Capstone Project - The Battle of Neighborhoods

Yasser Bigdeli
San Francisco, Ca

Location recommendation for a new vegetarian restaurant opening in San Francisco, California

Contents

1 Business Problem and Background

Restaurants are one of the most important business in the world. However, many restaurants will face a bankruptcy challenge during the first few years of opening. Most of the time, the newly opened restaurants are not able to compete with very successful ones and, so that, can't survive.

The main factor of having a restaurant business successful is the location of the business. It is more likely to fail if your business is very close to the old strong competitors. Thus, finding a right location for a business is very important. The aim of this article is to investigate the San Francisco neighborhoods for the density of the available restaurants at each region, and based on that knowledge, this article will recommend a location for opening a new restaurant for a businessman looking to open a new restaurant (vegetarian) in the area. In the San Francisco city, it is even more important to have a right restaurant in a right place to have a successful business, because the cost in the San Francisco is very high for a new business opening like a vegetarian restaurant.

The business plan is to open a new vegetarian restaurant in the City of San Francisco, California and the following are the very important questions to be answered properly before starting the business:

- The locations of vegetarian restaurants in san Francisco.
- Comparing vegetarian places to the total number of restaurants.
- Where is the densest area for restaurants?
- Where is the densest neighborhood for vegetarian restaurants?
- What is the rating of the Vegetarian restaurants?
- What is the percentages of vegetarian population in the area? (if possible)

2 Data Source and Description

San Francisco neighborhood data is obtained via web scraping (https://localwiki.org/sf/Neighborhoods). The geographical longitude and latitudes of 51 San Francisco neighborhoods are obtained using ‘geopy.geocoders’ python library. The location data is obtained from ‘foursquare.com’ dataset and is based on the geographical coordinates of the neighborhoods. The foursquare API is used for obtaining the venues and restaurants locations. These neighborhoods are shown on a map in the following figure.

Table 1 San Francisco neighborhoods

	Neighborhood	latitude	longitude
0	Alamo Square	37.776360	-122.434689
1	Bayview	40.772627	-124.183950
2	Bernal Heights	37.741001	-122.414214
3	Buena Vista	37.806532	-122.420648
5	Chinatown	37.794301	-122.406376
6	Civic Center	37.779026	-122.419906
7	Cole Valley	37.765813	-122.449962
8	Cow Hollow	37.797262	-122.436248
9	Crocker Amazon	37.709378	-122.438587
11	Dogpatch	37.760698	-122.389202

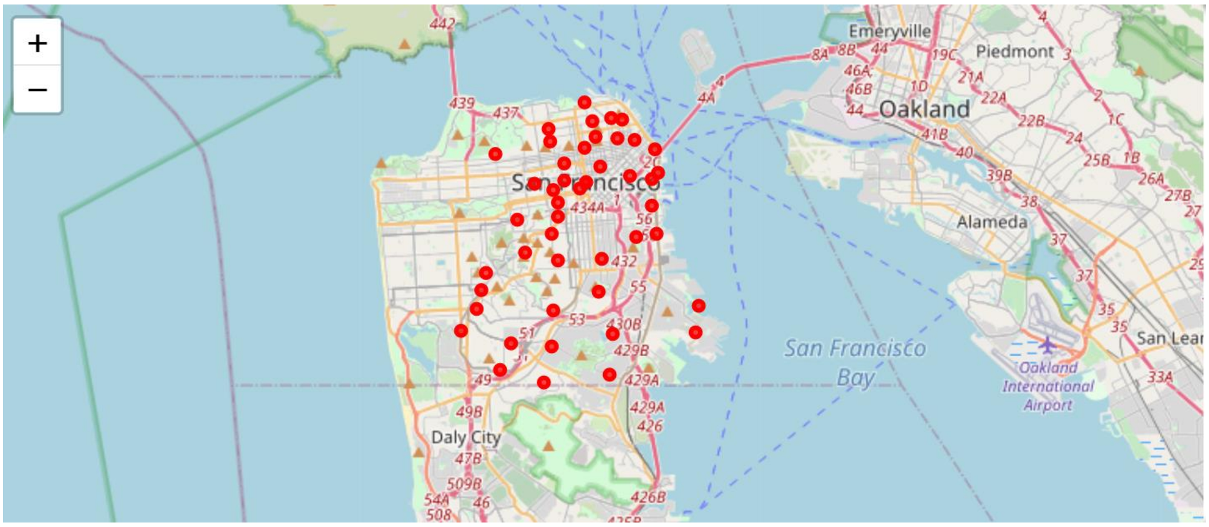


Figure 1 . Map for San Francisco neighborhoods

3 Methodology

The following tasks need to be performed:

1. Find the neighborhoods
2. Obtain the geographical coordinates
3. Get location data from foursquare API
4. Clean the data and create data frame
5. Explore the neighborhoods
6. Identify the number of all types of restaurants locations

7. Identify the number of vegetarian restaurants locations
8. Analyze each neighborhood per restaurants
9. Make a data frame for vegetarian restaurant
10. Group by neighborhood and restaurant types
11. Clustering Neighborhoods using K-means clustering per resturants
12. Visualization

4 Exploring the neighborhoods

In this section each neighborhood is explored and the venues at the 500m range are identified.

Table 2 Data set for all venues in SF neighbors

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Alamo Square	37.77636	-122.434689	Alamo Square	37.776045	-122.434363	Park
1	Alamo Square	37.77636	-122.434689	Alamo Square Dog Park	37.775878	-122.435740	Dog Run
2	Alamo Square	37.77636	-122.434689	Painted Ladies	37.776120	-122.433389	Historic Site
3	Alamo Square	37.77636	-122.434689	The Independent	37.775573	-122.437835	Rock Club
4	Alamo Square	37.77636	-122.434689	The Mill	37.776425	-122.437970	Bakery
5	Alamo Square	37.77636	-122.434689	Bar Crudo	37.775707	-122.438019	Seafood Restaurant
6	Alamo Square	37.77636	-122.434689	Fool's Errand	37.775512	-122.437961	Bar
7	Alamo Square	37.77636	-122.434689	Nopa	37.774888	-122.437532	New American Restaurant
8	Alamo Square	37.77636	-122.434689	Rare Device	37.775052	-122.437762	Gift Shop

The resultants locations are extracted from the venue data set and the bar graph is developed showing the number of total restaurants at each neighborhood.

Table 3 All type of Restaurants in SF neighborhoods

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
5	Alamo Square	37.77636	-122.434689	Bar Crudo	37.775707	-122.438019	Seafood Restaurant
7	Alamo Square	37.77636	-122.434689	Nopa	37.774888	-122.437532	New American Restaurant
19	Alamo Square	37.77636	-122.434689	jū-ni	37.776743	-122.438770	Sushi Restaurant
21	Alamo Square	37.77636	-122.434689	Tsunami	37.776869	-122.438486	Sushi Restaurant
27	Alamo Square	37.77636	-122.434689	Brenda's Meat & Three	37.778265	-122.438584	Southern / Soul Food Restaurant
30	Alamo Square	37.77636	-122.434689	Zaytoon	37.775185	-122.437896	Mediterranean Restaurant
32	Alamo Square	37.77636	-122.434689	Kung Food	37.777778	-122.438698	Hunan Restaurant
34	Alamo Square	37.77636	-122.434689	Namu Stonepot	37.774763	-122.437780	Korean Restaurant
39	Alamo Square	37.77636	-122.434689	Che Fico	37.777435	-122.438149	Italian Restaurant
41	Alamo Square	37.77636	-122.434689	Saffron Grill	37.776848	-122.437816	Indian Restaurant

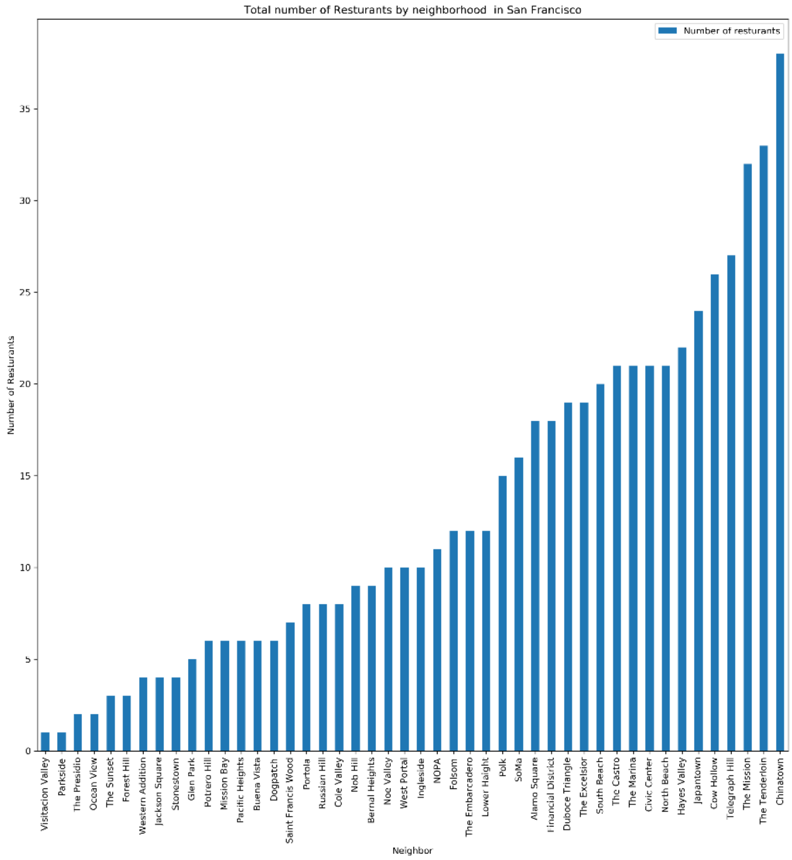


Figure 2 Number of restaurants in each Neighborhood

The vegetarian / Vegan restaurants were extracted out the dataset. The location of them are shown on san Francisco map with green points if Figure 4.

Table 4 Vegetarian restaurants in Sf neighborhoods

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
208	Chinatown	37.794301	-122.406376	Enjoy Vegetarian Restaurant	37.795833	-122.405093	Vegetarian / Vegan Restaurant
258	Chinatown	37.794301	-122.406376	Lucky Creation Restaurant	37.795056	-122.407271	Vegetarian / Vegan Restaurant
299	Civic Center	37.779026	-122.419906	Ananda Fuara	37.777693	-122.416353	Vegetarian / Vegan Restaurant
327	Civic Center	37.779026	-122.419906	Golden Era Vegan	37.781495	-122.416822	Vegetarian / Vegan Restaurant
354	Civic Center	37.779026	-122.419906	Thai Idea Vegetarian Restaurant	37.783395	-122.419141	Vegetarian / Vegan Restaurant
455	Cow Hollow	37.797262	-122.436248	Wildseed	37.797626	-122.432440	Vegetarian / Vegan Restaurant
487	Cow Hollow	37.797262	-122.436248	Vegan Picnic	37.797490	-122.431748	Vegetarian / Vegan Restaurant
1219	Nob Hill	37.794479	-122.415592	Nourish Cafe	37.790529	-122.417296	Vegetarian / Vegan Restaurant
2264	The Marina	37.779026	-122.419906	Ananda Fuara	37.777693	-122.416353	Vegetarian / Vegan Restaurant
2292	The Marina	37.779026	-122.419906	Golden Era Vegan	37.781495	-122.416822	Vegetarian / Vegan Restaurant

