Capstone Project - Battle of the Neighborhoods¶

Applied Data Science Specialization by IBM

1. Introduction
Definition of the business problem and the target audience.

2. Data
Description of the data and sources used to solve the problem.

3. Methodology
Discussion of the exploratory data analysis, the inferential statistical testing, and the machine learnings that were used.

4. Results
Presentation of the results.

5. Discussion
Observations and recommendations based on the results.

6. Conclusion
Final thoughts on the report.

1. Introduction

The owners of a successful gallery in Manhattan are looking at the prospect of expanding their business. Having considered many options for growth, they came to the conclusion that opening a second physical location is going to be their next move and San Fransisco was determined as the most ideal city for their target clientele.

Our objective is to use data science in order to **choose the ideal location for the new gallery.**

We will focus on some of the most *critical factors* that will contribute to the gallery's profitability, such as:

1. Neighborhood safety, accessibility and visibility
2. Surrounding businesses and possible competitors
3. Space size and affordability
4. Comparative analysis of the neighborhoods of the existing and proposed locations

2. Data

In order to approach the above-mentioned factors we will need to use different types of data from the available sources.

- To determine the safety factor, we will use the *San Francisco Police Department's dataset* that can provide us with incident reports filed by officers and by individuals through self-service online reporting for non-emergency cases.
- To estimate how accessible the potential location will be by public or private means, we will refer to *San Francisco's public transportation geodata* such as subway and bus stations as well as the *MTA's on- and off-street parking census*.
- Using *Foursquare* to extract the most popular venues in the city we will establish the visibility factor that different neighborhoods can provide.
- *Foursquare* analysis will be also used to avoid competition while simultaneously benefit from proximity with complimentary venues such as cafes and restaurants.
- Information regarding spaces sizes and rending costs will be gathered from *real estate agencies* and *accommodation search engines*.
- Since the existing gallery location is proved to have contributed to a successful business model in New York, we will critically compare the data for the two cities, in order to support our final location proposal.

3. Methodology

Firstly, download all the dependencies that are going to be used for the data analysis.

Import all the necessary libraries such as **numpy, matplotlib** and **pandas** as well as libraries to *convert an address into latitude and longitude values, to handle requests, transform JSON file into a pandas dataframe, map rendering library, data scraping library* etc.
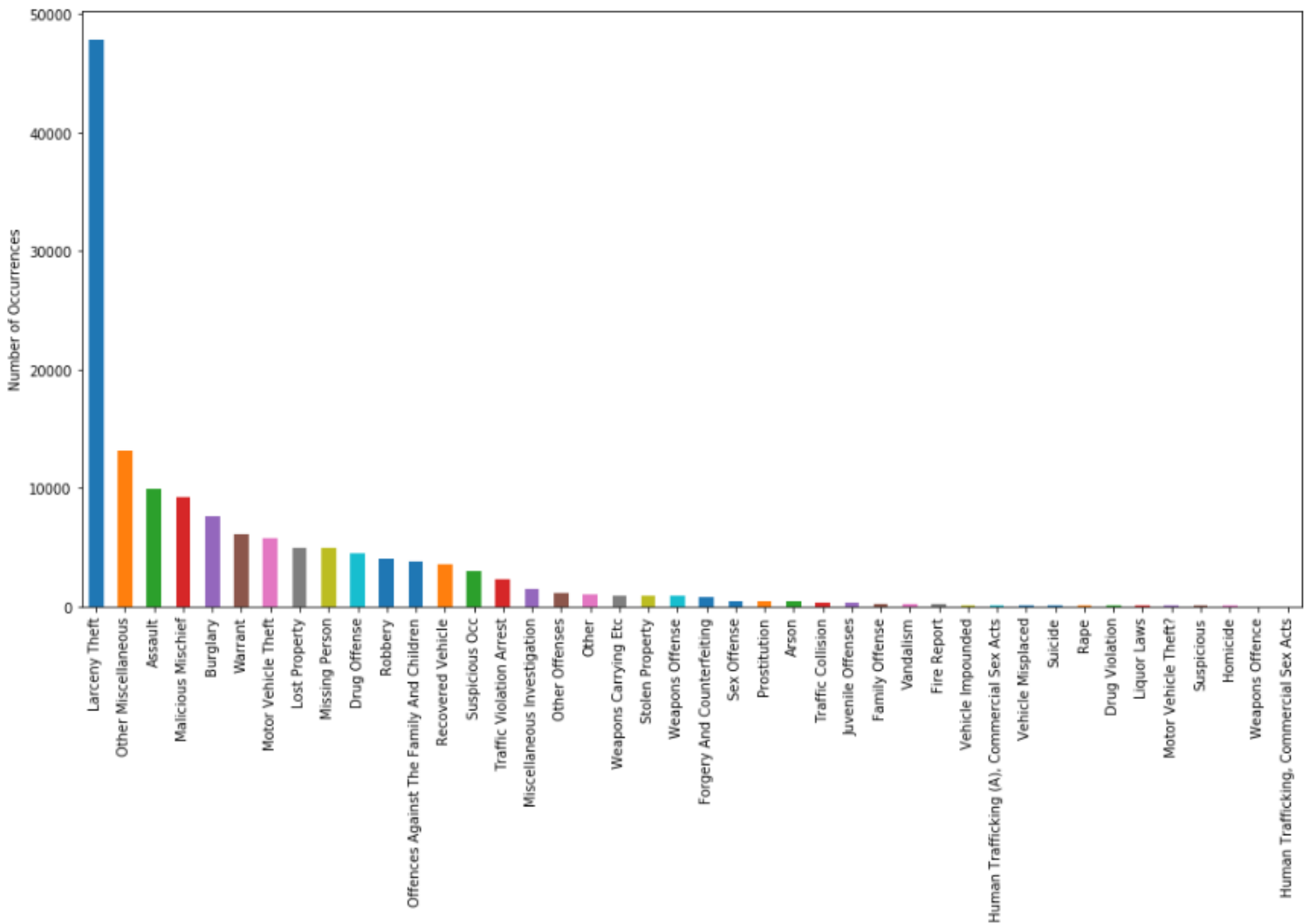
3.1 Safety factor analysis

