

The Battle of Neighborhoods



pinterest.com

Charmaine Lai
IBM Data Science Professional Certificate
lai.charmaine@gmail.com
15th July 2020

Abstract

San Francisco is the nation's leader in terms of property crime and also one of the most expensive cities in the US to live in. This report assesses the different neighborhoods of San Francisco by exploring (1) housing prices (2) safety (3) convenience through different data sets. This report is used to determine the best neighborhood to find housing in for recent college graduates and concludes that neighborhoods in the South side of San Francisco has the most ideal living conditions.

Table of Contents

1. Introduction	1
1.1 Background	1
1.2 Problem	1
1.3 Objective	1
2. Data Acquisition and Preprocessing	2
2.1 SF Find Neighborhoods	2
2.2 San Francisco Police Department Incident Reports	2
2.3 Zillow San Francisco Home Prices & Values	2
2.4 FourSquare API	2
3. Methodology	3
3.1 SF Find Neighborhoods	3
3.2 San Francisco Police Department Incident Reports	3
3.3 Zillow San Francisco Home Prices & Values	5
3.4 FourSquare API	6
4. Results	8
5. Discussion	9
6. Conclusion	10
7. Appendices / References	10

1. Introduction

1.1 Background

San Francisco is the nation's leader in terms of property crime and also one of the most expensive cities in the U.S. to live in. Hence, besides considering housing prices and the convenience of the neighborhood when looking for housing, we also have to consider how safe it is.

However, due to the Fair Housing Act, land lords usually steer away from answering the questions on neighborhood safety and crime rates. Answering these questions is considered as discrimination in Real Estate Sales since mentioning the safety of the neighborhood might sway your decision in renting the apartment.

1.2 Problem

As a recent college graduate in the Bay Area, one of the biggest concerns for many of my friends and me is finding housing in the city of San Francisco. San Francisco has many neighborhoods with its own distinct culture and charm, hence the crime rates, housing prices and nearby venues / amenities can be vastly different even for juxtapose neighborhoods.

In order to find the perfect neighborhood, I wrote down a list of criteria for the apartment search I should consider (highest priority first):

- Safety
- Moderately busy area (mix between commercial and residential)
- Close to the grocery store (walkable distance)
- 1 bed 1 bath apartment (avg price)

Distance to the office is usually an important factor in choosing housing too, but as many of us have a work from home option, it will be disregarded in this project.

1.3 Objective

This project aims to explore different neighborhoods in San Francisco and find the best area for recent college graduates in the Bay Area to move to. This study will mainly focus on the **safety**, **housing prices** and **neighborhood amenities** (such as grocery store, coffee shops etc).

2. Data Acquisition and Preprocessing

2.1 SF Find Neighborhoods

- **Purpose:** Define all neighborhoods in the city of San Francisco
- **Source:** Open data from San Francisco government
- **Summary:** This data set includes neighborhood boundaries defined in 2006 by the Mayor's Office of Neighbor Services. All boundaries are for the purpose of designating general locations of neighborhoods in the city of San Francisco
- **Note:** Boundaries of neighborhoods are not "hard" lines of division

2.2 San Francisco Police Department Incident Reports

- **Purpose:** Determine safety of neighborhood by evaluating frequency of crimes
- **Source:** Open data from San Francisco government
- **Summary:** This dataset includes police incident reports filed by officers and by individuals through self-service online reporting for non-emergency cases from 2019 to Present (6/30/20)
- **Note:** The San Francisco Police Department does not guarantee accuracy or completeness of the information as the data is subject to change as modifications and updates are completed

2.3 Zillow San Francisco Home Prices & Values

- **Purpose:** Determine the average rental price for each neighborhood
- **Source:** Zillow
- **Summary:** This dataset includes the median valuation (Zillow Home Value Index) for each neighborhood in San Francisco in the past year for 1 bedroom housings.
- **Note:** Prices on Zillow are not guaranteed accuracy or completeness of the information

2.4 FourSquare API

- **Purpose:** Find top 10 popular venue categories in each neighborhood and create a cluster to determine which neighborhood has the best amenities and environment
- **Source:** Foursquare API
- **Summary:** This dataset includes popular venues and its category from the Foursquare database

3. Methodology

3.1 SF Find Neighborhoods

The data contains the boundaries of each San Francisco neighborhood. Hence, the coordinates for the neighborhoods are shaped in multi polygons. In order to find a specific latitude and longitude for the neighborhood, I found the centroid points for each multi polygon (middle point of the neighborhood) and added that to my data.