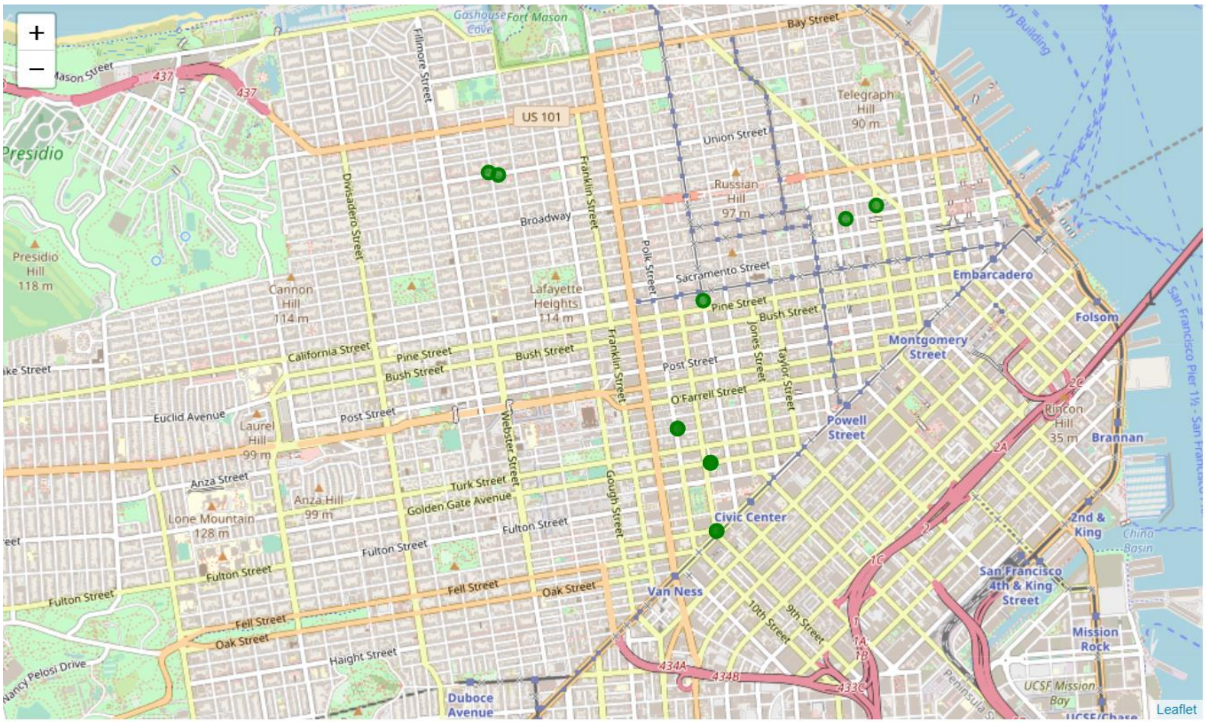


Figure 3 vegetarian restaurant locations

The neighborhood where at least one vegetarian/ vegan restaurant is located in are compared to the total number of restaurants located at that specific neighborhood.

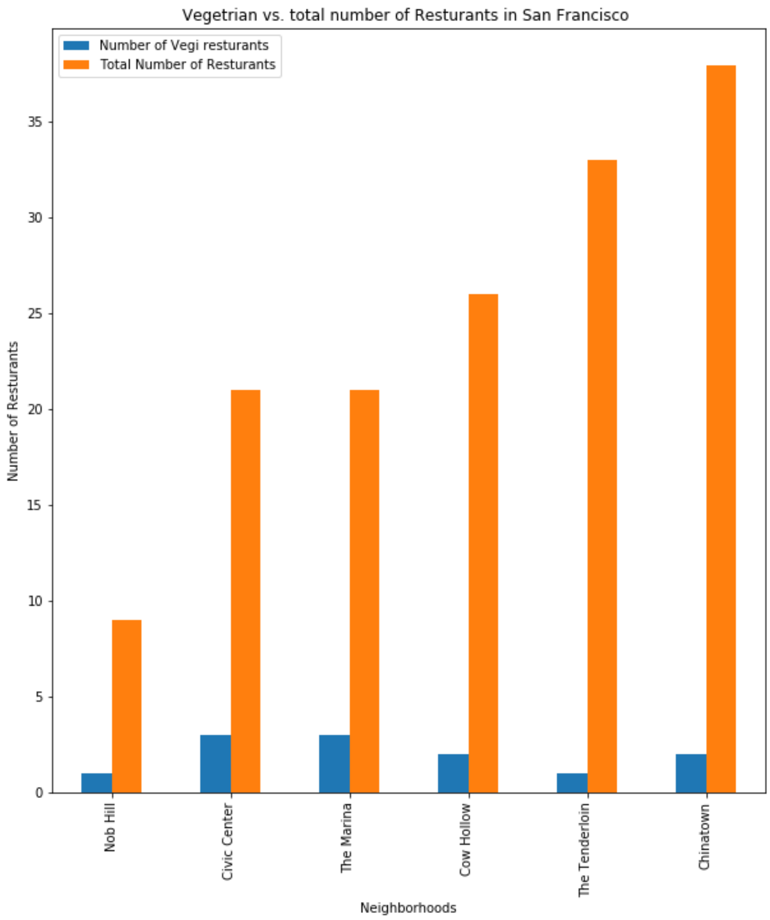


Figure 4 Number of Vegetarian Restaurants in SF Neighborhoods

5 Neighborhood clustering using K-Mean (K=5)

The 10 most common restaurants were ranged for each neighborhood as listed in the following table. The data set is used for clustering neighborhood using K-mean clustering into 5 categories. The class of clustering in also added to the data frame. The clustering results are shown on San Francisco map as shown in figure 5.

Table 10 Most common restaurants and identified clustering per neighborhood data frame

Neighborhood	latitude	longitude	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
Alamo Square	37.776360	-122.434689	1.0	Indian Restaurant	Seafood Restaurant	Ethiopian Restaurant	Sushi Restaurant	Hunan Restaurant	Mediterranean Restaurant	Mexican Restaurant	New American Restaurant
Bernal Heights	37.741001	-122.414214	2.0	Italian Restaurant	American Restaurant	Asian Restaurant	Mexican Restaurant	Indian Restaurant	Caribbean Restaurant	New American Restaurant	Peruvian Restaurant
Buena Vista	37.806532	-122.420648	2.0	Seafood Restaurant	Brazilian Restaurant	Fast Food Restaurant	Cantonese Restaurant	Korean Restaurant	Austrian Restaurant	Filipino Restaurant	Japanese Restaurant
Chinatown	37.794301	-122.406376	1.0	Chinese Restaurant	New American Restaurant	Italian Restaurant	Vietnamese Restaurant	Restaurant	Dim Sum Restaurant	Sushi Restaurant	Szechuan Restaurant
Civic Center	37.779026	-122.419906	1.0	Vegetarian / Vegan Restaurant	French Restaurant	Sushi Restaurant	Restaurant	Vietnamese Restaurant	Southern / Soul Food Restaurant	Mediterranean Restaurant	Mexican Restaurant
Cole Valley	37.765813	-122.449962	2.0	Vietnamese Restaurant	Thai Restaurant	Caribbean Restaurant	Mexican Restaurant	Tapas Restaurant	Middle Eastern Restaurant	Mediterranean Restaurant	Indian Restaurant
Cow Hollow	37.797262	-122.436248	1.0	Italian Restaurant	Mexican Restaurant	French Restaurant	American Restaurant	Vegetarian / Vegan Restaurant	Thai Restaurant	Sushi Restaurant	Caribbean Restaurant
Dogpatch	37.760698	-122.389202	1.0	Restaurant	Latin American Restaurant	Italian Restaurant	Southern / Soul Food Restaurant	Sushi Restaurant	Vietnamese Restaurant	Greek Restaurant	Filipino Restaurant
Duboce Triangle	37.767138	-122.432230	2.0	New American Restaurant	Seafood Restaurant	Mexican Restaurant	Sushi Restaurant	Vietnamese Restaurant	Indian Restaurant	Mediterranean Restaurant	Middle Eastern Restaurant
Financial District	37.793647	-122.398938	2.0	Restaurant	Japanese Restaurant	Mediterranean Restaurant	New American Restaurant	Dim Sum Restaurant	Latin American Restaurant	Mexican Restaurant	Middle Eastern Restaurant

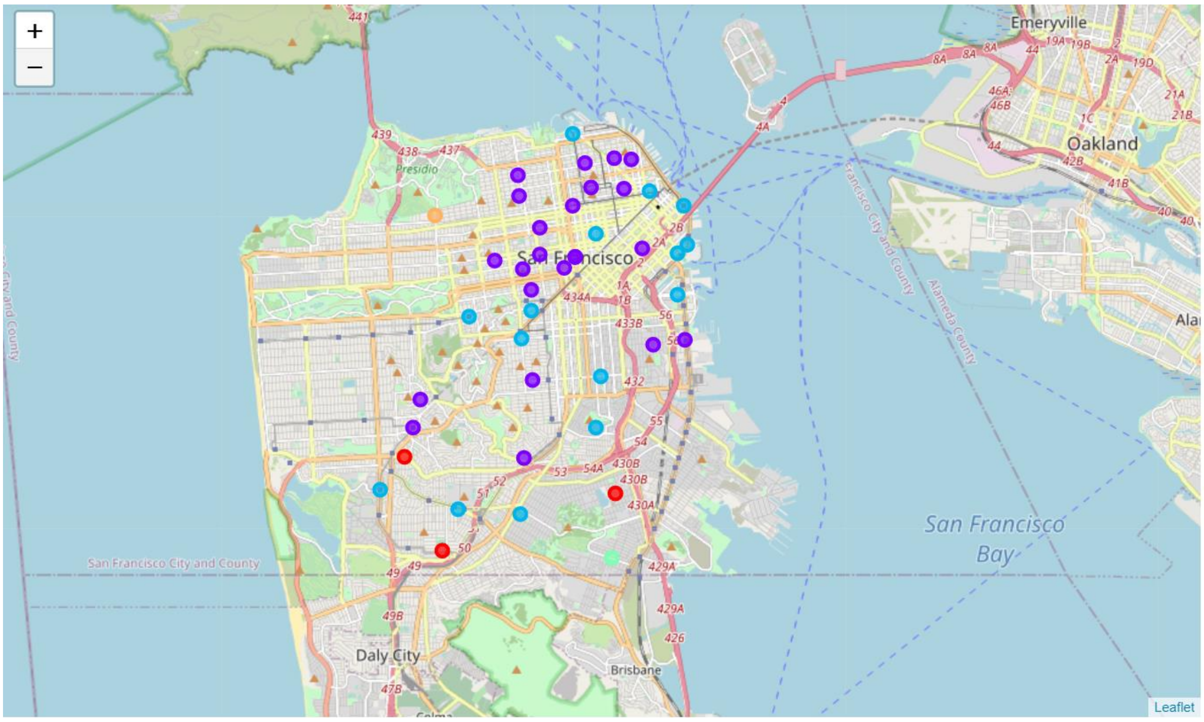


Figure 5 Class of clusters location on map (similar colors represent similar clusters)

To see more details on clustering please see the Appendix A.

6 Results and Discussion

From analyzing data, it is noticed that only 7 neighborhoods have vegetarian restaurants available. Therefore, opening a new vegetarian restaurant, in general, would be beneficial. It is argued that, the best neighborhood to open vegetarian place would be the neighborhood that has many restaurants types except vegetarian. In this case we recommend “The Mission” neighborhood. There are more factors to consider to get better recommendations, however, it is out of scope of this article.