Load and explore the San Francisco Police Department's dataset (from https://data.sfgov.org/d/wg3w-h783). Cleanup the dataframe, by removing unnecessary columns and values.

	Incident Date	Incident Time	Incident Number	Incident Category	Police District	Latitude	Longitude
0	2018/10/22	10:50	180801798	Burglary	Northern	37.775422	-122.419397
1	2018/08/24	14:50	180638466	Missing Person	Taraval	37.738900	-122.485074
2	2018/10/14	14:54	180807520	Larceny Theft	Taraval	37.720450	-122.483015
3	2018/10/14	14:54	180807520	Suspicious Occ	Taraval	37.720450	-122.483015
4	2018/10/18	16:40	180791014	Drug Offense	Southern	37.775161	-122.403636

Count the unique incident occurrences

Identify and analyze the different incidents, their occurrences and relevance to a neighborhood's safety.
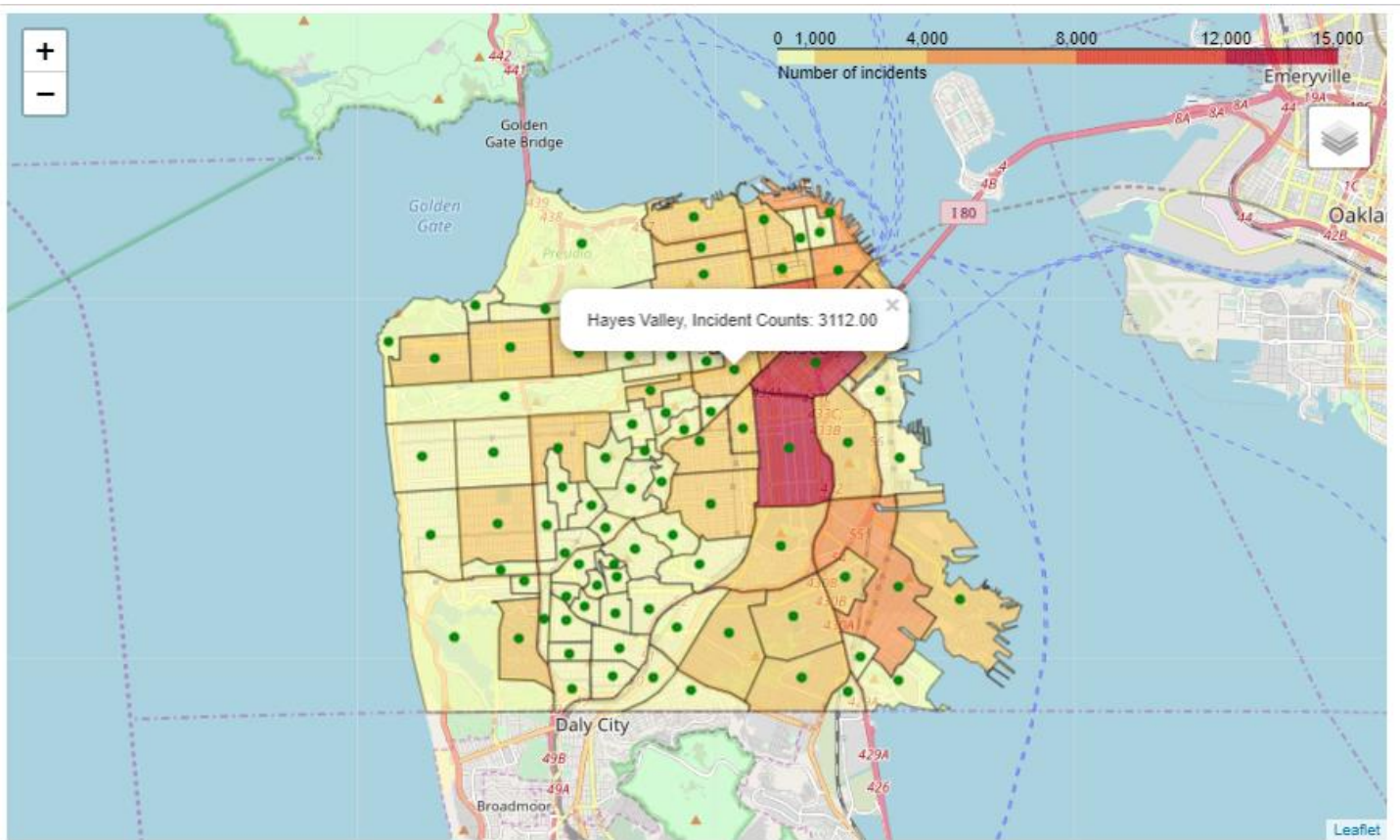
Import the neighborhoods from:

https://data.sfgov.org/Geographic-Locations-and-Boundaries/Realtor-Neighborhoods/5gzd-g9ns/data

 # San Francisco Neighborhoods

 # Calculate the number of incidents in each neighborhood

	nbrhood	incident_counts
0	Alamo Square	941
1	Anza Vista	403
2	Balboa Terrace	65
3	Bayview	4604
4	Bayview Heights	371

Generate a Choreoplath Map using Folium.

Hayes Value is a neighborhood that has a lot of cafes, bars and restaurants which are complementary venues to the gallery scene.

3.2 Accessibility analysis

Load and explore the San Francisco MUNI Transit Stops, MTA.meteredstreetblocks, MTA.motoparking_unmetered, MTA.parkingcensus_offstreet
 # make a comparative analysis of the different maps.

3.3 Visibility analysis

Using Foursquare to extract the most popular venues in the city we will establish the visibility factor that different neighborhoods can provide.

Get the 100 most popular venues at a 500m radius for all San Francisco neighborhoods.

0 Neighborhood: Alamo Square   , number of venues found: 71
1 Neighborhood: Anza Vista   , number of venues found: 18
2 Neighborhood: Balboa Terrace   , number of venues found: 17