		name	latitude	longitude
0		Seacliff	37.787944	-122.487274
1		Lake Street	37.786151	-122.480569
2		Presidio National Park	37.799268	-122.467974
3		Presidio Terrace	37.786930	-122.467957
4		Inner Richmond	37.781112	-122.468754
...	
112		Corona Heights	37.764403	-122.439767
113		Ashbury Heights	37.761947	-122.448376
114		Eureka Valley	37.758053	-122.441104
115		St. Francis Wood	37.737094	-122.463378
116		Sherwood Forest	37.737710	-122.456670

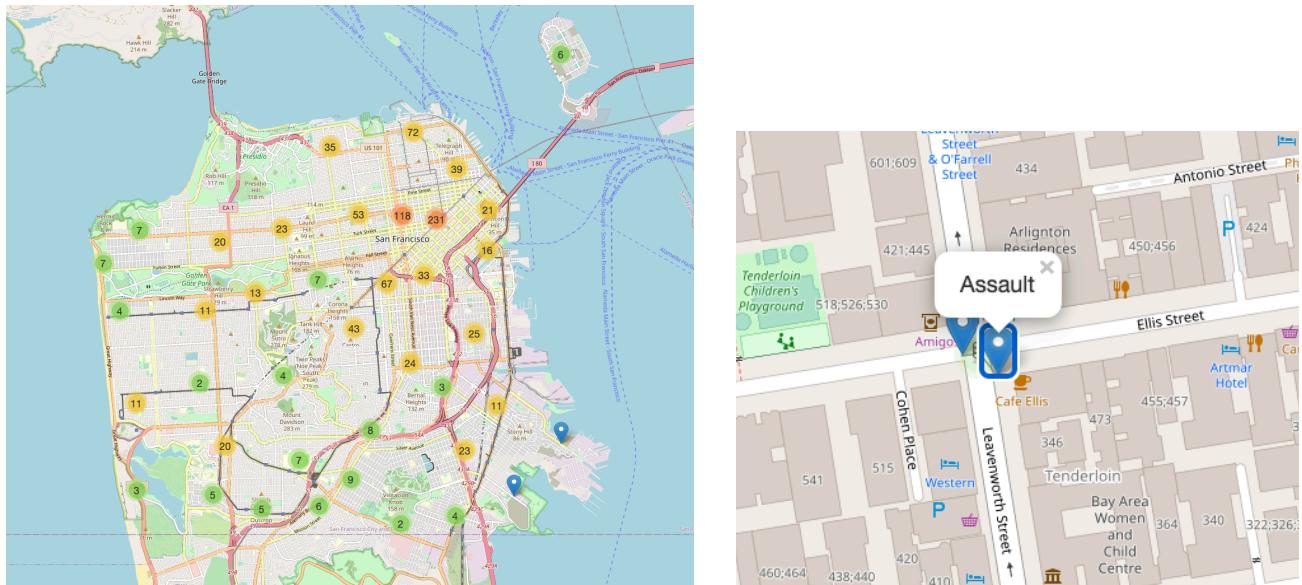
3.2 San Francisco Police Department Incident Reports

Non-violent crimes and unnecessary data are dropped from the table. For analysis, only “Open and Active”, or “Cite or Arrest” crimes was used and crimes were sorted by its Incident Category. Only violent and property crimes within the Incident Category was used and other variables, including “miscellaneous” was dropped. Data rows that contain empty cells were also removed.