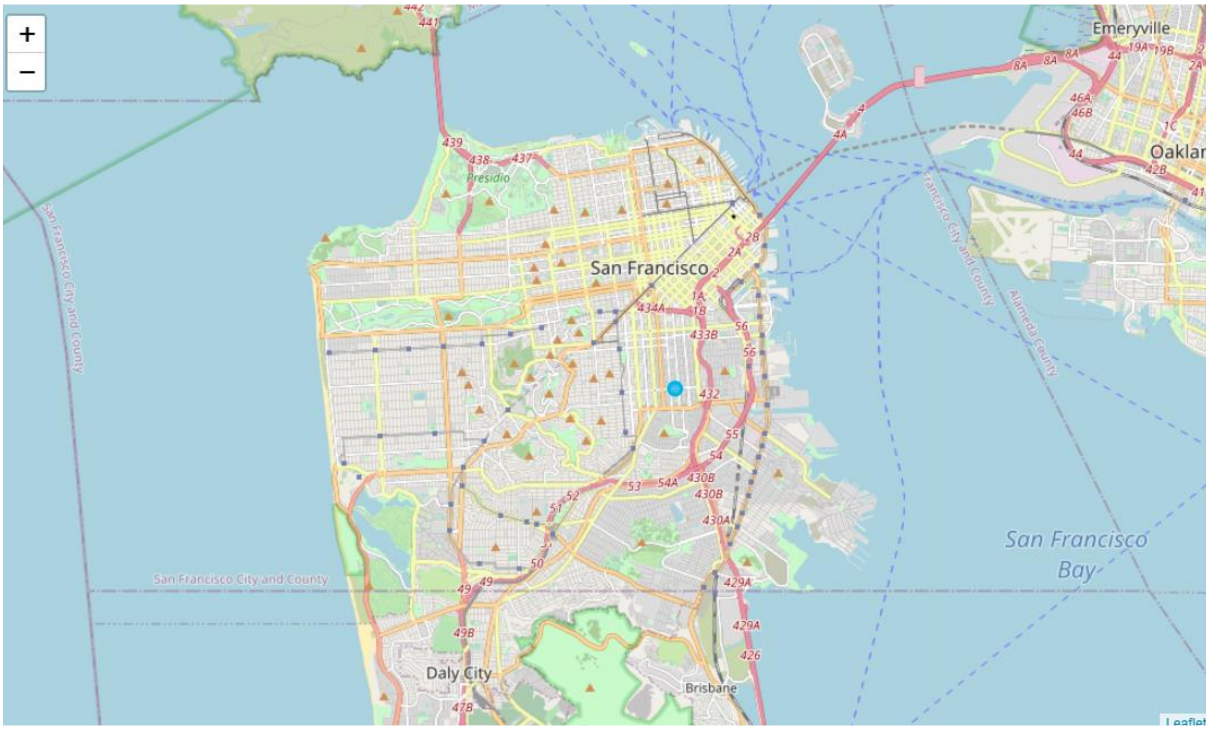


Figure 6 The recommended location (The Mission) for opening a Vegetarian restaurant

7 Conclusion

For a businessman, the most important factor is choosing a right place to start the business. This is more crucial for restaurant business. This article tries to recommend a good location at the city of San Francisco, California for a vegetarian eating place.
After analyzing all the neighborhoods and exploring the available restaurant and, in specific, vegetarian / vegan   eating places, the ‘The Mission’ neighborhood is recommended for opening new restaurant for vegetarians. Based on location data analyses, this neighborhood is the dense area of eating places, where, however, there is no single vegetarian restaurant in this region. Therefore, makes it a perfect place for opening a new restaurant for vegetarians. The neighborhoods are also classified based on the frequency of restaurants available in the area using K-Means clustering methodology. This could be used for further analyses which is out of this articles scope.

8 Future works

This project could be well expanded for many other types of business and sectors. There are a few tides that can be taken into exploration:

1. Where to open a fast-food restaurant
2. Where to open coffee shop
3. Where a police station might be needed to open
4. To spot the locations with higher crime rates

Appendix A:
The PDf of the analysis is appended. To have access to the Jupyter notebook version (.ipynb format) of this article please see the following link:

https://github.com/yasser64b/Coursera_Capstone/blob/master/Coursera%20Capstone%20Project_Final.ipynb

Capstone Project - The Battle of Neighborhoods

Yasser Bigdeli San Francisco, Ca

Location recommendation for a new vegetarian restaurant opening in San Francisco, California

1. Business Problem and Background:

Restaurants are one of the most important business in the world. However, many restaurants will face a bankruptcy challenge during the first few years of opening. Most of the time, the newly opened restaurants are not able to compete with very successful ones and, so that, can't survive. The main factor of having a restaurant business successful is the location of the business. It is more likely to fail if your business is very close to the old strong competitors. Thus, finding a right location for a business is very important. The aim of this article is to investigate the San Francisco neighborhoods for the density of the available restaurants at each region, and based on that knowledge, this article will recommend a location for opening a new restaurant for a businessman looking to open a new restaurant (vegetarian) in the area. In the San Francisco city, it is even more important to have a right restaurant in a right place to have a successful business, because the cost in the San Francisco is very high for a new business opening like a vegetarian restaurant. The business plan is to open a new vegetarian restaurant in the City of San Francisco, California and the following are the very important questions to be answered properly before starting the business:

- The locations of vegetarian restaurants in san Francisco.
- Comparing vegetarian places to the total number of restaurants.
- Where is the densest area for restaurants?
- Where is the densest neighborhood for vegetarian restaurants?
- What is the rating of the Vegetarian restaurants?
- What is the percentages of vegetarian population in the area? (if possible

2. Data Source and Description:

San Francisco neighborhood data is obtained via web scraping (https://localwiki.org/sf/Neighborhoods (https://localwiki.org/sf/Neighborhoods)). The geographical longitude and latitudes of each neighborhoods is obtained using 'geopy.geocoders' python library. The location data is obtained from 'foursquare.com' dataset and is based on the geographical coordinates of the neighborhoods. The foursquare API is used for obtaining the venues, user, rating, tips, etc.

3. Mthodology

The following task need to be perfomed:

1. Find the neighborhoods

2. Obtain the geographical coordinates
3. Get location data from foursquare API
4. Clean the data and create data frame
5. Explore the neighborhoods
6. Analyze each neighborhood
7. Make a data frame for vegetarian restaurant
8. Group by neighborhood
9. Clustering Neighborhoods using K-means clustering
10. Visualizing the resulting clusters
11. Making recommendation for a vegetarian restaurant location

Importing Libraries

In [1]:
```
!pip install geopy
```

```
Collecting geopy
  Downloading https://files.pythonhosted.org/packages/07/e1/9c72de674d5c2b8fcb0
738a5ceeb5424941fefa080bfe4e240d0bacb5a38/geopy-2.0.0-py3-none-any.whl (http
s://files.pythonhosted.org/packages/07/e1/9c72de674d5c2b8fcb0738a5ceeb5424941fe
fa080bfe4e240d0bacb5a38/geopy-2.0.0-py3-none-any.whl) (111kB)
Collecting geographiclib<2,>=1.49 (from geopy)
  Downloading https://files.pythonhosted.org/packages/8b/62/26ec95a98ba64299163
199e95ad1b0e34ad3f4e176e221c40245f211e425/geographiclib-1.50-py3-none-any.whl
 (https://files.pythonhosted.org/packages/8b/62/26ec95a98ba64299163199e95ad1b0e
34ad3f4e176e221c40245f211e425/geographiclib-1.50-py3-none-any.whl)
Installing collected packages: geographiclib, geopy
Successfully installed geographiclib-1.50 geopy-2.0.0
```

In [2]:
```
import pandas as pd
import numpy as np
pd.set_option('display.max_columns', None)
pd.set_option('display.max_rows', None)
from geopy.geocoders import Nominatim
from pandas.io.json import json_normalize
from sklearn.cluster import KMeans
import folium
import json
import matplotlib.cm as cm
import matplotlib.colors as colors
import requests
import matplotlib.pyplot as plt
import seaborn as sns
```

List of San Farancisco neighborhood: source: https://localwiki.org/sf/Neighborhoods (https://localwiki.org/sf/Neighborhoods)

In [3]:
```
neigh='''Alamo Square
Bayview
Bernal Heights
Buena Vista
Butcher Town
Chinatown
Civic Center
Cole Valley
Cow Hollow
Crocker Amazon
Diamond Heights
Dogpatch
Duboce Triangle
Fillmore District
Financial District
Folsom
Forest Hill
Glen Park
Golden Gate Heights
Hayes Valley
Hunters Point
India Basin
Ingleside
Jackson Square
Japantown
Lower Haight
Miraloma Park
Mission Bay
Nob Hill
Noe Valley
NOPA
North Beach
Ocean View
Pacific Heights
Park Merced
Parkside
Parnassus Heights
Polk
Portola
Potrero Hill
Rincon Hills
Russian Hill
Saint Francis Wood
SoMa
South Beach
Stonestown
Telegraph Hill
Theater District
The Castro
The Dolores Valley
The Embarcadero
The Excelsior
The Marina
The Mission
The Presidio
The Sunset
```

```
The Tenderloin
Twin Peaks
Upper Market
Visitacion Valley
Western Addition
West Portal'''
```

Data Cleaning

In [4]:
```
sf_neigh=pd.DataFrame(neigh.split('\n'), columns=['Neighborhood'])
```

In [5]:
```
sf_neigh.tail()
```

Out[5]:

	Neighborhood
57	Twin Peaks
58	Upper Market
59	Visitacion Valley
60	Western Addition
61	West Portal

In [6]:
```
sf_neigh['latitude']=np.nan
sf_neigh['longitude']=np.nan
# sf_neigh.head()
```

San Francisco Geograpgical coordinate

In [7]:
```
address = 'San Francisco City, CA, USA'

geolocator = Nominatim(user_agent="ca_explorer")
location = geolocator.geocode(address)
latitude_sf = location.latitude
longitude_sf = location.longitude
print('The geographical coordinate of San Francisco are {}, {}.'.format(latitude_s
```

```
The geograpical coordinate of San Francisco are 37.7790262, -122.4199061.
```

In [ ]:

Adding Neighborhood geographical coordinates to SF_neigh data frame

In [8]:
```
for i , neighbor in enumerate(sf_neigh['Neighborhood']):

    address = neighbor + ', ' + 'San Francisco City, CA, USA'
    try:
        geolocator = Nominatim(user_agent="California_explorer")
        location = geolocator.geocode(address, timeout=10000)
        latitude = location.latitude
        longitude = location.longitude
        sf_neigh.loc[i,'latitude']=latitude
        sf_neigh.loc[i,'longitude']=longitude
#         print(latitude, longitude)
    except:
        pass

sf_neigh.dropna(inplace=True)
```

In [11]:
```
sf_neigh.head(10)
```

Out[11]:

	Neighborhood	latitude	longitude
0	Alamo Square	37.776360	-122.434689
1	Bayview	40.772627	-124.183950
2	Bernal Heights	37.741001	-122.414214
3	Buena Vista	37.806532	-122.420648
5	Chinatown	37.794301	-122.406376
6	Civic Center	37.779026	-122.419906
7	Cole Valley	37.765813	-122.449962
8	Cow Hollow	37.797262	-122.436248
9	Crocker Amazon	37.709378	-122.438587
11	Dogpatch	37.760698	-122.389202