 3 Neighborhood: Bayview   , number of venues found: 16
 4 Neighborhood: Bernal Heights   , number of venues found: 39
 5 Neighborhood: Buena Vista/Ashbury Heights   , number of venues found: 54
 6 Neighborhood: Central Richmond   , number of venues found: 99
 7 Neighborhood: Central Sunset   , number of venues found: 13
 8 Neighborhood: Clarendon Heights   , number of venues found: 12
 9 Neighborhood: Corona Heights   , number of venues found: 57
10 Neighborhood: Cow Hollow   , number of venues found: 100
11 Neighborhood: Crocker Amazon   , number of venues found: 20
12 Neighborhood: Diamond Heights   , number of venues found: 13
13 Neighborhood: Downtown   , number of venues found: 100
14 Neighborhood: Duboce Triangle   , number of venues found: 88
15 Neighborhood: Eureka Valley / Dolores Height   , number of venues found: 100
16 Neighborhood: Excelsior   , number of venues found: 4
17 Neighborhood: Financial District/Barbary Coast   , number of venues found: 100
18 Neighborhood: Yerba Buena   , number of venues found: 100
19 Neighborhood: Forest Hill   , number of venues found: 6
20 Neighborhood: Forest Hills Extension   , number of venues found: 11
21 Neighborhood: Forest Knolls   , number of venues found: 11
22 Neighborhood: Glen Park   , number of venues found: 49
23 Neighborhood: Golden Gate Heights   , number of venues found: 14
24 Neighborhood: Golden Gate Park   , number of venues found: 6
25 Neighborhood: Haight Ashbury   , number of venues found: 100
26 Neighborhood: Hayes Valley   , number of venues found: 100
27 Neighborhood: Hunters Point   , number of venues found: 5
28 Neighborhood: Ingleside   , number of venues found: 24
29 Neighborhood: Ingleside Heights   , number of venues found: 11
30 Neighborhood: Ingleside Terrace   , number of venues found: 13
31 Neighborhood: Inner Mission   , number of venues found: 81
32 Neighborhood: Inner Parkside   , number of venues found: 27
33 Neighborhood: Inner Richmond   , number of venues found: 49
34 Neighborhood: Inner Sunset   , number of venues found: 16
35 Neighborhood: Jordan Park / Laurel Heights   , number of venues found: 34
36 Neighborhood: Lake Street   , number of venues found: 18
37 Neighborhood: Monterey Heights   , number of venues found: 4
38 Neighborhood: Lake Shore   , number of venues found: 6
39 Neighborhood: Lakeside   , number of venues found: 54
40 Neighborhood: Lone Mountain   , number of venues found: 13
41 Neighborhood: Lower Pacific Heights   , number of venues found: 100
42 Neighborhood: Marina   , number of venues found: 79

43 Neighborhood: Merced Heights , number of venues found: 5
44 Neighborhood: Merced Manor , number of venues found: 14
45 Neighborhood: Midtown Terrace , number of venues found: 4
46 Neighborhood: South Beach , number of venues found: 56
47 Neighborhood: Miraloma Park , number of venues found: 7
48 Neighborhood: Mission Bay , number of venues found: 58
49 Neighborhood: Mission Dolores , number of venues found: 100
50 Neighborhood: Mission Terrace , number of venues found: 23
51 Neighborhood: Mount Davidson Manor , number of venues found: 18
52 Neighborhood: Noe Valley , number of venues found: 59
53 Neighborhood: North Beach , number of venues found: 100
54 Neighborhood: North Panhandle , number of venues found: 47
55 Neighborhood: North Waterfront , number of venues found: 40
56 Neighborhood: Oceanview , number of venues found: 8
57 Neighborhood: Outer Mission , number of venues found: 18
58 Neighborhood: Outer Parkside , number of venues found: 22
59 Neighborhood: Outer Richmond , number of venues found: 43
60 Neighborhood: Outer Sunset , number of venues found: 32
61 Neighborhood: Pacific Heights , number of venues found: 61
62 Neighborhood: Parkside , number of venues found: 35
63 Neighborhood: Cole Valley/Parnassus Heights , number of venues found: 25
64 Neighborhood: Pine Lake Park , number of venues found: 9
65 Neighborhood: Portola , number of venues found: 5
66 Neighborhood: Potrero Hill , number of venues found: 42
67 Neighborhood: Presidio , number of venues found: 14
68 Neighborhood: Presidio Heights , number of venues found: 26
69 Neighborhood: Russian Hill , number of venues found: 70
70 Neighborhood: Saint Francis Wood , number of venues found: 54
71 Neighborhood: Sea Cliff , number of venues found: 7
72 Neighborhood: Silver Terrace , number of venues found: 5
73 Neighborhood: South of Market , number of venues found: 88
74 Neighborhood: Stonestown , number of venues found: 15
75 Neighborhood: Sunnyside , number of venues found: 15
76 Neighborhood: Telegraph Hill , number of venues found: 100
77 Neighborhood: Twin Peaks , number of venues found: 7
78 Neighborhood: Van Ness/Civic Center , number of venues found: 71
79 Neighborhood: Visitacion Valley , number of venues found: 6
80 Neighborhood: West Portal , number of venues found: 56
81 Neighborhood: Western Addition , number of venues found: 63
82 Neighborhood: Westwood Highlands , number of venues found: 9