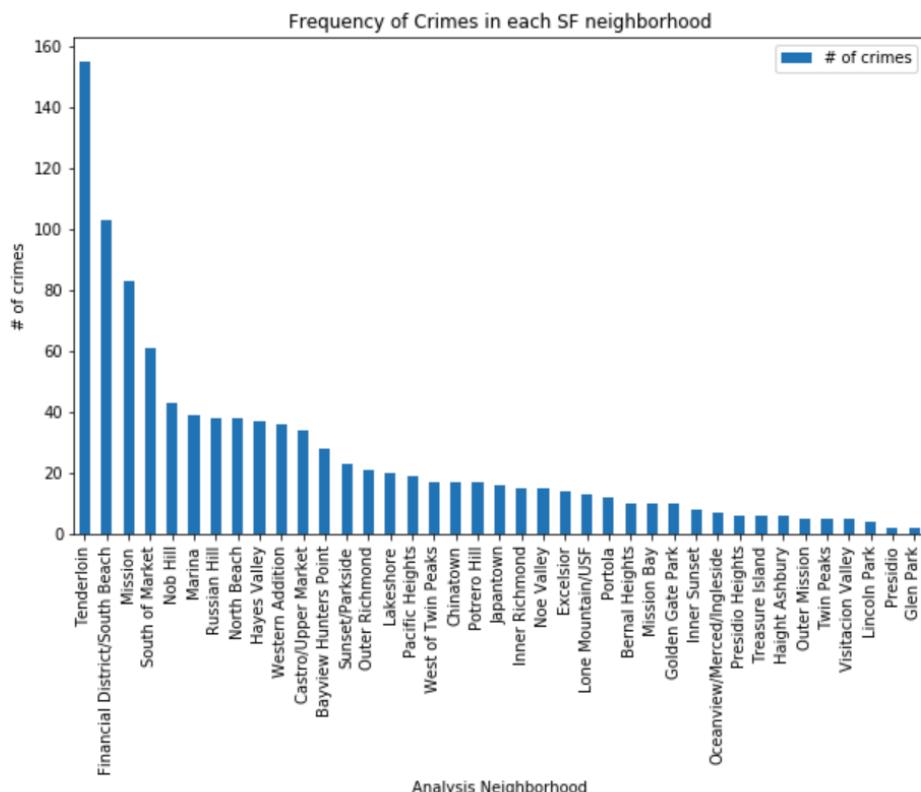
The data frame consists of 65,538 crimes, which took place in the past year. In order to reduce computational cost, we will only be working with the **first 1000 incidents** in the data set.

index	Date	Time	Number	Category	Description	Resolution	Analysis Neighborhood	Latitude	Longitude	point
0	1 2019/06/30 00:00		190473917	Larceny Theft	Theft, From Locked Vehicle, >\$950	Open or Active	South of Market	37.782119	-122.407255	(37.78211941829001, -122.40725467595034)
1	2 2019/06/30 00:00		196142437	Larceny Theft	Theft, Other Property, >\$950	Open or Active	Inner Richmond	37.783189	-122.458998	(37.783188663923646, -122.45899840209343)
2	3 2019/06/30 00:00		196133721	Larceny Theft	Theft, From Locked Vehicle, >\$950	Open or Active	Lincoln Park	37.781577	-122.494520	(37.78157729391555, -122.49451975275728)
3	7 2019/06/30 00:00		196129974	Larceny Theft	Theft, From Locked Vehicle, >\$950	Open or Active	Lakeshore	37.718956	-122.500221	(37.71895555784868, -122.50022095522831)
4	8 2019/06/30 00:00		196127837	Larceny Theft	Theft, Other Property, 50–200	Open or Active	Nob Hill	37.788223	-122.420210	(37.788222699672026, -122.42021043633935)
...
995	2106 2019/07/04 09:25		190483615	Assault	Battery	Open or Active	Tenderloin	37.784657	-122.414430	(37.78465714011983, -122.41443019938555)
996	2112 2019/07/04 09:30		190484704	Motor Vehicle Theft	Vehicle, Stolen, Truck	Open or Active	Oceanview/Merced/Ingleside	37.714301	-122.463500	(37.714300778842734, -122.46349969867914)
997	2114 2019/07/04 09:45		190483819	Larceny Theft	Theft, From Locked Vehicle, >\$950	Open or Active	Potrero Hill	37.757884	-122.389248	(37.757883662734024, -122.38924822480189)
998	2116 2019/07/04 10:00		190504998	Motor Vehicle Theft	Vehicle, Stolen & Recovered	Open or Active	Bayview Hunters Point	37.721408	-122.393737	(37.72140845975932, -122.3937369294708)
999	2118 2019/07/04 10:00		190498410	Motor Vehicle Theft	Vehicle, Stolen, Auto	Open or Active	South of Market	37.777372	-122.399763	(37.77737199450485, -122.39976258647586)

I then used the **python folium library** to visualize geographic details of the city of San Francisco and where the crime took place. Each cluster represents the number of crimes in each approximate neighborhood. And if you zoom into the map, the location and category of the crime will show.