In [10]:
```
sf_neigh.shape
```

Out[10]:
```
(51, 3)
```

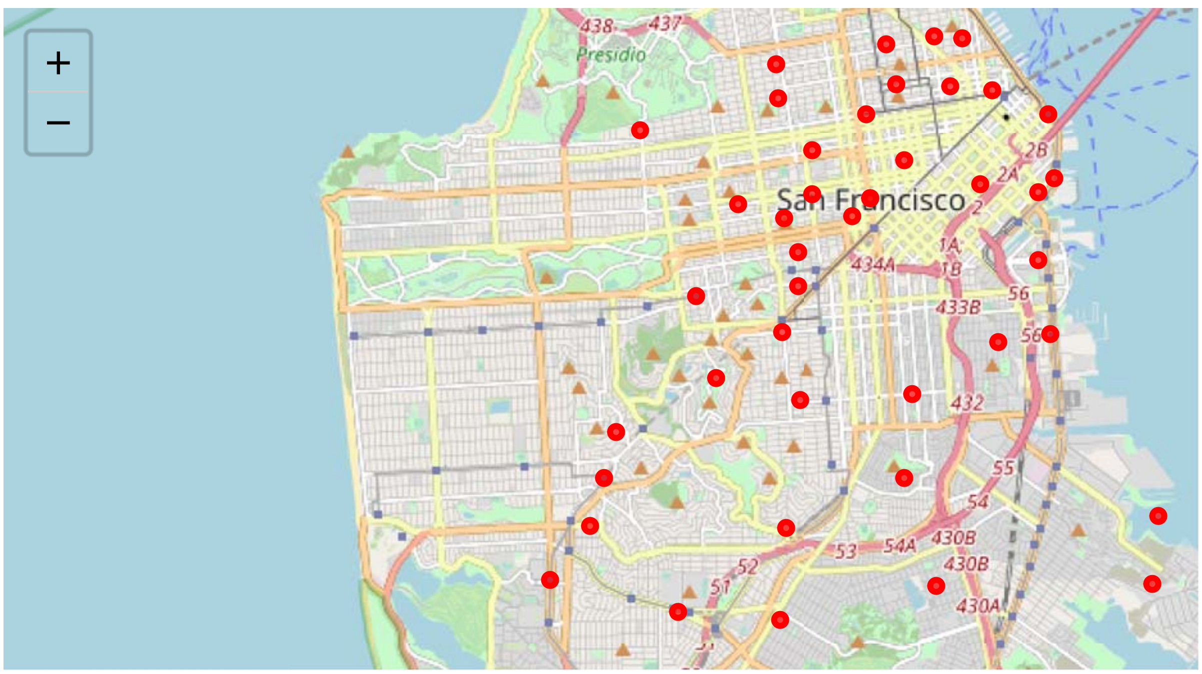
Create SF map and neighbors

In [12]:
```python
# create map of SF using latitude and longitude values
map_SF = folium.Map(location=[latitude_sf, longitude_sf], zoom_start=12)

# add markers to map
for lat, lng, neighborhood in zip(sf_neigh['latitude'], sf_neigh['longitude'], s
    label = '{}'.format(neighborhood)
    label = folium.Popup(label, parse_html=True)
    folium.CircleMarker(
        [lat, lng],
        radius=3,
        popup=label,
        color='red',
        fill=True,
#         fill_color='#3186cc',
        fill_opacity=0.7,
        parse_html=False).add_to(map_SF)

map_SF
```

Out[12]:



Defining Foursquare Credentials and Version

In [13]:
```python
CLIENT_ID = 'PLCXAA0GUXKEDINFEIBBOK2VDEROMUSYNCLBVJE35J2ARXX2' # your Foursquare
CLIENT_SECRET = 'AKXSSXH2OLJDE0DLJXRQEQSOEWHLKTDUC24FEGQIHI4W0RXW' # your Foursq
VERSION = '20180605' # Foursquare API version
LIMIT = 100 # A default Foursquare API limit value

print('Your credentails:')
print('CLIENT_ID: ' + CLIENT_ID)
print('CLIENT_SECRET:' + CLIENT_SECRET)
```

```
Your credentails:
CLIENT_ID: PLCXAA0GUXKEDINFEIBBOK2VDEROMUSYNCLBVJE35J2ARXX2
CLIENT_SECRET:AKXSSXH2OLJDE0DLJXRQEQSOEWHLKTDUC24FEGQIHI4W0RXW
```

Explore Neighborhoods in SF

In [14]:
```
def getNearbyVenues(names, latitudes, longitudes, radius=500):

    venues_list=[]
    for name, lat, lng in zip(names, latitudes, longitudes):
        print(name)

        # create the API request URL
        url = 'https://api.foursquare.com/v2/venues/explore?&client_id={}&client_
            CLIENT_ID,
            CLIENT_SECRET,
            VERSION,
            lat,
            lng,
            radius,
            LIMIT)

        # make the GET request
        results = requests.get(url).json()["response"]['groups'][0]['items']

        # return only relevant information for each nearby venue
        venues_list.append([(
            name,
            lat,
            lng,
            v['venue']['name'],
            v['venue']['location']['lat'],
            v['venue']['location']['lng'],
            v['venue']['categories'][0]['name']) for v in results])

    nearby_venues = pd.DataFrame([item for venue_list in venues_list for item in
    nearby_venues.columns = ['Neighborhood',
                  'Neighborhood Latitude',
                  'Neighborhood Longitude',
                  'Venue',
                  'Venue Latitude',
                  'Venue Longitude',
                  'Venue Category']

    return(nearby_venues)
```

In [15]:

```
SF_venues = getNearbyVenues(names=sf_neigh['Neighborhood'],
                                   latitudes=sf_neigh['latitude'],
                                   longitudes=sf_neigh['longitude']
                                  )
```

```
Alamo Square
Bayview
Bernal Heights
Buena Vista
Chinatown
Civic Center
Cole Valley
Cow Hollow
Crocker Amazon
Dogpatch
Duboce Triangle
Financial District
Folsom
Forest Hill
Glen Park
Hayes Valley
Hunters Point
India Basin
Ingleside
Jackson Square
Japantown
Lower Haight
Mission Bay
Nob Hill
Noe Valley
NOPA
North Beach
Ocean View
Pacific Heights
Parkside
Polk
Portola
Potrero Hill
Russian Hill
Saint Francis Wood
SoMa
South Beach
Stonestown
Telegraph Hill
The Castro
The Embarcadero
The Excelsior
The Marina
The Mission
The Presidio
The Sunset
The Tenderloin
Twin Peaks
Visitacion Valley
Western Addition
West Portal
```

In [16]:
```
SF_venues.shape
```

Out[16]:
```
(2636, 7)
```

In [17]:
```
SF_venues.head(10)
```

Out[17]:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Alamo Square	37.77636	-122.434689	Alamo Square	37.776045	-122.434363	Park
1	Alamo Square	37.77636	-122.434689	Alamo Square Dog Park	37.775878	-122.435740	Dog Run
2	Alamo Square	37.77636	-122.434689	Painted Ladies	37.776120	-122.433389	Historic Site
3	Alamo Square	37.77636	-122.434689	The Independent	37.775573	-122.437835	Rock Club
4	Alamo Square	37.77636	-122.434689	The Mill	37.776425	-122.437970	Bakery
5	Alamo Square	37.77636	-122.434689	Bar Crudo	37.775707	-122.438019	Seafood Restaurant
6	Alamo Square	37.77636	-122.434689	Fool's Errand	37.775512	-122.437961	Bar
7	Alamo Square	37.77636	-122.434689	Nopa	37.774888	-122.437532	New American Restaurant
8	Alamo Square	37.77636	-122.434689	Rare Device	37.775052	-122.437762	Gift Shop
9	Alamo Square	37.77636	-122.434689	4505 Burgers & BBQ	37.776125	-122.438142	BBQ Joint

In [18]:
```
SF_venues.groupby('Neighborhood').count()
```

Out[18]:

Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Alamo Square	77	77	77	77	77	77
Bernal Heights	43	43	43	43	43	43
Buena Vista	51	51	51	51	51	51
Chinatown	100	100	100	100	100	100
Civic Center	100	100	100	100	100	100
Cole Valley	49	49	49	49	49	49
Cow Hollow	100	100	100	100	100	100
Crocker Amazon	2	2	2	2	2	2
Dogpatch	56	56	56	56	56	56
Duboce Triangle	71	71	71	71	71	71
Financial District	77	77	77	77	77	77
Folsom	67	67	67	67	67	67
Forest Hill	5	5	5	5	5	5
Glen Park	30	30	30	30	30	30
Hayes Valley	100	100	100	100	100	100
Hunters Point	6	6	6	6	6	6
India Basin	2	2	2	2	2	2
Ingleside	34	34	34	34	34	34
Jackson Square	12	12	12	12	12	12
Japantown	78	78	78	78	78	78
Lower Haight	59	59	59	59	59	59
Mission Bay	62	62	62	62	62	62
NOPA	36	36	36	36	36	36
Nob Hill	43	43	43	43	43	43
Noe Valley	52	52	52	52	52	52
North Beach	93	93	93	93	93	93
Ocean View	13	13	13	13	13	13
Pacific Heights	61	61	61	61	61	61
Parkside	3	3	3	3	3	3
Polk	81	81	81	81	81	81

	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Neighborhood						
Portola	25	25	25	25	25	25
Potrero Hill	33	33	33	33	33	33
Russian Hill	45	45	45	45	45	45
Saint Francis Wood	26	26	26	26	26	26
SoMa	93	93	93	93	93	93
South Beach	100	100	100	100	100	100
Stonestown	37	37	37	37	37	37
Telegraph Hill	99	99	99	99	99	99
The Castro	98	98	98	98	98	98
The Embarcadero	71	71	71	71	71	71
The Excelsior	46	46	46	46	46	46
The Marina	100	100	100	100	100	100
The Mission	90	90	90	90	90	90
The Presidio	11	11	11	11	11	11
The Sunset	18	18	18	18	18	18
The Tenderloin	100	100	100	100	100	100
Twin Peaks	11	11	11	11	11	11
Visitacion Valley	3	3	3	3	3	3
West Portal	44	44	44	44	44	44
Western Addition	23	23	23	23	23	23

In [19]:
```
SF_venues.groupby('Venue Category').count()
```

Venue Category						
Asian Restaurant	8	8	8	8	8	8
Athletics & Sports	5	5	5	5	5	5
Austrian Restaurant	1	1	1	1	1	1
Automotive Shop	1	1	1	1	1	1
BBQ Joint	6	6	6	6	6	6
Bagel Shop	7	7	7	7	7	7
Bakery	55	55	55	55	55	55
Bank	11	11	11	11	11	11
Bar	38	38	38	38	38	38
Baseball Field	2	2	2	2	2	2
Baseball Stadium	16	16	16	16	16	16
Basketball Stadium	1	1	1	1	1	1
Bath House	1	1	1	1	1	1

number of unique categories

In [20]:
```
print('There are {} uniques categories.'.format(len(SF_venues['Venue Category'].
```

```
There are 312 uniques categories.
```

Extracting restaurants from the data set

In [21]:
```
Rest=[]
for i in range (SF_venues.shape[0]):
   Rest.append('Restaurant' in SF_venues['Venue Category'].apply(lambda x: x.spl
```

In [22]:
```
SF_rest=SF_venues[Rest]
SF_rest.shape
```

Out[22]:
```
(596, 7)
```

In [50]:
```
SF_rest.head(10)
```

Out[50]:

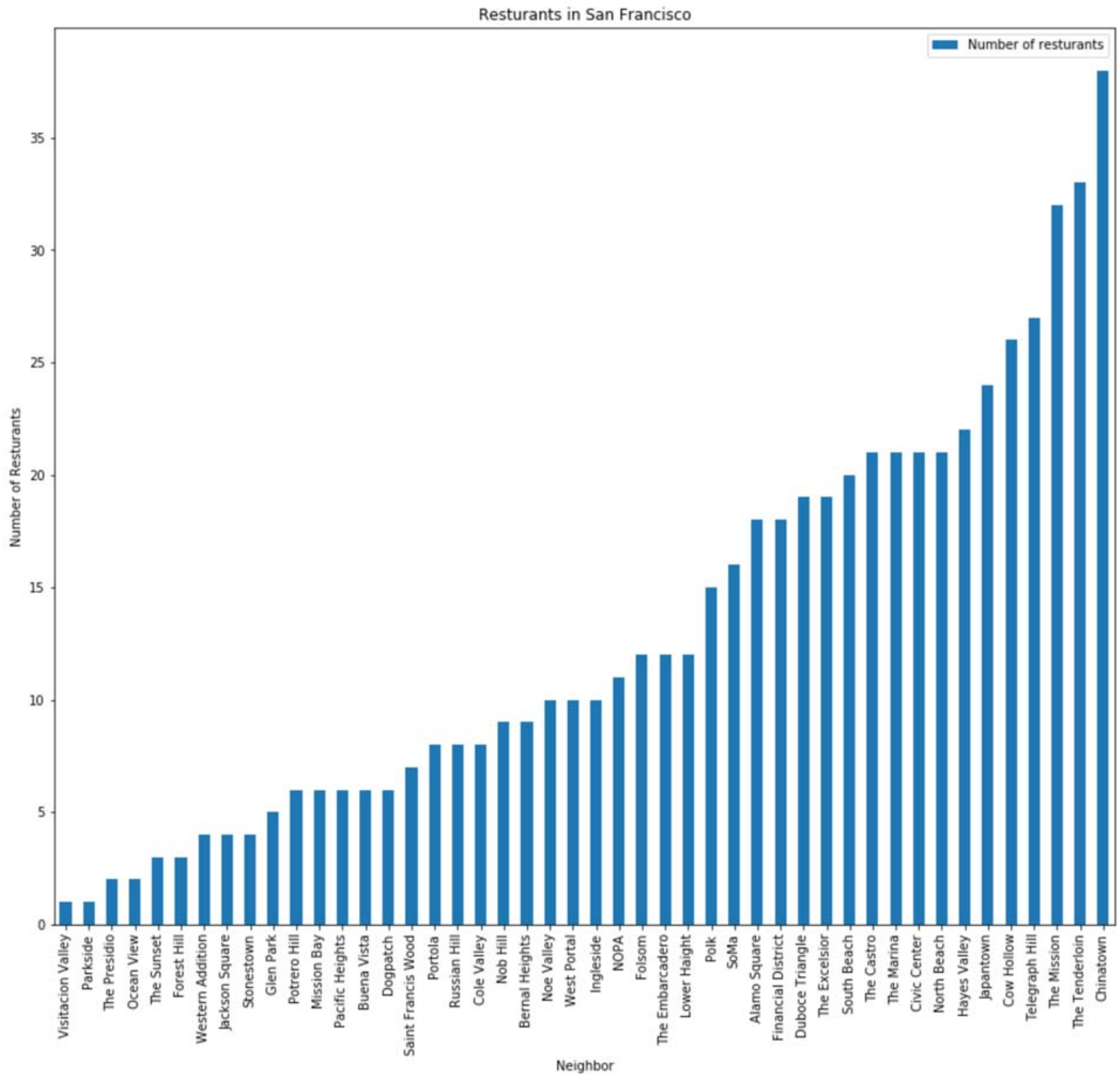
	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
5	Alamo Square	37.77636	-122.434689	Bar Crudo	37.775707	-122.438019	Seafood Restaurant
7	Alamo Square	37.77636	-122.434689	Nopa	37.774888	-122.437532	New American Restaurant
19	Alamo Square	37.77636	-122.434689	jū-ni	37.776743	-122.438770	Sushi Restaurant
21	Alamo Square	37.77636	-122.434689	Tsunami	37.776869	-122.438486	Sushi Restaurant
27	Alamo Square	37.77636	-122.434689	Brenda's Meat & Three	37.778265	-122.438584	Southern / Soul Food Restaurant
30	Alamo Square	37.77636	-122.434689	Zaytoon	37.775185	-122.437896	Mediterranean Restaurant
32	Alamo Square	37.77636	-122.434689	Kung Food	37.777778	-122.438698	Hunan Restaurant
34	Alamo Square	37.77636	-122.434689	Namu Stonepot	37.774763	-122.437780	Korean Restaurant
39	Alamo Square	37.77636	-122.434689	Che Fico	37.777435	-122.438149	Italian Restaurant
41	Alamo Square	37.77636	-122.434689	Saffron Grill	37.776848	-122.437816	Indian Restaurant

In [23]:
```
SF_rest_count=SF_rest.groupby('Neighborhood').count()[['Venue']]
SF_rest_count.rename(columns={'Venue': 'Number of resturants'}, inplace=True)
```

In [24]:
```
SF_rest_count.sort_values(by='Number of resturants', inplace=True)
```

In [34]:
```
SF_rest_count.plot(kind='bar', figsize=(15, 13))
plt.xlabel('Neighbor') # add to x-label to the plot
plt.ylabel('Number of Resturants') # add y-label to the plot
plt.title('Resturants in San Francisco') # add title to the plot

# plt.show()
plt.savefig('Resturant.png', dpi=300)
```



Extracting Vegetarian Restaurants

In [35]:
```
Rest_v=[]
for i in range (SF_venues.shape[0]):
    Rest_v.append('Vegan' in SF_venues['Venue Category'].apply(lambda x: x.split(
```

In [36]:
```
SF_Veg=SF_venues[Rest_v]
print('There are total of {} vegeterian resturant in San Francisco'. format(SF_V
```

```
There are total of 12 vegeterian resturant in San Francisco
```

In [37]:
```
SF_Veg.shape
```

Out[37]:
```
(12, 7)
```

In [51]:
```
SF_Veg.head(10)
```

258	Chinatown	37.794301	-122.406376	Lucky Creation Restaurant	37.795056	-122.407271	Vegetariar / Vegar Restauran
299	Civic Center	37.779026	-122.419906	Ananda Fuara	37.777693	-122.416353	Vegetariar / Vegar Restauran
327	Civic Center	37.779026	-122.419906	Golden Era Vegan	37.781495	-122.416822	Vegetariar / Vegar Restauran
354	Civic Center	37.779026	-122.419906	Thai Idea Vegetarian Restaurant	37.783395	-122.419141	Vegetariar / Vegar Restauran
455	Cow Hollow	37.797262	-122.436248	Wildseed	37.797626	-122.432440	Vegetariar / Vegar Restauran
487	Cow Hollow	37.797262	-122.436248	Vegan Picnic	37.797490	-122.431748	Vegetariar / Vegar Restauran

In [39]:
```
Veg_count=SF_Veg.groupby('Neighborhood').count()[['Venue']]
```

In [40]:
```
Veg_count.rename(columns={'Venue': 'Number of Vegi resturants'}, inplace=True)
```

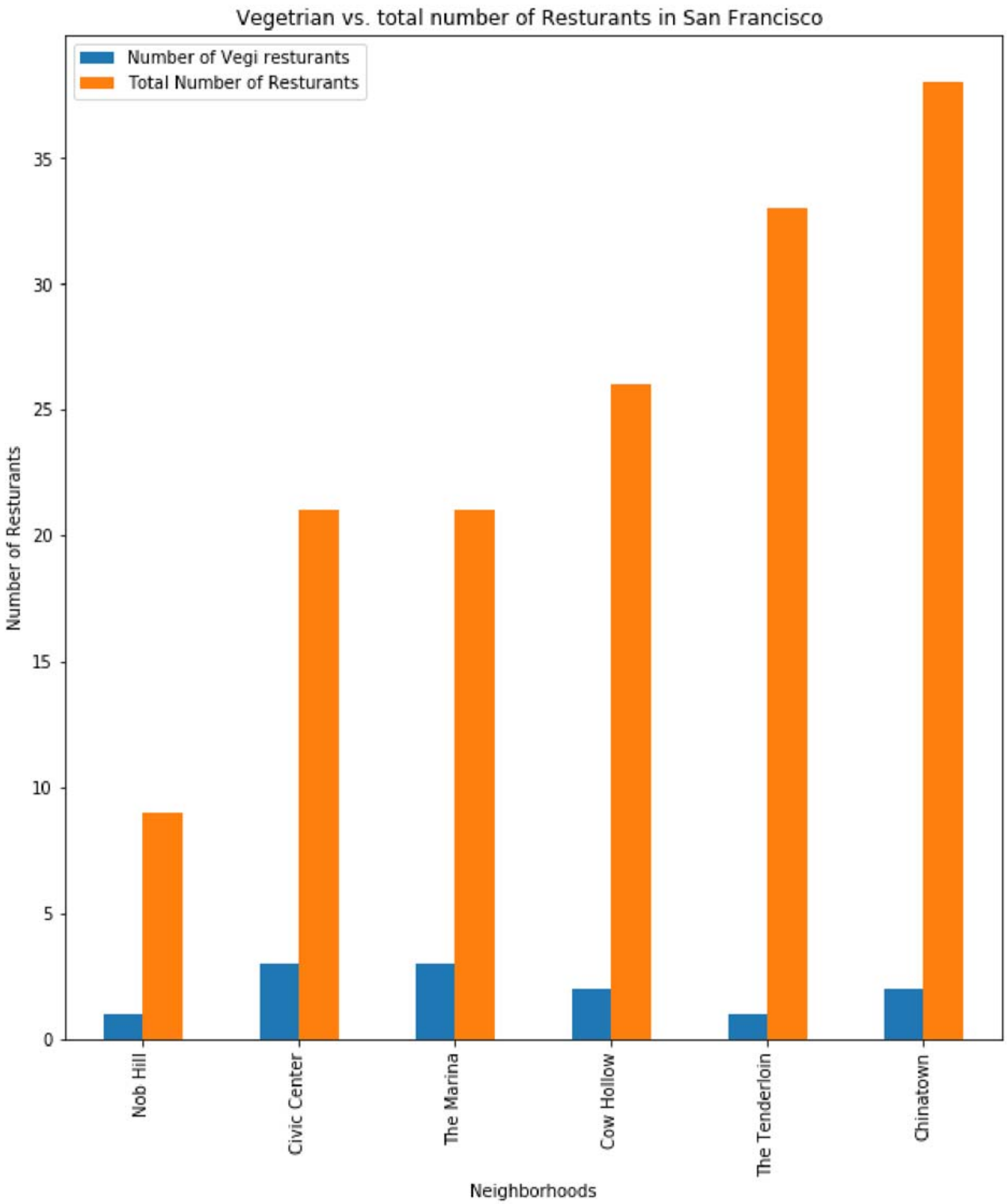
In [41]:
```
Veg_count['Total Number of Resturants']=SF_rest_count.loc[Veg_count.index]
```

In [42]:
```
Veg_count.sort_values(by='Total Number of Resturants', inplace=True)
```

In [43]:
```
labels=SF_rest_count.index
```

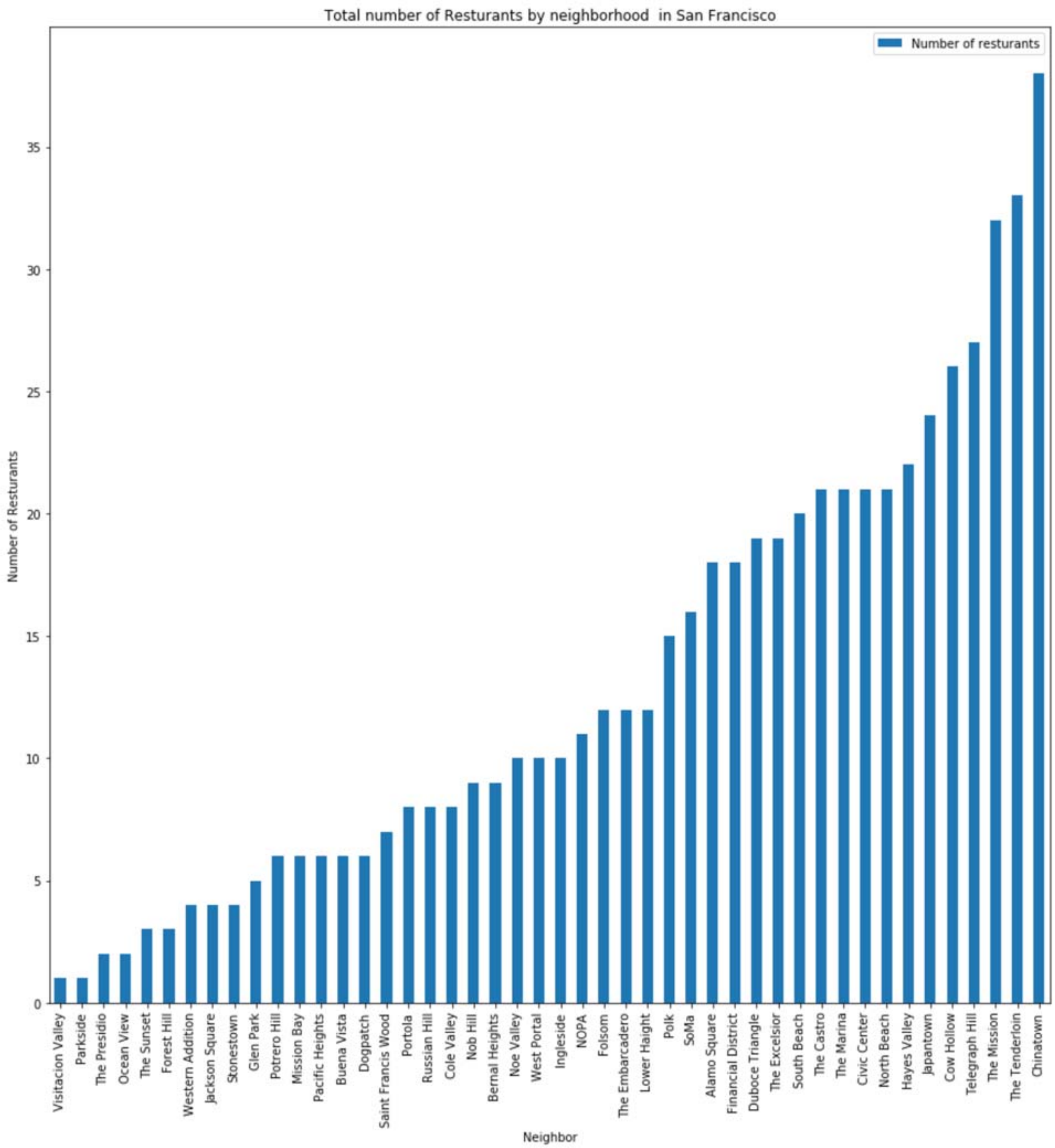
In [59]:
```
Veg_count.plot(kind='bar', figsize=(10, 11))
plt.xlabel('Neighborhoods') # add to x-label to the plot
plt.ylabel('Number of Resturants') # add y-label to the plot
plt.title('Vegetrian vs. total number of Resturants in San Francisco ') # add ti

plt.savefig('Veg Resturant.png')
```