83 Neighborhood: Westwood Park   , number of venues found: 28

84 Neighborhood: Lincoln Park   , number of venues found: 22

85 Neighborhood: Sherwood Forest   , number of venues found: 4

86 Neighborhood: Tenderloin   , number of venues found: 100

87 Neighborhood: Central Waterfront/Dogpatch   , number of venues found: 55

88 Neighborhood: Candlestick Point   , number of venues found: 8

89 Neighborhood: Bayview Heights   , number of venues found: 4

90 Neighborhood: Little Hollywood   , number of venues found: 12

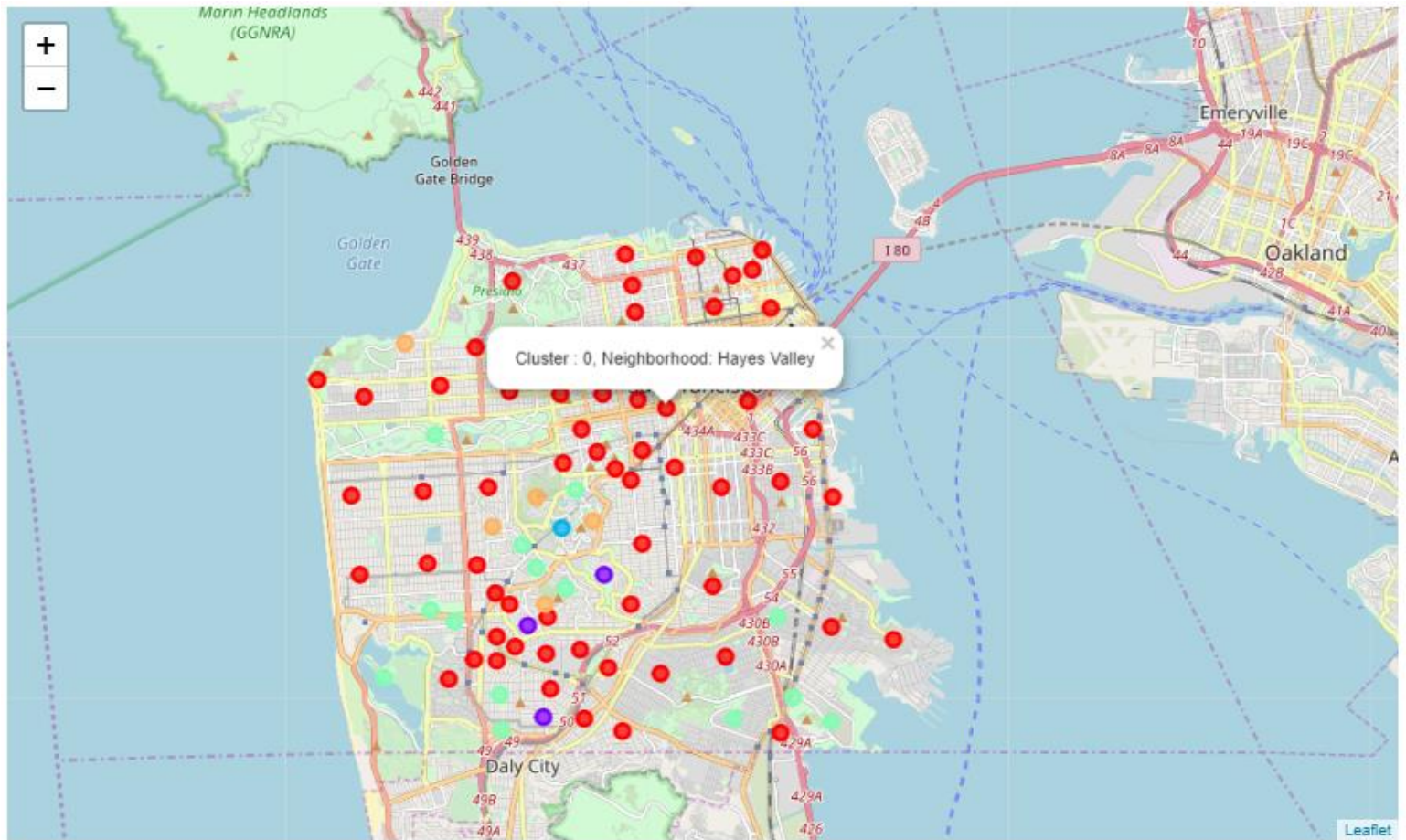91 Neighborhood: Nob Hill   , number of venues found: 46

The search has returned **3568 venues in total** and there are **344 uniques categories**.