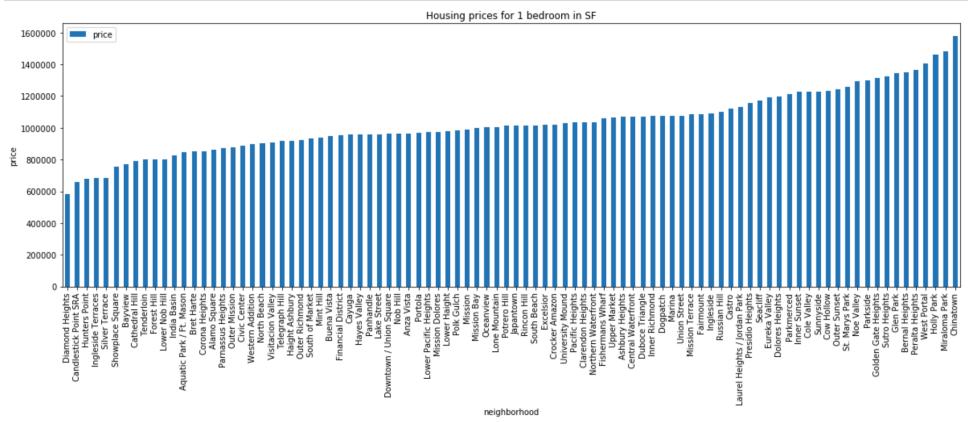
To further visualize the data and the safety of the neighborhood, I created a bar graph to show the **frequency of crimes** occurring per neighborhood in descending order.



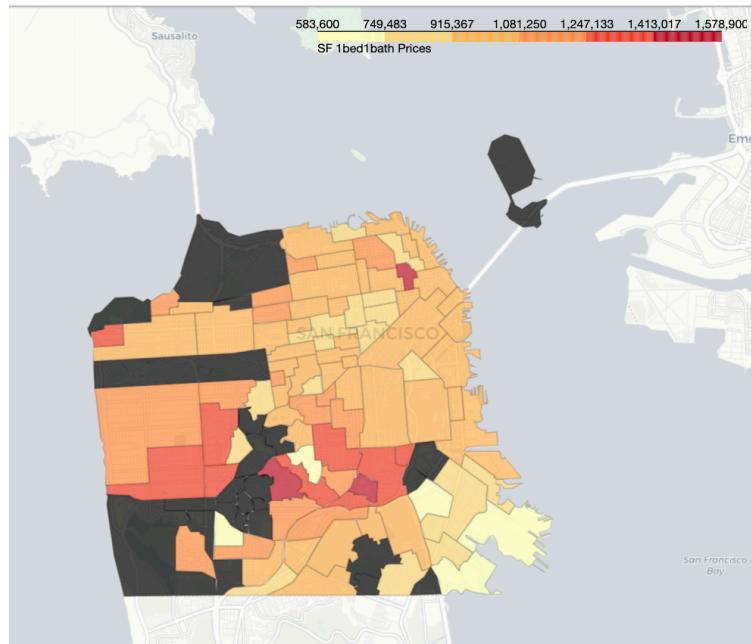
3.3 Zillow San Francisco Home Prices & Values

Zillow provides a comprehensive list for the average 1 bedroom home prices for each neighborhood in the city of San Francisco. However, the naming of the neighborhood differs slightly from the one the government provided in the SF Find Neighborhoods dataset. After cleaning the dataset, I **joined** the data frame from SF Find Neighborhoods with this dataset, combining **location** and **price**. We can then visualize the **housing prices** of different neighborhoods in a bar graph shown below.

	name	latitude	longitude	price
0	Diamond Heights	37.743098	-122.440356	583600
1	Candlestick Point SRA	37.714110	-122.388192	656800
2	Hunters Point	37.725608	-122.371341	677700
3	Ingleside Terraces	37.726545	-122.467861	682400
4	Silver Terrace	37.734842	-122.400505	683700
...
87	Peralta Heights	37.745488	-122.407347	1364200
88	West Portal	37.740790	-122.467424	1406400
89	Holly Park	37.736433	-122.421655	1461100
90	Miraloma Park	37.738863	-122.448326	1481800
91	Chinatown	37.794466	-122.406601	1578900



To further analyze the data and visualize where the neighborhoods are, I used the python folium library to create a **choropleth map** by combining the **boundaries of San Francisco** provided by the government and the Zillow dataset for housing prices. The grey areas are not listed on the Zillow dataset, hence there is no pricing information for it.