In [56]:
```
SF_rest_count.plot(kind='bar', figsize=(15, 15))
# Veg_count.plot(kind='bar', figsize=(10, 6), ax=ax)
plt.xlabel('Neighbor') # add to x-label to the plot
plt.ylabel('Number of Resturants') # add y-label to the plot
plt.title('Total number of Resturants by neighborhood  in San Francisco') # add 

plt.savefig('Resturant.png')
```



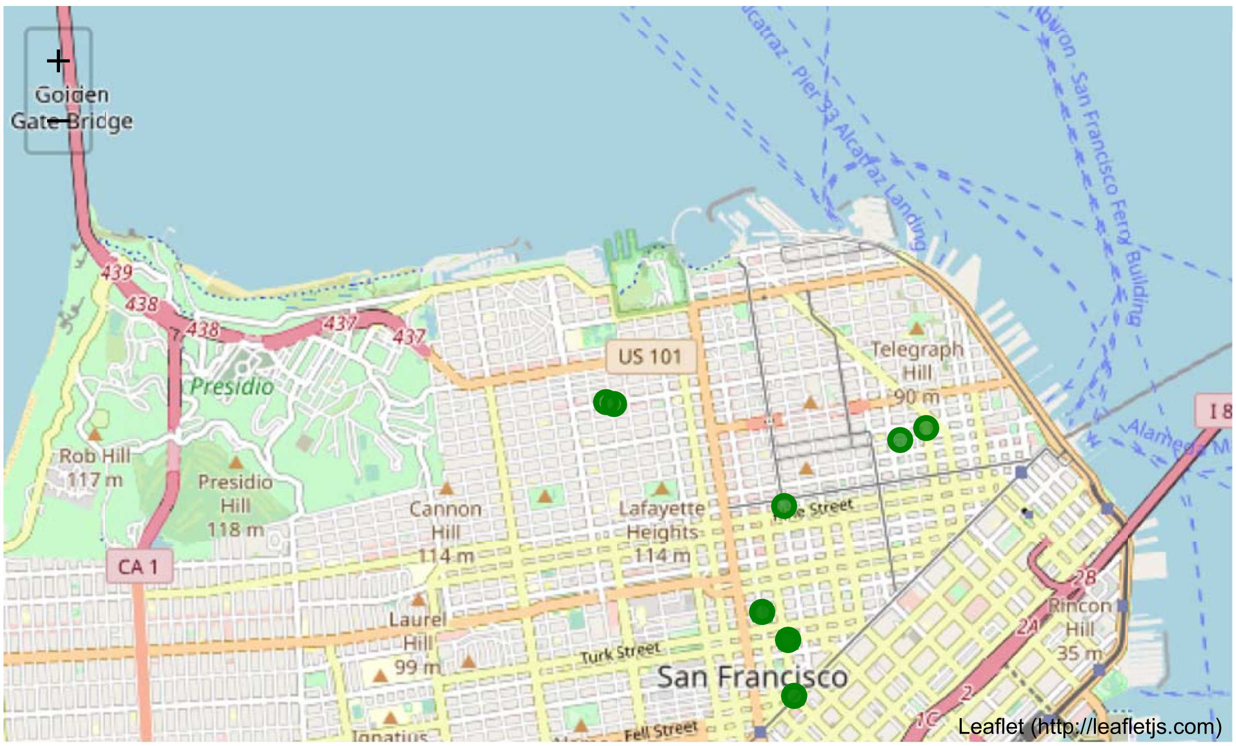
Visualizing Vegeterian restaurants on Map

In [60]:
```
# create map of SF using latitude and longitude values
map_SF = folium.Map(location=[latitude_sf, longitude_sf], zoom_start=12)

# add markers to map
for lat, lng, neighborhood in zip(SF_Veg['Venue Latitude'], SF_Veg['Venue Longit
    label = '{}'.format(neighborhood)
    label = folium.Popup(label, parse_html=True)
    folium.CircleMarker(
        [lat, lng],
        radius=5,
        popup=label,
        color='green',
        fill=True,
#         fill_color='#3186cc',
        fill_opacity=0.7,
        parse_html=False).add_to(map_SF)

map_SF
```

Out[60]:



Analyzing Each Neighborhood for all resturants

In [61]:
```
# one hot encoding
SF_onehot = pd.get_dummies(SF_rest[['Venue Category']], prefix="", prefix_sep=""

# add neighborhood column back to dataframe
SF_onehot['Neighborhood'] = SF_rest['Neighborhood']

# move neighborhood column to the first column
fixed_columns = [SF_onehot.columns[-1]] + list(SF_onehot.columns[:-1])
SF_onehot = SF_onehot[fixed_columns]

SF_onehot.head()
```

Out[61]:

	Neighborhood	African Restaurant	American Restaurant	Argentinian Restaurant	Asian Restaurant	Austrian Restaurant	Brazilian Restaurant	Burm Restau
5	Alamo Square	0	0	0	0	0	0	
7	Alamo Square	0	0	0	0	0	0	
19	Alamo Square	0	0	0	0	0	0	
21	Alamo Square	0	0	0	0	0	0	
27	Alamo Square	0	0	0	0	0	0	

In [62]:
```
SF_onehot.shape
```

Out[62]: (596, 61)

grouping rows by neighborhood and by taking the mean of the frequency of occurrence of each restaurant category

In [63]:
```
SF_grouped = SF_onehot.groupby('Neighborhood').mean().reset_index()
SF_grouped
```

Out[63]:

	Neighborhood	African Restaurant	American Restaurant	Argentinian Restaurant	Asian Restaurant	Austrian Restaurant	Brazilian Restaurant	Burm Restau
0	Alamo Square	0.000000	0.000000	0.000000	0.000000	0.000	0.000000	0.00(
1	Bernal Heights	0.000000	0.111111	0.000000	0.111111	0.000	0.000000	0.00(
2	Buena Vista	0.000000	0.000000	0.000000	0.000000	0.000	0.166667	0.00(
3	Chinatown	0.000000	0.000000	0.000000	0.000000	0.000	0.000000	0.00(
4	Civic Center	0.000000	0.047619	0.000000	0.000000	0.000	0.000000	0.00(
5	Cole Valley	0.000000	0.000000	0.000000	0.000000	0.000	0.000000	0.00(
6	Cow Hollow	0.000000	0.076923	0.000000	0.038462	0.000	0.000000	0.03ε
7	Dogpatch	0.000000	0.000000	0.000000	0.000000	0.000	0.000000	0.00(
8	Duboce Triangle	0.000000	0.000000	0.000000	0.052632	0.000	0.000000	0.00(
9	Financial District	0.000000	0.000000	0.000000	0.000000	0.000	0.000000	0.00(
10	Folsom	0.083333	0.083333	0.000000	0.000000	0.000	0.000000	0.00(
11	Forest Hill	0.000000	0.000000	0.000000	0.000000	0.000	0.000000	0.00(
12	Glen Park	0.000000	0.000000	0.000000	0.000000	0.000	0.000000	0.00(
13	Hayes Valley	0.000000	0.045455	0.000000	0.000000	0.000	0.000000	0.00(
14	Ingleside	0.000000	0.100000	0.000000	0.100000	0.000	0.000000	0.00(
15	Jackson Square	0.000000	0.000000	0.000000	0.000000	0.000	0.000000	0.00(
16	Japantown	0.000000	0.083333	0.000000	0.041667	0.000	0.000000	0.00(
17	Lower Haight	0.000000	0.000000	0.000000	0.000000	0.000	0.000000	0.00(
18	Mission Bay	0.000000	0.166667	0.000000	0.166667	0.000	0.000000	0.00(
19	NOPA	0.000000	0.000000	0.000000	0.000000	0.000	0.000000	0.00(
20	Nob Hill	0.000000	0.111111	0.000000	0.000000	0.000	0.000000	0.00(
21	Noe Valley	0.000000	0.100000	0.000000	0.000000	0.000	0.000000	0.10(
22	North Beach	0.000000	0.000000	0.047619	0.000000	0.000	0.000000	0.00(
23	Ocean View	0.000000	0.000000	0.000000	0.000000	0.000	0.000000	0.00(
24	Pacific Heights	0.000000	0.166667	0.000000	0.000000	0.000	0.000000	0.00(
25	Parkside	0.000000	0.000000	0.000000	0.000000	0.000	0.000000	0.00(
26	Polk	0.000000	0.066667	0.000000	0.000000	0.000	0.000000	0.00(
27	Portola	0.000000	0.000000	0.000000	0.000000	0.000	0.000000	0.00(
28	Potrero Hill	0.000000	0.000000	0.000000	0.000000	0.000	0.000000	0.00(
29	Russian Hill	0.000000	0.000000	0.000000	0.000000	0.125	0.000000	0.00(

	Neighborhood	African Restaurant	American Restaurant	Argentinian Restaurant	Asian Restaurant	Austrian Restaurant	Brazilian Restaurant	Burm Restau
30	Saint Francis Wood	0.000000	0.000000	0.000000	0.000000	0.000	0.000000	0.00(
31	SoMa	0.000000	0.062500	0.000000	0.000000	0.000	0.125000	0.00(
32	South Beach	0.000000	0.100000	0.000000	0.000000	0.000	0.000000	0.00(
33	Stonestown	0.000000	0.000000	0.000000	0.000000	0.000	0.000000	0.00(
34	Telegraph Hill	0.000000	0.074074	0.000000	0.000000	0.000	0.000000	0.00(
35	The Castro	0.000000	0.047619	0.000000	0.000000	0.000	0.000000	0.00(
36	The Embarcadero	0.000000	0.166667	0.000000	0.083333	0.000	0.000000	0.00(
37	The Excelsior	0.000000	0.000000	0.000000	0.000000	0.000	0.000000	0.00(
38	The Marina	0.000000	0.047619	0.000000	0.000000	0.000	0.000000	0.00(
39	The Mission	0.000000	0.031250	0.000000	0.031250	0.000	0.000000	0.00(
40	The Presidio	0.000000	1.000000	0.000000	0.000000	0.000	0.000000	0.00(
41	The Sunset	0.000000	0.000000	0.000000	0.000000	0.000	0.000000	0.00(
42	The Tenderloin	0.000000	0.060606	0.000000	0.000000	0.000	0.000000	0.00(
43	Visitacion Valley	0.000000	0.000000	0.000000	0.000000	0.000	0.000000	0.00(
44	West Portal	0.000000	0.000000	0.000000	0.000000	0.000	0.100000	0.00(
45	Western Addition	0.000000	0.000000	0.000000	0.000000	0.000	0.000000	0.00(

