The Battle of Neighborhoods

IBM_Data_Science_Certificate_Capstone

INTRODUCTION

San Francisco, officially the City and County of San Francisco is the cultural, commercial, and financial center Northern California. Living and working in San Francisco is a dream for many young people and new grads, including me, not only because of the booming job opportunities, but also the cultural inclusiveness. Here comes to my question, **what is the best neighborhood to live in SF**? My audience are the young people/couples who have decent jobs in Bay area. What they consider the most are first fun places and good restaurants to hang out, second, a safe community.

DATA

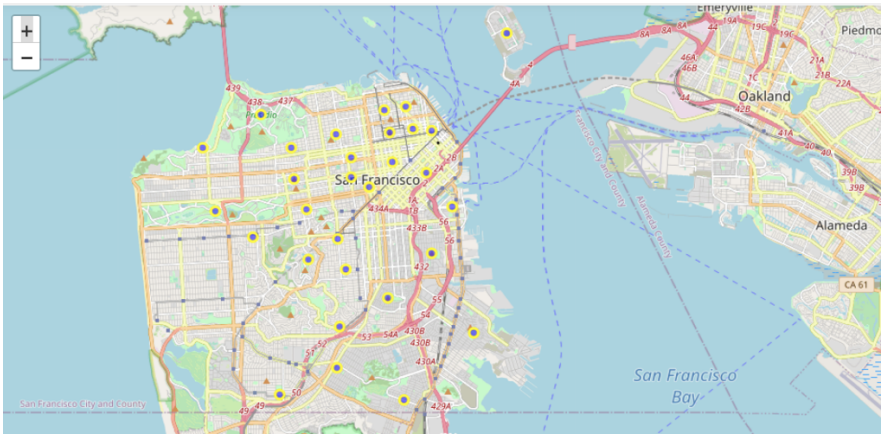
I will use the following data sets:

- A file named San_Francisco_Analysis_Neighborhood.csv downloaded from https://data.sfgov.org/Geographic-Locations-and-Boundaries/Analysis-Neighborhoods/p5b7-5n3h
  - This file will give me the list of neighborhoods in San Francisco.
- Foursquare location data
  - Where I can find the latitudes and longitudes of each neighborhood.
  - Where I can get the venues of each neighborhood.
- A file named Police_Department_Incidents_-_Previous_Year__2016.csv provided in IBM_Applied Data Science Capstone course Week_2
  - By maping the data, I can get a general idea of the crime rate of each neighborhood.

METHODOLOGY

1. We begin by collecting and cleaning the neighborhoods of San Francisco (SF).

The neighborhood 'West of Twin Peaks' cannot be found through geolocator, therefore, we delete this entry. Since we are focusing on the Bay area, we also delete neighborhoods who latitude is greater than 39.0 or less than 34.0, whose longitude is greater than -122.0. Finally, we got 32 neighborhoods in total in SF.