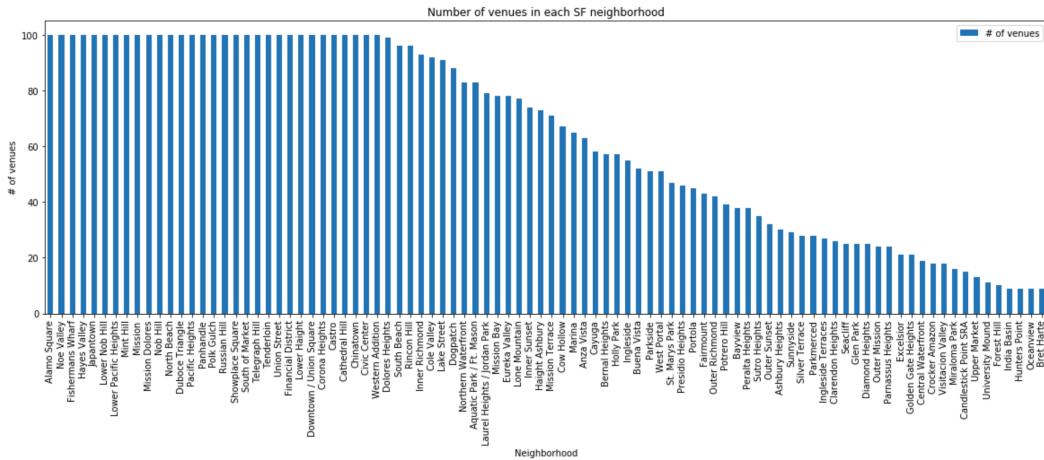


3.4 FourSquare API

I then used the Foursquare API to explore the neighborhoods and what it provides. The limit is set to **100 venues** with a **radius of 700 meters** as SF is a very dense city. **5920** venues and **367** unique categories were returned by FourSquare for all neighborhoods combined:

Neighborhood		Neighborhood Latitude		Neighborhood Longitude		Venue	Venue Latitude	Venue Longitude	Venue Category
0	Diamond Heights	37.743098		-122.440356	Glen Canyon South Rim Cliff Trail	37.742359	-122.441423		Trail
1	Diamond Heights	37.743098		-122.440356	Harbor Villa Restaurant	37.743755	-122.439404	Dim Sum Restaurant	
2	Diamond Heights	37.743098		-122.440356	Upper Douglass Dog Park	37.746493	-122.438764		Dog Run
3	Diamond Heights	37.743098		-122.440356	Christopher Playground	37.743496	-122.440077		Playground
4	Diamond Heights	37.743098		-122.440356	Great Clips	37.743857	-122.439989	Salon / Barbershop	
...
5916	Chinatown	37.794466		-122.406601	Pure Organic Spa	37.789286	-122.409191		Spa
5917	Chinatown	37.794466		-122.406601	La Fromagerie	37.790580	-122.402360	French Restaurant	
5918	Chinatown	37.794466		-122.406601	Serenity MedSpa	37.789102	-122.404292		Spa
5919	Chinatown	37.794466		-122.406601	Glow Yoga & Wellness	37.800125	-122.409114		Yoga Studio
5920	Chinatown	37.794466		-122.406601	Baonecci Ristorante	37.799697	-122.407783	Italian Restaurant	

We can also visualize the **number of venues** that each San Francisco neighborhood has with a limit of 100 venues:



In order to find out the type of neighborhood, I created a data frame displaying the **top 10 venue categories** for each neighborhood. This will give me a good sense of what is around and how convenient everything is.

	name	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	Alamo Square	Coffee Shop	Café	Bar	Sushi Restaurant	Record Shop	Dive Bar	Nightclub	Boutique	Ethiopian Restaurant	Mediterranean Restaurant
1	Anza Vista	Coffee Shop	Café	Sandwich Place	Sushi Restaurant	Mexican Restaurant	Bus Line	Gym / Fitness Center	Cosmetics Shop	Grocery Store	Burrito Place
2	Aquatic Park / Ft. Mason	Art Gallery	Food Truck	Theater	Chocolate Shop	Park	Historic Site	National Park	Vegetarian / Vegan Restaurant	Gift Shop	Bike Rental / Bike Share
3	Ashbury Heights	Park	Trail	Breakfast Spot	Bus Station	Wine Bar	Convenience Store	Organic Grocery	Mexican Restaurant	Cheese Shop	Grocery Store
4	Bayview	Southern / Soul Food Restaurant	Bakery	Light Rail Station	Café	Mexican Restaurant	Market	Thrift / Vintage Store	Theater	Chinese Restaurant	Gym

Since San Francisco is a very dense city with many neighborhoods, I decided to run **K-Means** to cluster the general neighborhood into **5 clusters**. Each neighborhood is special in its own way but for simplicity, this report will only evaluate the atmosphere of 5 "neighborhoods". Below is my merge table with the cluster labels.

name	latitude	longitude	price	Cluster Labels	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
Diamond Heights	37.743098	-122.440356	583600	2	Park	Trail	Bus Station	Scenic Lookout	Playground	Coffee Shop	Shopping Mall	Outdoors & Recreation	Pharmacy	Athletics & Sports
Candlestick Point SRA	37.714110	-122.388192	656800	4	Football Stadium	American Restaurant	Park	Mountain	Soccer Field	Food & Drink Shop	Stadium	Campground	Health & Beauty Service	Business Service
Hunters Point	37.725608	-122.371341	677700	0	Harbor / Marina	Food Service	Restaurant	Grocery Store	Scenic Lookout	Home Service	Spa	Park	Art Gallery	Field
Ingleside Terraces	37.726545	-122.467861	682400	0	Chinese Restaurant	Yoga Studio	Food Truck	Playground	Gym	Noodle House	Coffee Shop	College Gym	Café	Comic Shop
Silver Terrace	37.734842	-122.400505	683700	0	Chinese Restaurant	Grocery Store	Nightclub	Bubble Tea Shop	Diner	Gas Station	Soccer Field	Pet Service	Baseball Field	Pharmacy

I then examine each cluster to give it proper cluster labels.

Cluster 1: "Social Venues / Downtown"

	name	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	Hunters Point	Harbor / Marina	Food Service	Restaurant	Grocery Store	Scenic Lookout	Home Service	Spa	Park	Art Gallery	Field
1	Ingleside Terraces	Chinese Restaurant	Yoga Studio	Food Truck	Playground	Gym	Noodle House	Coffee Shop	College Gym	Café	Comic Shop
2	Silver Terrace	Chinese Restaurant	Grocery Store	Nightclub	Bubble Tea Shop	Diner	Gas Station	Soccer Field	Pet Service	Baseball Field	Pharmacy
3	Showplace Square	Furniture / Home Store	Art Gallery	Coffee Shop	Food Truck	Gay Bar	Gym	Music Venue	Clothing Store	Bar	Grocery Store
4	Bayview	Southern / Soul Food Restaurant	Bakery	Light Rail Station	Café	Mexican Restaurant	Market	Thrift / Vintage Store	Theater	Chinese Restaurant	Gym
...
72	Bernal Heights	Coffee Shop	Bakery	Italian Restaurant	Furniture / Home Store	Playground	Pizza Place	Gourmet Shop	Park	Farmers Market	Mexican Restaurant
73	Peralta Heights	Brewery	Fast Food Restaurant	Convenience Store	Park	Grocery Store	Mexican Restaurant	Liquor Store	Deli / Bodega	Nightclub	Café
74	West Portal	Coffee Shop	Indian Restaurant	Mexican Restaurant	Burger Joint	Chinese Restaurant	Pub	Pizza Place	Park	Grocery Store	Italian Restaurant
75	Holly Park	Coffee Shop	Mexican Restaurant	Latin American Restaurant	Italian Restaurant	Gourmet Shop	Playground	Pizza Place	Grocery Store	Café	Bakery
76	Chinatown	Coffee Shop	Hotel	Men's Store	New American Restaurant	Italian Restaurant	Bubble Tea Shop	Gym	Café	Bakery	Restaurant

Cluster 2: "Home & Lifestyle"

	name	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	Bret Harte	Brewery	Breakfast Spot	Burger Joint	Food Truck	Latin American Restaurant	Park	Light Rail Station	Mexican Restaurant	Grocery Store	Fish Market
1	Visitation Valley	Coffee Shop	Grocery Store	Park	Construction & Landscaping	Vietnamese Restaurant	Marijuana Dispensary	Light Rail Station	Trail	Sandwich Place	Café
2	Oceanview	Light Rail Station	Supermarket	Park	Grocery Store	Food	Spanish Restaurant	Thai Restaurant	Playground	Coffee Shop	Filipino Restaurant
3	Sunnyside	Park	Light Rail Station	Tennis Court	Food Truck	Deli / Bodega	Pool	Mountain	Dance Studio	Baseball Field	Bar

Cluster 3: "Parks and Scenic Trails"

	name	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	Diamond Heights	Park	Trail	Bus Station	Scenic Lookout	Playground	Coffee Shop	Shopping Mall	Outdoors & Recreation	Pharmacy	Athletics & Sports
1	Parnassus Heights	Trail	Park	Garden	Coffee Shop	ATM	Mountain	Cafeteria	Monument / Landmark	Gift Shop	Salad Place
2	Clarendon Heights	Trail	Scenic Lookout	Park	Bus Station	Grocery Store	Monument / Landmark	Tailor Shop	Art Gallery	Bus Stop	Road
3	Upper Market	Trail	Scenic Lookout	Park	Coffee Shop	Hill	Reservoir	American Restaurant	Thai Restaurant	Zoo Exhibit	Fabric Shop
4	Ashbury Heights	Park	Trail	Breakfast Spot	Bus Station	Wine Bar	Convenience Store	Organic Grocery	Mexican Restaurant	Cheese Shop	Grocery Store
5	Golden Gate Heights	Park	Trail	Playground	Garden	Scenic Lookout	Event Space	Dance Studio	Food Court	Fabric Shop	Falafel Restaurant
6	Glen Park	Park	Trail	Playground	Coffee Shop	Shopping Mall	Grocery Store	Tennis Court	Outdoors & Recreation	Pharmacy	Salon / Barbershop
7	Miraloma Park	Park	Trail	Monument / Landmark	Bus Line	Mountain	Grocery Store	Athletics & Sports	Playground	Tree	Outdoors & Recreation

Cluster 4: "Laid Back Social Venues"

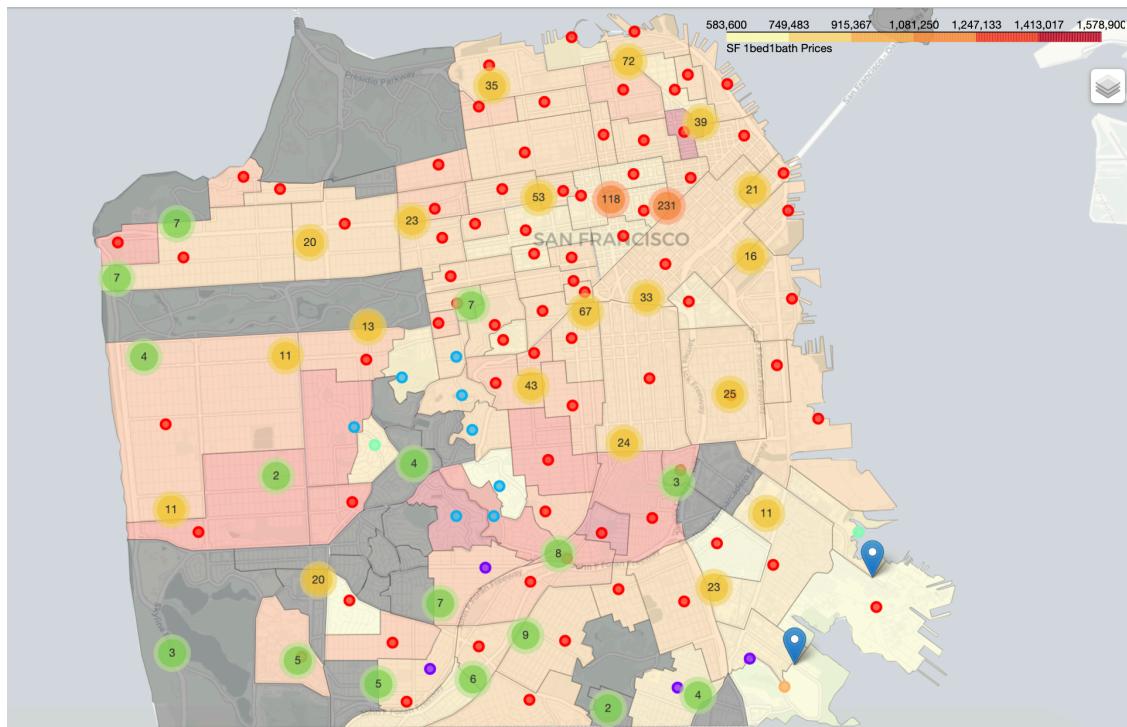
	name	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Forest Hill	Park	French Restaurant	Garden	Scenic Lookout	Playground	Hotpot Restaurant	Japanese Restaurant	Food Service	Food Court	Fabric Shop
1	India Basin	Park	Spa	Coffee Shop	Brewery	Garden Center	Waterfront	Home Service	Farm	Farmers Market	Fast Food Restaurant

Cluster 5: "Sports"

	name	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	Candlestick Point SRA	Football Stadium	American Restaurant	Park	Mountain	Soccer Field	Food & Drink Shop	Stadium	Campground	Health & Beauty Service	Business Service

4. Results

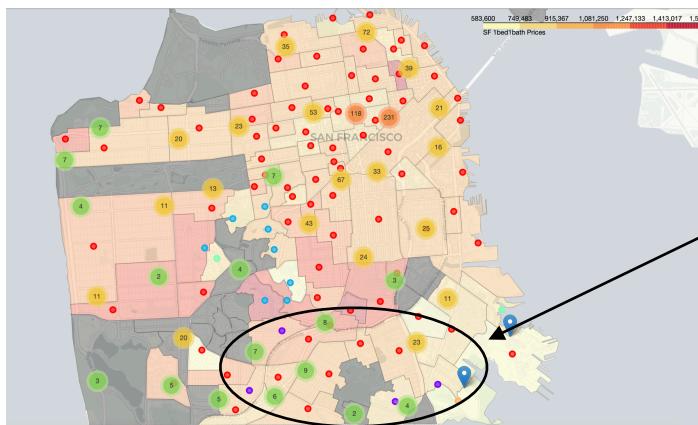
I combined all the data above into a Choropleth Map so we can visualize the results:



From the map above, we can see a relationship between **crime rates**, **housing prices** and **neighborhood atmosphere**. The **housing prices are the lowest** where the **crime rates are the highest** (neighborhoods north of San Francisco, i.e. Tenderloin). The highest housing prices surround the blue dots (cluster 3) which are **scenic trails and parks**. Large landscapes and nearby green spaces tend to be more desirable places to live in. The Financial District of San Francisco also has the highest housing prices. **Housing prices and crime rates lower** as we **move away from downtown** San Francisco (north) and towards the **South**. The outskirts of San Francisco have a lot less crime with average housing prices. However, venues and amenities nearby are **more scattered** and not as convenient as the neighborhoods located downtown.

5. Discussion

After analyzing the data above, the ideal neighborhood to move to as a recent college graduation seems to be neighborhoods located in the South of San Francisco, near Cluster 2 (purple dots).



Neighborhoods include: Sunnyside, Oceaview, Visitacion Valley, Bret Harte

I then evaluated the neighborhoods against the objectives I made towards the beginning:

- Safety
 - Low crime rates around the area
- Moderately busy area (mix between commercial and residential)
 - Many transportation options to the city
- Small shops and boutiques
 - Sunnyside and Visitacion Valley has more venues
- Close to the grocery store (walkable distance)
 - Grocery store in top 5 common venues
- 1 bed 1 bath apartment (avg price)
 - Average housing price is towards the low end (~\$800,000), assuming that rental prices are in proportion to housing prices

6. Conclusion

Sunnyside, Oceanview, Visitacion Valley and Bret Harte seem to be good neighborhoods for recent college graduates to move to as it fits the criteria I set in the beginning. There are relatively low crime rates and moderately priced 1 bedroom housing. The area is also a mix between residential and commercial shops, with many groceries stores within walkable distance.

For a more detailed report in the future, the full data set should be utilized and further expanded. Housing prices data that was drawn from the Zillow Housing Index might not be an accurate reflection of real time rental prices. In future studies, I could find rental prices for 1 bed room apartments per neighborhood online and calculate the average rental price per neighborhood. This will be a better reflection of real time housing prices.

7. Appendices / References

Datasets:

[1] SF Find Neighborhoods

https://data.sfgov.org/d/pty2-tcw4?category=Geographic-Locations-and-Boundaries&view_name=SF-Find-Neighborhoods

[2] San Francisco Police Department Incident Reports

<https://data.sfgov.org/Public-Safety/Map-of-Police-Department-Incident-Reports-2018-to-/jq29-s5wp>

[3] Zillow San Francisco Home Prices & Values

<https://www.zillow.com/san-francisco-ca/home-values/>

[4] FourSquare API

<https://foursquare.com/>

References:

[5] Housing Rights

http://www.housingrights.com/?page_id=68

[6] Coursera IBM Data Science Professional Certificate Course Modules and Labs