In [64]:
```
SF_grouped.shape
```

Out[64]:
```
(46, 61)
```

printing each neighborhood along with the top 5 most common restaurants

In [65]:
```
num_top_venues = 5

for hood in SF_grouped['Neighborhood']:
    print("----"+hood+"----")
    temp = SF_grouped[SF_grouped['Neighborhood'] == hood].T.reset_index()
    temp.columns = ['venue','freq']
    temp = temp.iloc[1:]
    temp['freq'] = temp['freq'].astype(float)
    temp = temp.round({'freq': 2})
    print(temp.sort_values('freq', ascending=False).reset_index(drop=True).head(
    print('\n')
```

```
4      Japanese Restaurant  0.05


----Civic Center----
                           venue  freq
0  Vegetarian / Vegan Restaurant  0.14
1               Sushi Restaurant  0.14
2              French Restaurant  0.14
3                     Restaurant  0.10
4          Vietnamese Restaurant  0.05


----Cole Valley----
                       venue  freq
0      Vietnamese Restaurant  0.12
1           Tapas Restaurant  0.12
2  Middle Eastern Restaurant  0.12
3          Indian Restaurant  0.12
4   Mediterranean Restaurant  0.12
```

Puting that into a pandas dataframe

In [66]:
```
def return_most_common_venues(row, num_top_venues):
    row_categories = row.iloc[1:]
    row_categories_sorted = row_categories.sort_values(ascending=False)

    return row_categories_sorted.index.values[0:num_top_venues]
```

In [67]:
```
num_top_venues = 10

indicators = ['st', 'nd', 'rd']

# create columns according to number of top venues
columns = ['Neighborhood']
for ind in np.arange(num_top_venues):
    try:
        columns.append('{}{} Most Common Venue'.format(ind+1, indicators[ind]))
    except:
        columns.append('{}th Most Common Venue'.format(ind+1))

# create a new dataframe
neighborhoods_venues_sorted = pd.DataFrame(columns=columns)
neighborhoods_venues_sorted['Neighborhood'] = SF_grouped['Neighborhood']

for ind in np.arange(SF_grouped.shape[0]):
    neighborhoods_venues_sorted.iloc[ind, 1:] = return_most_common_venues(SF_grou
```

In [68]:
```
neighborhoods_venues_sorted.head(10)
```

Out[68]:

	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	
0	Alamo Square	Indian Restaurant	Seafood Restaurant	Ethiopian Restaurant	Sushi Restaurant	Hunan Restaurant	Mediterranean Restaurant	I
1	Bernal Heights	Italian Restaurant	American Restaurant	Asian Restaurant	Mexican Restaurant	Indian Restaurant	Caribbean Restaurant	New I
2	Buena Vista	Seafood Restaurant	Brazilian Restaurant	Fast Food Restaurant	Cantonese Restaurant	Korean Restaurant	Austrian Restaurant	I
3	Chinatown	Chinese Restaurant	New American Restaurant	Italian Restaurant	Vietnamese Restaurant	Restaurant	Dim Sum Restaurant	I
4	Civic Center	Vegetarian / Vegan Restaurant	French Restaurant	Sushi Restaurant	Restaurant	Vietnamese Restaurant	Southern / Soul Food Restaurant	Mec I
5	Cole Valley	Vietnamese Restaurant	Thai Restaurant	Caribbean Restaurant	Mexican Restaurant	Tapas Restaurant	Middle Eastern Restaurant	Mec I
6	Cow Hollow	Italian Restaurant	Mexican Restaurant	French Restaurant	American Restaurant	Vegetarian / Vegan Restaurant	Thai Restaurant	I
7	Dogpatch	Restaurant	Latin American Restaurant	Italian Restaurant	Southern / Soul Food Restaurant	Sushi Restaurant	Vietnamese Restaurant	I
8	Duboce Triangle	New American Restaurant	Seafood Restaurant	Mexican Restaurant	Sushi Restaurant	Vietnamese Restaurant	Indian Restaurant	Mec I
9	Financial District	Restaurant	Japanese Restaurant	Mediterranean Restaurant	New American Restaurant	Dim Sum Restaurant	Latin American Restaurant	I

Clustering Neighborhoods

In [69]:
```
kclusters = 5
SF_grouped_clustering = SF_grouped.drop('Neighborhood', 1)
kmeans = KMeans(n_clusters=kclusters, random_state=0).fit(SF_grouped_clustering)
```

In [70]:
```
kmeans.labels_
```

Out[70]:
```
array([1, 2, 2, 1, 1, 2, 1, 1, 2, 2, 2, 1, 1, 1, 2, 2, 1, 1, 2, 1, 1, 1,
       1, 0, 1, 0, 1, 0, 1, 1, 0, 1, 2, 2, 1, 2, 2, 2, 1, 2, 4, 2, 2, 3,
       1, 1])
```

adding cluster labels to a new dataframe

In [ ]:
```
# add clustering labels

neighborhoods_venues_sorted.insert(0, 'Cluster Labels', kmeans.labels_)

SF_merged = sf_neigh

# merge manhattan_grouped with manhattan_data to add latitude/longitude for each
SF_merged = SF_merged.join(neighborhoods_venues_sorted.set_index('Neighborhood')
SF_merged.dropna(inplace=True)
```

In [73]:
```
SF_merged.head(10) # check the last columns!
```

Out[73]:

Neighborhood	latitude	longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	
Alamo Square	37.776360	-122.434689	1.0	Indian Restaurant	Seafood Restaurant	Ethiopian Restaurant	Sushi Restaurant	F
Bernal Heights	37.741001	-122.414214	2.0	Italian Restaurant	American Restaurant	Asian Restaurant	Mexican Restaurant	F
Buena Vista	37.806532	-122.420648	2.0	Seafood Restaurant	Brazilian Restaurant	Fast Food Restaurant	Cantonese Restaurant	F
Chinatown	37.794301	-122.406376	1.0	Chinese Restaurant	New American Restaurant	Italian Restaurant	Vietnamese Restaurant	F
Civic Center	37.779026	-122.419906	1.0	Vegetarian / Vegan Restaurant	French Restaurant	Sushi Restaurant	Restaurant	Vi F
Cole Valley	37.765813	-122.449962	2.0	Vietnamese Restaurant	Thai Restaurant	Caribbean Restaurant	Mexican Restaurant	F
Cow Hollow	37.797262	-122.436248	1.0	Italian Restaurant	Mexican Restaurant	French Restaurant	American Restaurant	\ F
Dogpatch	37.760698	-122.389202	1.0	Restaurant	Latin American Restaurant	Italian Restaurant	Southern / Soul Food Restaurant	F
Duboce Triangle	37.767138	-122.432230	2.0	New American Restaurant	Seafood Restaurant	Mexican Restaurant	Sushi Restaurant	Vi F
Financial District	37.793647	-122.398938	2.0	Restaurant	Japanese Restaurant	Mediterranean Restaurant	New American Restaurant	F

Visulizing the clusters

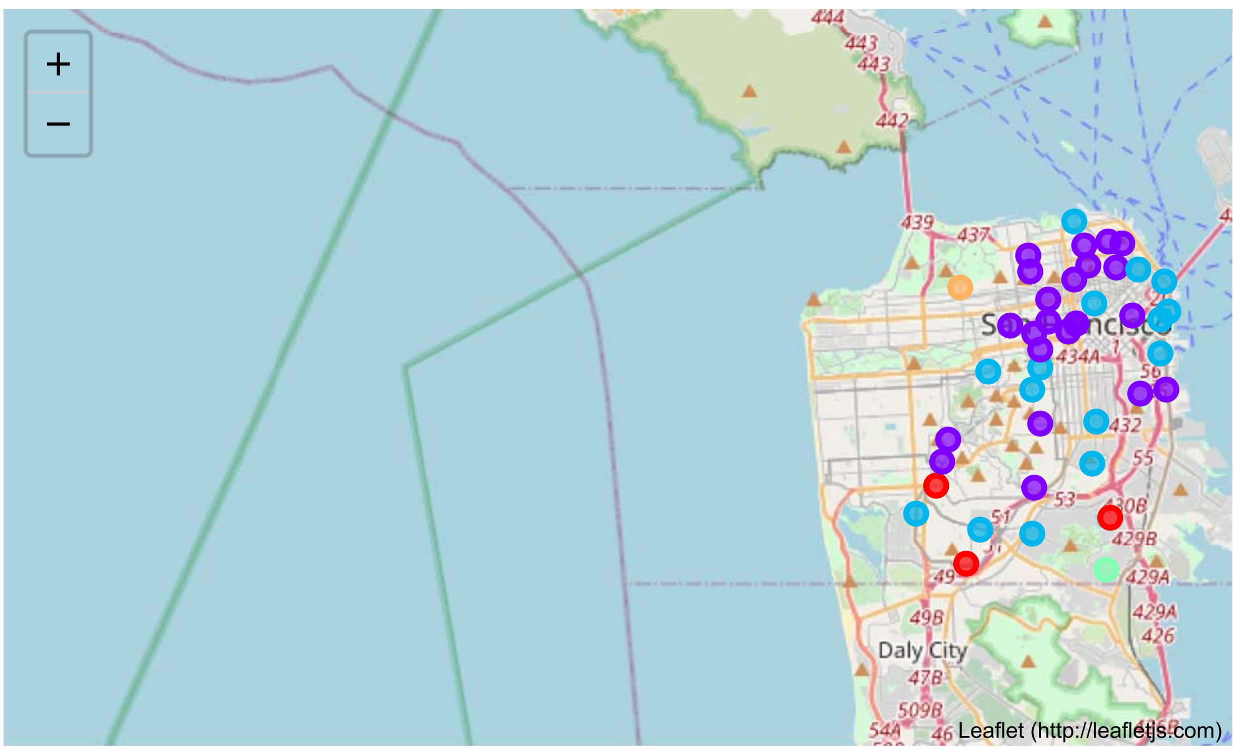
In [72]:
```python
# create map
map_clusters = folium.Map(location=[latitude_sf, longitude_sf], zoom_start=11)

# set color scheme for the clusters
x = np.arange(kclusters)
ys = [i + x + (i*x)**2 for i in range(kclusters)]
colors_array = cm.rainbow(np.linspace(0, 1, len(ys)))
rainbow = [colors.rgb2hex(i) for i in colors_array]

# add markers to the map
markers_colors = []
for lat, lon, poi, cluster in zip(SF_merged['latitude'], SF_merged['longitude'],
    label = folium.Popup(str(poi) + ' Cluster ' + str(cluster), parse_html=True)
    cluster=int(cluster)
    folium.CircleMarker(
        [lat, lon],
        radius=5,
        popup=label,
        color=rainbow[cluster-1],
        fill=True,
        fill_color=rainbow[cluster-1],
        fill_opacity=0.7).add_to(map_clusters)

map_clusters
```

Out[72]:



Examining clusters

Cluster 1

In [288]:

```
SF_merged.loc[SF_merged['Cluster Labels'] == 0, SF_merged.columns[[0] + list(ran
```

Out[288]:

	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 Mos Commo Venu
32	Ocean View	Thai Restaurant	Chinese Restaurant	Vietnamese Restaurant	Ethiopian Restaurant	Japanese Restaurant	Japanese Curry Restaurant	Italia Restaurar
35	Parkside	Chinese Restaurant	Vietnamese Restaurant	Fast Food Restaurant	Jewish Restaurant	Japanese Restaurant	Japanese Curry Restaurant	Italia Restaurar
38	Portola	Vietnamese Restaurant	Chinese Restaurant	Cantonese Restaurant	Dim Sum Restaurant	Filipino Restaurant	Jewish Restaurant	Japanes Restaurar
42	Saint Francis Wood	Sushi Restaurant	Chinese Restaurant	Vietnamese Restaurant	Indian Restaurant	Greek Restaurant	Ethiopian Restaurant	Japanes Restaurar

Cluster 2

In [289]:
```
SF_merged.loc[SF_merged['Cluster Labels'] == 1, SF_merged.columns[[0] + list(ran
```

Out[289]:

	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
0	Alamo Square	Indian Restaurant	Seafood Restaurant	Ethiopian Restaurant	Sushi Restaurant	Hunan Restaurant	Mediterranean Restaurant
5	Chinatown	Chinese Restaurant	New American Restaurant	Italian Restaurant	Vietnamese Restaurant	Restaurant	Dim Sum Restaurant
6	Civic Center	Vegetarian / Vegan Restaurant	French Restaurant	Sushi Restaurant	Restaurant	Vietnamese Restaurant	Southern / Soul Food Restaurant
8	Cow Hollow	Italian Restaurant	Mexican Restaurant	French Restaurant	American Restaurant	Vegetarian / Vegan Restaurant	Thai Restaurant
11	Dogpatch	Restaurant	Latin American Restaurant	Italian Restaurant	Southern / Soul Food Restaurant	Sushi Restaurant	Vietnamese Restaurant
16	Forest Hill	Japanese Restaurant	Hotpot Restaurant	French Restaurant	Vietnamese Restaurant	Vegetarian / Vegan Restaurant	Jewish Restaurant
17	Glen Park	Italian Restaurant	Mexican Restaurant	Thai Restaurant	Chinese Restaurant	Sushi Restaurant	Vietnamese Restaurant
19	Hayes Valley	French Restaurant	Sushi Restaurant	New American Restaurant	Italian Restaurant	Restaurant	Indian Restaurant
24	Japantown	Vietnamese Restaurant	American Restaurant	Thai Restaurant	Italian Restaurant	Japanese Restaurant	Sushi Restaurant
25	Lower Haight	Sushi Restaurant	Vietnamese Restaurant	Indian Restaurant	Italian Restaurant	Japanese Restaurant	Mediterranean Restaurant
28	Nob Hill	Italian Restaurant	French Restaurant	Vegetarian / Vegan Restaurant	American Restaurant	Thai Restaurant	Fast Food Restaurant
29	Noe Valley	Italian Restaurant	Sushi Restaurant	Restaurant	American Restaurant	Peruvian Restaurant	Mexican Restaurant
30	NOPA	Sushi Restaurant	Eastern European Restaurant	Italian Restaurant	Indian Restaurant	Hunan Restaurant	Mexican Restaurant
31	North Beach	Italian Restaurant	Chinese Restaurant	Mexican Restaurant	Persian Restaurant	French Restaurant	Restaurant
33	Pacific Heights	French Restaurant	American Restaurant	Turkish Restaurant	Mediterranean Restaurant	Italian Restaurant	Vietnamese Restaurant
37	Polk	Sushi Restaurant	Thai Restaurant	Vietnamese Restaurant	Italian Restaurant	American Restaurant	Mediterranean Restaurant
39	Potrero Hill	Vietnamese Restaurant	Mediterranean Restaurant	Mexican Restaurant	Sushi Restaurant	Peruvian Restaurant	French Restaurant

	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
41	Russian Hill	Italian Restaurant	Sushi Restaurant	Austrian Restaurant	Indian Restaurant	Vietnamese Restaurant	Ethiopian Restaurant
43	SoMa	Vietnamese Restaurant	Brazilian Restaurant	French Restaurant	New American Restaurant	Restaurant	American Restaurant
46	Telegraph Hill	Italian Restaurant	New American Restaurant	American Restaurant	Seafood Restaurant	Chinese Restaurant	Mexican Restaurant
52	The Marina	Vegetarian / Vegan Restaurant	French Restaurant	Sushi Restaurant	Restaurant	Vietnamese Restaurant	Southern / Soul Food Restaurant
60	Western Addition	Seafood Restaurant	Pakistani Restaurant	German Restaurant	French Restaurant	Ethiopian Restaurant	Japanese Restaurant
61	West Portal	Italian Restaurant	Mexican Restaurant	Indian Restaurant	Mediterranean Restaurant	Thai Restaurant	Brazilian Restaurant

Cluster 3

In [290]:
```
SF_merged.loc[SF_merged['Cluster Labels'] == 2, SF_merged.columns[[0] + list(ran
```

Out[290]:

	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	
2	Bernal Heights	Italian Restaurant	American Restaurant	Asian Restaurant	Mexican Restaurant	Indian Restaurant	Caribbean Restaurant	Ne
3	Buena Vista	Seafood Restaurant	Brazilian Restaurant	Fast Food Restaurant	Cantonese Restaurant	Korean Restaurant	Austrian Restaurant	
7	Cole Valley	Vietnamese Restaurant	Thai Restaurant	Caribbean Restaurant	Mexican Restaurant	Tapas Restaurant	Middle Eastern Restaurant	Me
12	Duboce Triangle	New American Restaurant	Seafood Restaurant	Mexican Restaurant	Sushi Restaurant	Vietnamese Restaurant	Indian Restaurant	Me
14	Financial District	Restaurant	Japanese Restaurant	Mediterranean Restaurant	New American Restaurant	Dim Sum Restaurant	Latin American Restaurant	
15	Folsom	Seafood Restaurant	Japanese Restaurant	Vietnamese Restaurant	American Restaurant	Dim Sum Restaurant	French Restaurant	Me
22	Ingleside	Vietnamese Restaurant	Pakistani Restaurant	American Restaurant	Asian Restaurant	Dim Sum Restaurant	Japanese Restaurant	
23	Jackson Square	Fast Food Restaurant	Mexican Restaurant	Italian Restaurant	Vietnamese Restaurant	Jiangsu Restaurant	Japanese Restaurant	
27	Mission Bay	Ethiopian Restaurant	Asian Restaurant	Dumpling Restaurant	Fast Food Restaurant	Mediterranean Restaurant	American Restaurant	
44	South Beach	Mexican Restaurant	Thai Restaurant	New American Restaurant	American Restaurant	Mediterranean Restaurant	Korean Restaurant	
45	Stonestown	Japanese Restaurant	Udon Restaurant	Mexican Restaurant	Taiwanese Restaurant	Ethiopian Restaurant	Japanese Curry Restaurant	
48	The Castro	Thai Restaurant	New American Restaurant	Mediterranean Restaurant	Indian Restaurant	Seafood Restaurant	Japanese Restaurant	
50	The Embarcadero	Mexican Restaurant	American Restaurant	New American Restaurant	Indian Restaurant	Chinese Restaurant	Italian Restaurant	
51	The Excelsior	Mexican Restaurant	Vietnamese Restaurant	Chinese Restaurant	Latin American Restaurant	Restaurant	Fast Food Restaurant	
53	The Mission	Mexican Restaurant	Latin American Restaurant	Sushi Restaurant	Salvadoran Restaurant	South Indian Restaurant	Jewish Restaurant	
55	The Sunset	Fast Food Restaurant	Mexican Restaurant	New American Restaurant	Vietnamese Restaurant	Japanese Restaurant	Japanese Curry Restaurant	
56	The Tenderloin	Thai Restaurant	Vietnamese Restaurant	Mexican Restaurant	Sushi Restaurant	American Restaurant	Filipino Restaurant	

Cluster 4

In [291]:
```
SF_merged.loc[SF_merged['Cluster Labels'] == 3, SF_merged.columns[[0] + list(ran
```

Out[291]:

	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
59	Visitacion Valley	Vietnamese Restaurant	Vegetarian / Vegan Restaurant	Jewish Restaurant	Japanese Restaurant	Japanese Curry Restaurant	Italian Restaurant	Indian Restaurant

Cluster 5

In [292]:
```
SF_merged.loc[SF_merged['Cluster Labels'] == 4, SF_merged.columns[[0] + list(ran
```

Out[292]:

	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
54	The Presidio	American Restaurant	Vietnamese Restaurant	Fast Food Restaurant	Jewish Restaurant	Japanese Restaurant	Japanese Curry Restaurant	Italian Restaurant

Discussion

In [74]:
```
The_mission=SF_merged[SF_merged['Neighborhood']== 'The Mission']
```

In [75]:
```
The_mission
```

Out[75]:

	Neighborhood	latitude	longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	
53	The Mission	37.752498	-122.412826	2.0	Mexican Restaurant	Latin American Restaurant	Sushi Restaurant	Salvadoran Restaurant	F

Recomended Location

In [76]:

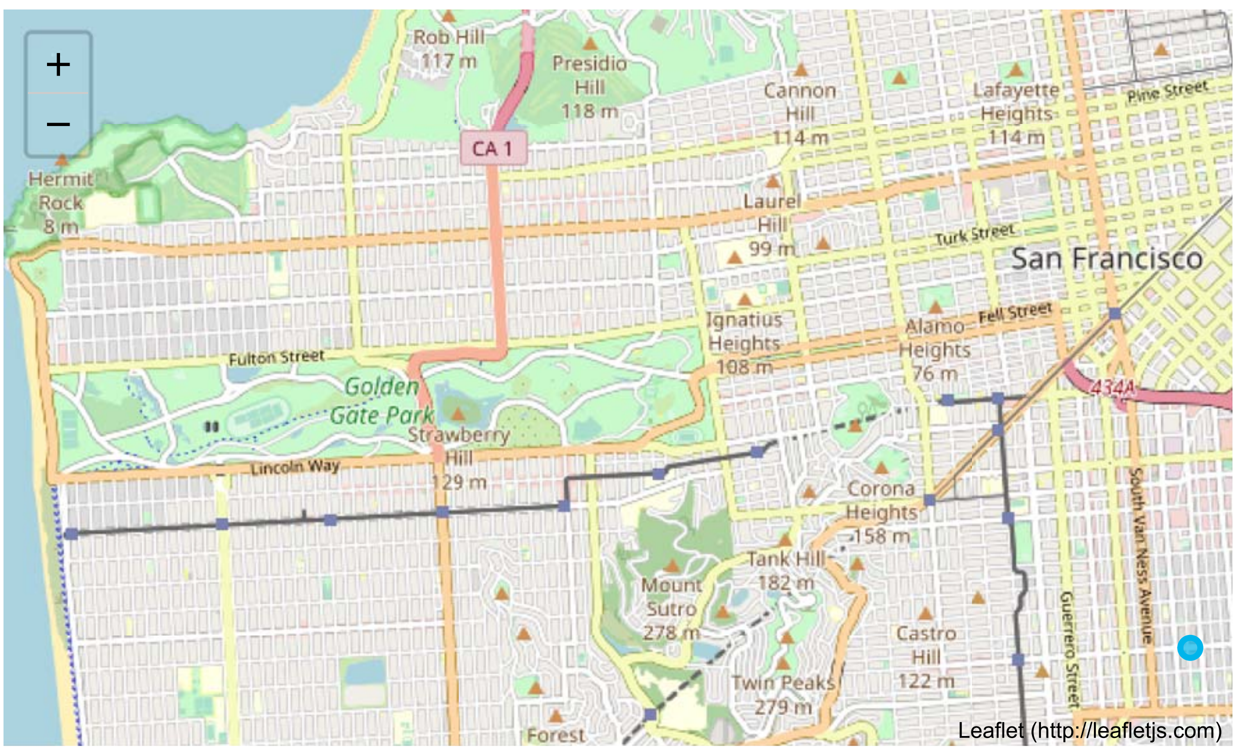
```
# create map
map_clusters = folium.Map(location=[latitude_sf, longitude_sf], zoom_start=11)

# set color scheme for the clusters
x = np.arange(kclusters)
ys = [i + x + (i*x)**2 for i in range(kclusters)]
colors_array = cm.rainbow(np.linspace(0, 1, len(ys)))
rainbow = [colors.rgb2hex(i) for i in colors_array]

# add markers to the map
markers_colors = []
for lat, lon, poi, cluster in zip(The_mission['latitude'], The_mission['longitude
    label = folium.Popup(str(poi) + ' Cluster ' + str(cluster), parse_html=True)
    cluster=int(cluster)
    folium.CircleMarker(
        [lat, lon],
        radius=5,
        popup=label,
        color=rainbow[cluster-1],
        fill=True,
        fill_color=rainbow[cluster-1],
        fill_opacity=0.7).add_to(map_clusters)

map_clusters
```

Out[76]:



In [ ]: