, "lng": ...} | 709 | 94117 | US | San Francisco | CA | United States | [722 Duboce Ave (Castro), San Francisco, CA 94... | |
| 3 | Twist Restaurant and Bar | Food | 1066 14th St | 37.767356 | -122.437492 | [{"label": "display", "lat": 37.767356, "lng": ...} | 458 | 94114 | US | San Francisco | CA | United States | [1066 14th St, San Francisco, CA 94114, United... | |
| 4 | Thailand Restaurant | Thai Restaurant | 438 Castro St | 37.761955 | -122.435128 | [{"label": "display", "lat": 37.761955, "lng": ...} | 742 | 94114 | US | San Francisco | CA | United States | [438 Castro St (btwn Market & 18th), San Franc... | bt |

And the same applies if we want to examine the accessibility of the neighborhood.

Methodology

Due to my limited ability to acquire more detailed data in each neighborhood, the following analysis will aim to find the best neighborhood to set up an office, instead of a more specific location.

Exploratory Analysis

Let's first look at the crime data first since it is vital to ensure the security of the neighborhood and reassure the clients, employees as well as the entrepreneurs themselves.

The first step is to clean the dataset and convert it into a dataframe as shown below using *groupby* function.

| Neighborhood | Count |
|-----------------------|-------|
| Bayview Hunters Point | 9370 |
| Bernal Heights | 2783 |
| Castro/Upper Market | 4683 |
| Chinatown | 3220 |
| Excelsior | 2426 |

Figure 1 Cleaned dataframe

Then sort the values in descending order to see some pattern.

| Neighborhood | Count |
|--------------------------------|-------|
| Mission | 17573 |
| Tenderloin | 16517 |
| Financial District/South Beach | 14773 |
| South of Market | 13431 |
| Bayview Hunters Point | 9370 |

Figure 2 Sorted dataframe

A bar chart is a great way to visualize the pattern in data. Here is the bar chart of crime rate in each neighborhood.

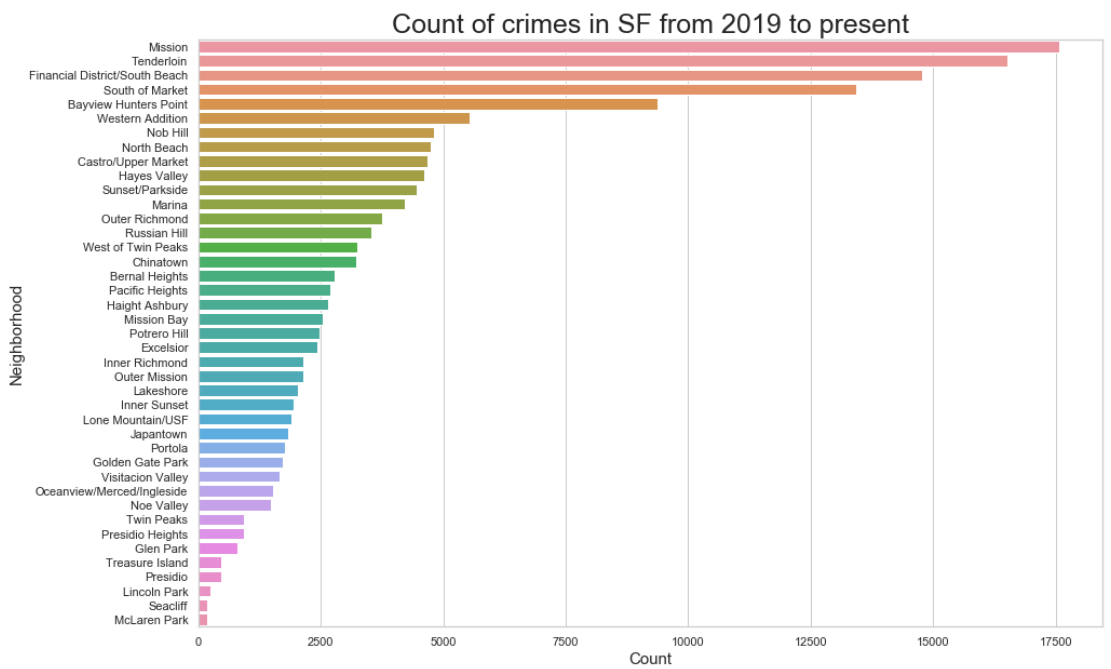


Figure 3 Bar chart of crime rate in neighborhoods

Also, a choropleth is great way to get intuitional feelings for the data above.

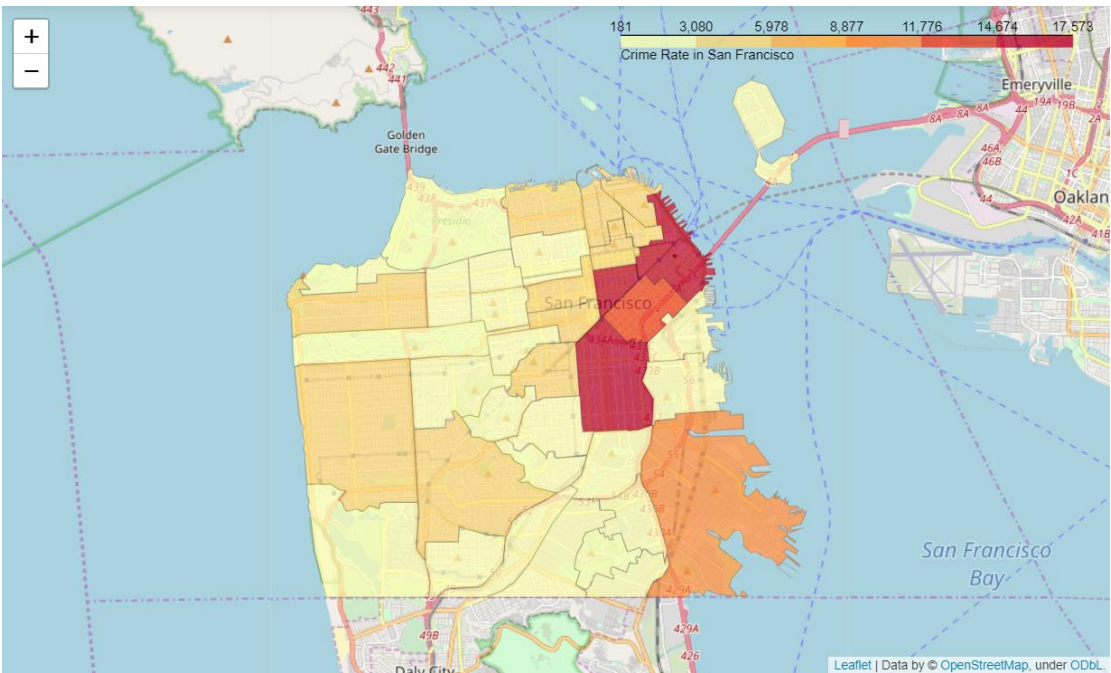


Figure 4 Interactive Choropleth of crime rate in neighborhoods

As the data shows, we can first exclude the neighborhoods with distinctly high

frequency of crimes such as Mission and Tenderloin. While it might be arbitrary to exclude some neighborhoods based on crime rate since crime rate might happen more often because the neighborhood has more business, let us simplify the problem by finding the best neighborhood that meet the previous criteria(low crime rate, convenient transportation and adequate restaurants), and leave more consideration for further discussion or a more advanced analysis.

Another thing to take into consideration as said previously is number of restaurants that are nearby the neighborhood. Here I set the radius to 1000 since it is quite accessible if the restaurant is within 1 kilometer. Then we get the following neighborhoods having at least one restaurant nearby.

| Neighborhood |
|------------------|
| Chinatown |
| Golden Gate Park |
| Inner Richmond |
| Inner Sunset |
| Japantown |
| Outer Richmond |
| Russian Hill |
| Seacliff |
| Sunset/Parkside |

After this, it's time to see if the neighborhood has high accessibility. If employees or clients want to take public transportation to the office, the station should be within 300m, then we can consider it is convenient to reach via public transportation. Using Foursquare API, the number of bus stops can be accessed easily. Here is the according data frame.

| | | | | Venue |
|------------------|-----------------------|------------------------|----|-------|
| Neighborhood | Neighborhood Latitude | Neighborhood Longitude | | |
| Sunset/Parkside | 37.7692 | -122.4481 | 33 | |
| Outer Richmond | 37.7941 | -122.4078 | 25 | |
| Inner Richmond | 37.7798 | -122.4901 | 6 | |
| Inner Sunset | 37.8011 | -122.4194 | 6 | |
| Japantown | 37.7406 | -122.4589 | 6 | |
| Golden Gate Park | 37.8037 | -122.4368 | 5 | |
| Seacliff | 37.7925 | -122.4382 | 5 | |
| Russian Hill | 37.7389 | -122.4152 | 2 | |
| Chinatown | 37.7425 | -122.4896 | 1 | |

Figure 5 Number of bus stops near each neighborhood

Also, we can also see how many subways are near the neighborhoods

| | | | | Venue |
|----------------|-----------------------|------------------------|---|-------|
| Neighborhood | Neighborhood Latitude | Neighborhood Longitude | | |
| Outer Richmond | 37.7941 | -122.4078 | 3 | |

Unfortunately, there is only one neighborhood that has subway station nearby, so I did not take subway into consideration.

From the data frame above, we can see that China town only has one bus stop nearby, so we can exclude it for further analysis.

Before we continue, it is rather obvious that Golden State Park and Seacliff are not good locations to set up an office (unless they want to work outside or rent an ocean view house as office). Thus, we can exclude them too.

Now we only have six neighborhoods. I then compared the ZRI of the neighborhoods. The Zillow Rent Index (ZRI) is a dollar-valued index intended to capture typical market

rent for a given segment. I use it to see how expensive rooms in the neighborhood are.

Here is the processed data frame. (The value of Japantown is missing)

| | RegionName | Zri |
|---|----------------|------|
| 0 | Outer Sunset | 3933 |
| 1 | Outer Richmond | 3993 |
| 4 | Inner Sunset | 4092 |
| 2 | Inner Richmond | 4135 |
| 3 | Russian Hill | 5074 |

This is not surprising since Russian Hill is famous for its upscale residential community. The other four have similar ZRI. Comparing this with previous dataframe of transportation and restaurants, we can consider **Outer Sunset, Outer Richmond and Inner Sunset** to be the optimal neighborhoods to set up an office since they have restaurants and adequate public transportation nearby, and they are relatively low-priced.

Results

From the charts and tables shown above, we can see that Outer Sunset, Outer Richmond and Inner Sunset meet criteria to be a good location for setting up an office.

Discussion

This analysis is rather contrived and simple. First, the threshold for determining whether a neighborhood is safe based on its past number of crimes is not clear, so it might be arbitrary of me to exclude 25% of the neighborhood. Second, what Foursquare API provides may not be true in real life. For example, a lot of

neighborhoods don't even have one restaurant within radius of 1km from the coordinates, which probably because the restaurant is not recorded in Foursquare's database. Third, data cannot be the only source of information. To actually determine whether a neighborhood is good to set up an office we need a lot more additional information such as neighborhood business environment, whether there are competitors and whether the population density is suitable for the business. And we should also take the business itself into consideration, since it requires different things to help the business grow. Lastly, this analysis is simple because there is not any inferential statistics involved. Due to the accessibility of data, I could not apply any machine learning algorithm to help make better decision on finding the optimal neighborhood. But further analysis can surely attempt to make this possible with more data and effort.

Conclusion

From the result shown above, we can see that Outer Sunset, Outer Richmond and Inner Sunset are the optimal neighborhoods I find that meet the criteria previously defined. And it is surely possible to improve this analysis by using machine learning algorithm or finding more data.