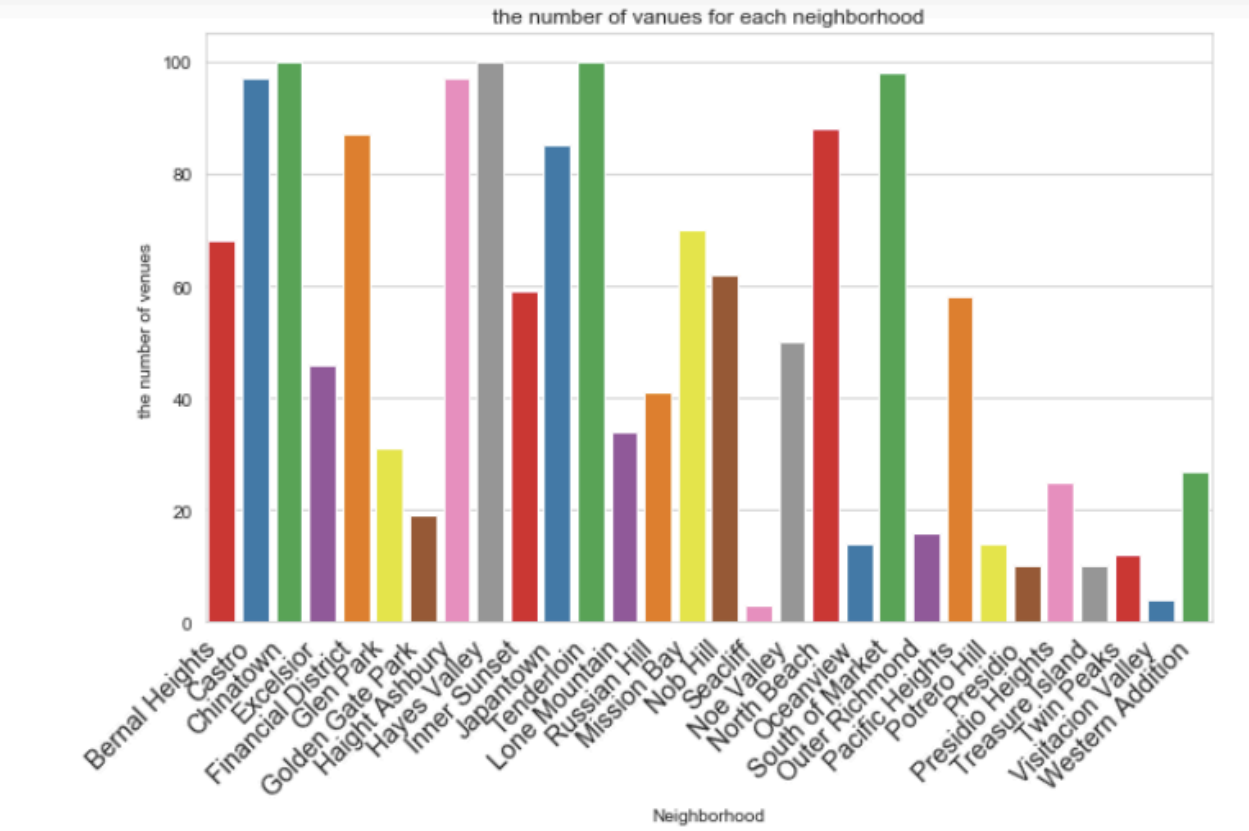


2. We then use Foursquare API to filter out the venues in the radius of 500 meters (no more than 100) for each neighborhood. We keep name, geographical coordinate, category, and id for each venue.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category	Venue id
0	Bernal Heights	37.742986	-122.415804	Bernal Heights Park	37.743265	-122.414957	Park	4aa6b24ff964a520b24a20e3
1	Bernal Heights	37.742986	-122.415804	Esmerelda Slide Park	37.743508	-122.418121	Playground	4c35116266e40f47fea5ca8b
2	Bernal Heights	37.742986	-122.415804	Bernal Heights Swing	37.742916	-122.415010	Playground	5782e416498e8701c54cbdec
3	Bernal Heights	37.742986	-122.415804	Coleridge Mini Park	37.743786	-122.419545	Scenic Lookout	4aeb400bf964a52020c021e3
4	Bernal Heights	37.742986	-122.415804	The Epicurean Trader	37.739257	-122.417712	Gourmet Shop	54e76132498e5435366425f5

The first five lines

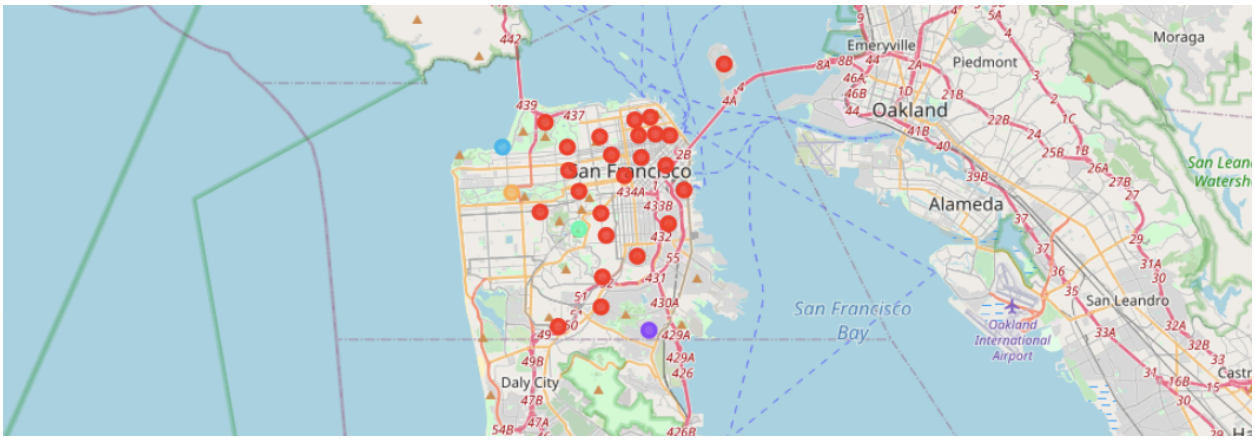
We found that some neighborhoods have quite a few venues in the radius of 500 meters, while some do not.



3. After getting a general picture, we start analyzing each neighborhood through looking at the categories of venues. We filter out 273 categories in total, such as coffee shop, Mexican restaurant, park etc, and then calculate what are the top five categories of venues for each neighborhood. For example, by looking at the top five most common venues in Berbal Heights, we could find that the top five kinds of venues are coffee shops, Mexican restaurants, Italian restaurants, and pizza place.

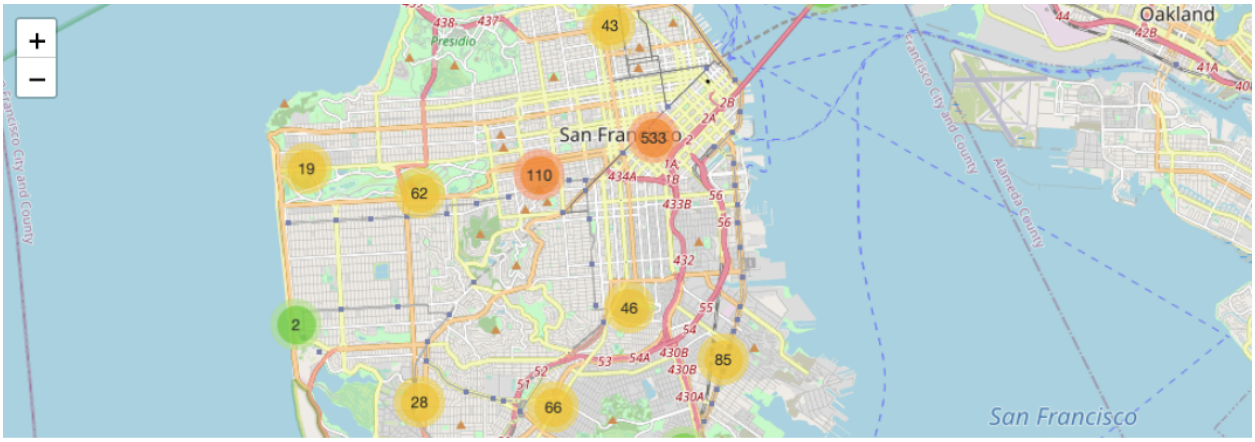
	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Bernal Heights	Coffee Shop	Mexican Restaurant	Italian Restaurant	Playground	Gourmet Shop	Grocery Store	Food Truck	Cocktail Bar	Trail	Bakery
1	Castro	Gay Bar	Coffee Shop	Thai Restaurant	Indian Restaurant	Pet Store	Yoga Studio	New American Restaurant	Mediterranean Restaurant	Deli / Bodega	Convenience Store
2	Chinatown	Italian Restaurant	Chinese Restaurant	Coffee Shop	New American Restaurant	Cocktail Bar	Bakery	Hotel	Tea Room	Bubble Tea Shop	Szechuan Restaurant
3	Excelsior	Mexican Restaurant	Bakery	Vietnamese Restaurant	Chinese Restaurant	Deli / Bodega	Pizza Place	Latin American Restaurant	Pharmacy	Sandwich Place	Bank
4	Financial District	Coffee Shop	Gym	Food Truck	Men's Store	Café	Gym / Fitness Center	Restaurant	Park	New American Restaurant	Japanese Restaurant

4. We then use K-means cluster machine learning algorithm to cluster the neighborhoods based. We set up five labels in total and get five clusters back.



5. We then use the Police_Department_Incidents_-_Previous_Year__2016.csv data to visualize the crime rate in SF area. We first download the file and find that there are 150500 entries. To make it more efficient, we only use the first 1000 entries for investigation.

6. Then we use folium to visualize the number of incidents for different areas.



RESULTS

We clustered the neighborhoods into five different clusters. The biggest cluster is cluster 1 (red). In this cluster, we can see the most common venues are restaurants, bars, gyms etc.

For the other clusters, there are only one or two neighborhoods fitting into. Those neighborhoods are first far away from the downtown SF, and have many parks, trails, or fields.

Let's look at the incidents in SF area. We can see that the closer to SF downtown, the more incidents we can see. The tourist spots such as the Golden Gate Park, the beaches are relatively safer. The southeast area out of SF downtown have fewer incidents as well.

DISCUSSION

We can see from the clustering map, the majority of the neighborhoods are good for living because you can find a lot of venues such as coffee shops, various kinds of restaurants, and gyms and yoga studios. But as far as the safety's concern, the southeast of the sf downtown, such as the neighborhood Mission bay, could be a better choice for living.

CONCLUSION

In conclusion, considering both the entertainment and safety, we think the best neighborhood to live is Mission Bay. However, we have a lot of further work to do, such as also considering data like rents and house price, school districts, using word cloud to visualize the frequency of categories.