Cluster the neighborhoods using the k-means algorithm.

Create a dataframe that contains venues in descending order for each zip code area.

	nbrhood	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	Bar	Record Shop	Gift Shop	BBQ Joint	Ethiopian Restaurant	Italian Restaurant	Mediterranean Restaurant	Sushi Restaurant	Cocktail Bar	Pizza Place
1	Anza Vista	Café	Coffee Shop	Health & Beauty Service	Burger Joint	Cosmetics Shop	Mexican Restaurant	Grocery Store	Tunnel	Sandwich Place	Big Box Store
2	Balboa Terrace	Light Rail Station	Japanese Restaurant	Vietnamese Restaurant	Gym	Thai Restaurant	Pharmacy	Bakery	Circus	Dessert Shop	Playground
3	Bayview	Southern / Soul Food Restaurant	BBQ Joint	Gym	Café	Park	Food	Plaza	Bakery	Theater	African Restaurant
4	Bayview Heights	Breakfast Spot	Park	Burger Joint	Light Rail Station	Food	Exhibit	Farmers Market	Fast Food Restaurant	Filipino Restaurant	Flower Shop

Use k-means to cluster the results into 5 clusters and generate the San Francisco neighborhood clusters map.

There are some galleries nearby but not too close to Fell Street (which is the central street in Hayes Neighborhood).
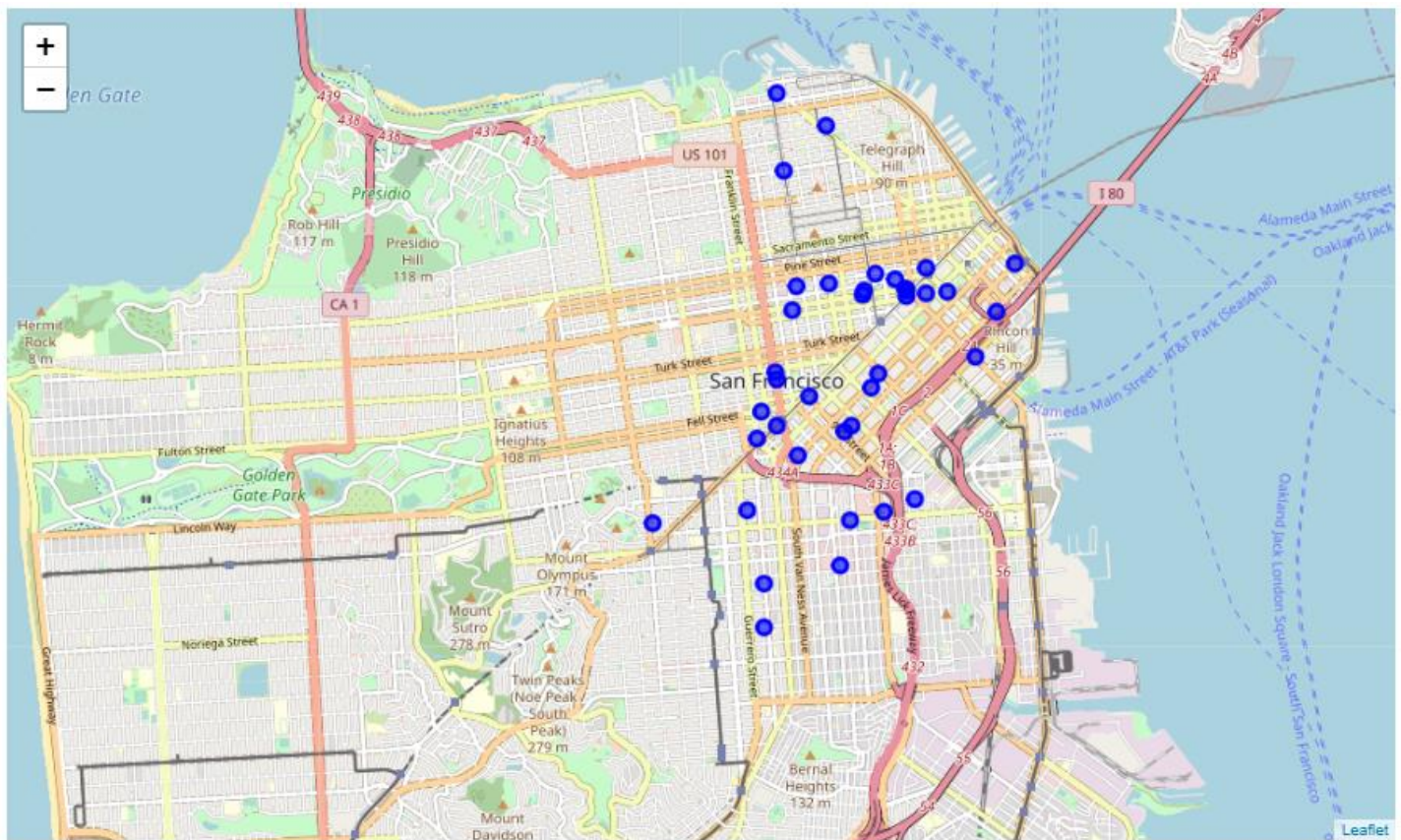
3.4 Competitor analysis

Foursquare analysis will be also used to avoid competition while simultaneously benefit from proximity with complimentary venues such as cafes and restaurants.
Use the Foursquare API to get all the art galleries in San Francisco.

	name	categories	address	cc	city	country	crossStreet	distance	formattedAddress	labeledLatLngs	lat	lng	postalCode	s
0	SFAC Main Gallery	Art Gallery	401 Van Ness Ave Ste 126	US	San Francisco	United States	at McAllister St	97	[401 Van Ness Ave Ste 126 (at McAllister St), ...	[{'label': 'display', 'lat': 37.77976899820566...	37.779769	-122.420157	94102	
1	111 Minna Gallery	Art Gallery	111 Minna St	US	San Francisco	United States	at 2nd St	1979	[111 Minna St (at 2nd St), San Francisco, CA 9...	[{'label': 'display', 'lat': 37.78748829404249...	37.787488	-122.399274	94105	
2	1:AM Gallery	Art Gallery	1000 Howard St	US	San Francisco	United States	at 6th St	1024	[1000 Howard St (at 6th St), San Francisco, CA...	[{'label': 'display', 'lat': 37.779704, 'lng':...	37.779704	-122.407605	94103	
3	Martin Lawrence Gallery	Art Gallery	366 Geary St	US	San Francisco	United States	btwn Mason & Powell St.	1239	[366 Geary St (btwn Mason & Powell St.), San F...	[{'label': 'display', 'lat': 37.78728755511516...	37.787288	-122.409449	94102	
4	Terra Gallery & Event Space	Art Gallery	511 Harrison St	US	San Francisco	United States	at 1st	2396	[511 Harrison St (at 1st), San Francisco, CA 9...	[{'label': 'display', 'lat': 37.78569906230929...	37.785699	-122.393240	94105	

Visualize the galleries on the map.

There are some galleries nearby but not too close to Fell Street (which is the central street in Hayes Neighborhood).

We can further explore the places with the highest foot traffic in the city (trending places) using Foursquare API.

3.5 Further analysis

Information regarding spaces sizes and rending costs can be gathered from real estate agencies and accommodation search engines in order to perform affordability analysis.

Since the existing gallery location is proved to have contributed to a successful business model in New York, we can critically compare the data for the two cities, in order to support our final location proposal.

4. Results

The most common crime according to the safety analysis is larceny theft, with miscallenous and assault following relatively far behind.

The neighborhoods with the most incidents (more than 15,000) are South of Market and Inner Mission.

Mission Dolores, Hayes Valley and Western Addition are central and have a relatively low number of incidents (between 1,000 and 4,000).

:

Hayes Value is a neighborhood that has a lot of cafes, bars and restaurants which are complementary venues to the gallery scene.

There are some galleries nearby but not too close to Fell Street (which is the central street in Hayes Neighborhood).

The proposed location for the new gallery would be **Fell Street in Hayes Neighborhood**.