

# Capstone Project - The Battle of Neighborhoods

## Part 1: Introduction

San Francisco is one of the financial, cultural and technology centers on the west coast, it has population close to 900,000. San Francisco is a diverse and culturally rich city, where you expect people to live in different lifestyles across different neighborhoods. Indeed, some of its neighborhoods are cozy and relaxing, while some others are busy and commercialized. Therefore, it is very important for either a business to choose where to open a new store or a person to pick where to live with his or her lifestyle.

I want to utilize Foursquare location to identify venues within each neighborhood, and then use venues' frequencies within each neighborhood to create clusters that provide insightful information for business and people to choose the target neighborhoods to open a new business or live with a desired lifestyle.

## Part 2: Data Acquisition and Cleaning

### 1. Data Sources

The data is from the following sources:

- San Francisco neighborhood list: [Wikipedia SF Neighborhoods](#);
- Location data: [Opencage Geocoder](#);
- Venues data: [Foursquare](#);

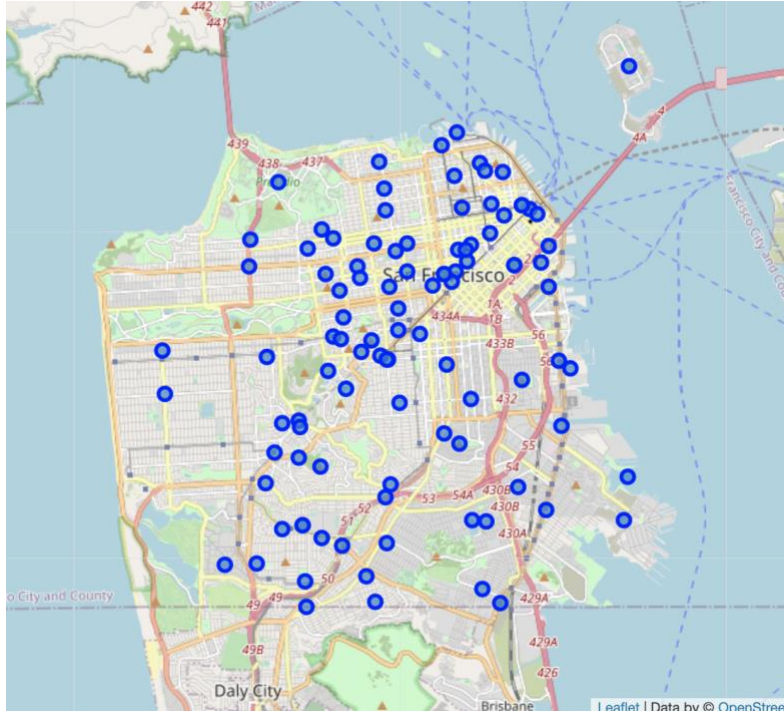
### 2. Data Cleaning

- For San Francisco neighborhood data, I used "mw-headline" class in Beautiful Soup Python library to extract the neighborhood list from the Wikipedia website;

| Neighborhood |                    |
|--------------|--------------------|
| 0            | Alamo Square       |
| 1            | Anza Vista         |
| 2            | Ashbury Heights    |
| 3            | Balboa Park        |
| 4            | Balboa Terrace     |
| ...          | ...                |
| 114          | West Portal        |
| 115          | Western Addition   |
| 116          | Westwood Highlands |
| 117          | Westwood Park      |
| 118          | Yerba Buena        |

119 rows × 1 columns

df



- b. For the location data, I used the Opencage Geocoder Python library and free API key from [Opencage](#) to create a data frame that includes the latitudes and longitudes for each of the neighborhood in San Francisco. I have noticed the data frame generated from the Opencage Python library are not 100% accurate, so I deleted the rows that contained inaccurate location data;

|     | Neighborhood       | lat       | lng         |
|-----|--------------------|-----------|-------------|
| 0   | Alamo Square       | 37.776360 | -122.434688 |
| 1   | Anza Vista         | 37.780836 | -122.443149 |
| 2   | Ashbury Heights    | 37.775599 | -122.448068 |
| 3   | Balboa Park        | 37.721427 | -122.447547 |
| 4   | Bayview            | 37.728889 | -122.392500 |
| ... | ...                | ...       | ...         |
| 101 | West Portal        | 37.741141 | -122.465634 |
| 102 | Western Addition   | 37.779559 | -122.429810 |
| 103 | Westwood Highlands | 37.725726 | -122.458199 |
| 104 | Westwood Park      | 37.725726 | -122.458199 |
| 105 | Butchertown        | 37.784827 | -122.727802 |

106 rows × 3 columns

df

- c. Once I had the location data, I used [Foursquare](#) to generate 200 venues with a radius of 1,000 meters from the coordinate for each neighborhood. Then I created a

funcation called getNearbyVenues, which generates a data frame that contains each neighborhood and venue information.

|      | Neighborhood  | Neighborhood Latitude | Neighborhood Longitude | Venue   | Venue Latitude | Venue Longitude | Venue Category |
|------|---------------|-----------------------|------------------------|---|----------------|-----------------|----------------|
| 0    | Alamo Square  | 37.776360             | -122.434688            | Alamo Square                                      | 37.776045      | -122.434363     | Park           |
| 1    | Alamo Square  | 37.776360             | -122.434688            | Alamo Square Dog Park                             | 37.775878      | -122.435740     | Dog Run        |
| 2    | Alamo Square  | 37.776360             | -122.434688            | Painted Ladies                                    | 37.776120      | -122.433389     | Historic Site  |
| 3    | Alamo Square  | 37.776360             | -122.434688            | The Independent                                   | 37.775573      | -122.437835     | Rock Club      |
| 4    | Alamo Square  | 37.776360             | -122.434688            | The Mill  | 37.776425      | -122.437970     | Bakery         |
| ...  | ...           | ...                   | ...                    | ...   | ...            | ...             | ...            |
| 5102 | Westwood Park | 37.725726             | -122.458199            | Old Balboa Reservoir Berm                         | 37.727362      | -122.453289     | Dog Run        |
| 5103 | Westwood Park | 37.725726             | -122.458199            | Orchids Cafe                                      | 37.723141      | -122.453710     | Cha Chaan Teng |
| 5104 | Westwood Park | 37.725726             | -122.458199            | Wiley's No Limit Liquor & Food Mart               | 37.723354      | -122.453505     | Liquor Store   |
| 5105 | Butchertown   | 37.784827             | -122.727802            | JazzDeck  | 37.784800      | -122.727800     | Music Venue    |
| 5106 | Butchertown   | 37.784827             | -122.727802            | Newly Formed Emerald Park to Benefit San Franc... | 37.783734      | -122.729690     | Park           |

5107 rows x 7 columns

sf\_venues

### 3. Data Exploration

After I obtained venue data from each neighborhood, I wanted to count the number of venues within each neighborhood and to display top 5 neighborhoods, and I found the following 5 neighborhoods have the largest numbers of the venues:

|    | Neighborhood              | count |
|----|---------------------------|-------|
| 89 | South of Market           | 196   |
| 37 | Hayes Valley              | 100   |
| 35 | Golden Gate Heights       | 100   |
| 13 | Chinatown                 | 100   |
| 9  | Butchertown (Old and New) | 100   |

I created dummy variables for the venue categories and calculated the mean for each venue category within each neighborhood to reveal how frequent those venue categories appeared.

|     | Neighborhood       | Yoga Studio | ATM | Acai House | Accessories Store | Acupuncturist | Adult Boutique | Alternative Healer | American Restaurant | Animal Shelter | ... | Video Store | Vietnamese Restaurant | Vineyard | Wareh |
|-----|--------------------|-------------|-----|------------|-------------------|---------------|----------------|--------------------|---------------------|----------------|-----|-------------|-----------------------|----------|-------|
| 0   | Alamo Square       | 0.013514    | 0.0 | 0.0        | 0.0               | 0.0           | 0.0            | 0.0                | 0.000000            | 0.0            | ... | 0.000000    | 0.000000              | 0.0      |       |
| 1   | Anza Vista         | 0.000000    | 0.0 | 0.0        | 0.0               | 0.0           | 0.0            | 0.0                | 0.000000            | 0.0            | ... | 0.000000    | 0.000000              | 0.0      |       |
| 2   | Ashbury Heights    | 0.038462    | 0.0 | 0.0        | 0.0               | 0.0           | 0.0            | 0.0                | 0.000000            | 0.0            | ... | 0.038462    | 0.000000              | 0.0      |       |
| 3   | Balboa Park        | 0.000000    | 0.0 | 0.0        | 0.0               | 0.0           | 0.0            | 0.0                | 0.000000            | 0.0            | ... | 0.000000    | 0.055556              | 0.0      |       |
| 4   | Bayview            | 0.000000    | 0.0 | 0.0        | 0.0               | 0.0           | 0.0            | 0.0                | 0.000000            | 0.0            | ... | 0.000000    | 0.000000              | 0.0      |       |
| ... | ...                | ...         | ... | ...        | ...               | ...           | ...            | ...                | ...                 | ...            | ... | ...         | ...                   | ...      | ...   |
| 100 | Vista del Mar      | 0.000000    | 0.0 | 0.0        | 0.0               | 0.0           | 0.0            | 0.0                | 0.010526            | 0.0            | ... | 0.000000    | 0.000000              | 0.0      |       |
| 101 | West Portal        | 0.023810    | 0.0 | 0.0        | 0.0               | 0.0           | 0.0            | 0.0                | 0.000000            | 0.0            | ... | 0.000000    | 0.000000              | 0.0      |       |
| 102 | Western Addition   | 0.000000    | 0.0 | 0.0        | 0.0               | 0.0           | 0.0            | 0.0                | 0.000000            | 0.0            | ... | 0.000000    | 0.000000              | 0.0      |       |
| 103 | Westwood Highlands | 0.056604    | 0.0 | 0.0        | 0.0               | 0.0           | 0.0            | 0.0                | 0.000000            | 0.0            | ... | 0.000000    | 0.037736              | 0.0      |       |
| 104 | Westwood Park      | 0.056604    | 0.0 | 0.0        | 0.0               | 0.0           | 0.0            | 0.0                | 0.000000            | 0.0            | ... | 0.000000    | 0.037736              | 0.0      |       |

105 rows x 338 columns

`sf_grouped`

I was curious about the top 10 venues within each neighborhood, so I created a data frame called `neighborhoods_venues_sorted` to obtain top 10 frequent venues within each neighborhood, following is the head of the `neighborhoods_venues_sorted`:

|   | Neighborhood    | 1st Most Common Venue | 2nd Most Common Venue           | 3rd Most Common Venue | 4th Most Common Venue | 5th Most Common Venue           | 6th Most Common Venue     | 7th Most Common Venue | 8th Most Common Venue | 9th Most Common Venue   | 10th Most Common Venue |
|---|-----------------|-----------------------|---------------------------------|-----------------------|-----------------------|---------------------------------|---------------------------|-----------------------|-----------------------|-------------------------|------------------------|
| 0 | Alamo Square    | Bar                   | Hotel                           | Café                  | Record Shop           | Sushi Restaurant                | Liquor Store              | BBQ Joint             | Seafood Restaurant    | Park                    | Wine Bar               |
| 1 | Anza Vista      | Café                  | Cosmetics Shop                  | Tunnel                | Burger Joint          | Southern / Soul Food Restaurant | Big Box Store             | Mexican Restaurant    | Grocery Store         | Health & Beauty Service | Liquor Store           |
| 2 | Ashbury Heights | Café                  | Bank                            | Coffee Shop           | Dog Run               | Supermarket                     | Middle Eastern Restaurant | Mexican Restaurant    | Outdoor Sculpture     | Sculpture Garden        | Massage Studio         |
| 3 | Balboa Park     | Baseball Field        | Café                            | Breakfast Spot        | Bus Station           | Flower Shop                     | BBQ Joint                 | Asian Restaurant      | Light Rail Station    | Skate Park              | Bus Stop               |
| 4 | Bayview         | Bakery                | Southern / Soul Food Restaurant | Café                  | Mexican Restaurant    | Dumpling Restaurant             | Pharmacy                  | Light Rail Station    | Home Service          | Coffee Shop             | Piercing Parlor        |

`neighborhoods_venues_sorted`

### Part 3: Methodology: K-means Clustering

I used K-means from `sklearn.cluster` package as my methodology to cluster the neighborhoods. I deleted Neighborhood column in the `sf_grouped` data frame and then used it to fit my clustering model. I discovered that a 4 clusters model would best cluster my data and provide the most insightful information.

After the aforementioned process, I was able to obtain the following result:

|     | Neighborhood       | lat       | lng         | 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 | Cluster |
|-----|--------------------|-----------|-------------|--------|-----------------------|---------------------------------|-----------------------|-----------------------|---------------------------------|---------------------------|-----------------------|-----------------------|-------------------------|------------------------|---------|
| 0   | Alamo Square       | 37.776360 | -122.434688 | 1      | Bar                   | Hotel                           | Café                  | Record Shop           | Sushi Restaurant                | Liquor Store              | BBQ Joint             | Seafood Restaurant    | Park                    |                        | V       |
| 1   | Anza Vista         | 37.780836 | -122.443149 | 1      | Café                  | Cosmetics Shop                  | Tunnel                | Burger Joint          | Southern / Soul Food Restaurant | Big Box Store             | Mexican Restaurant    | Grocery Store         | Health & Beauty Service |                        |         |
| 2   | Ashbury Heights    | 37.775599 | -122.448068 | 1      | Café                  | Bank                            | Coffee Shop           | Dog Run               | Supermarket                     | Middle Eastern Restaurant | Mexican Restaurant    | Outdoor Sculpture     | Sculpture Garden        |                        | M       |
| 3   | Balboa Park        | 37.721427 | -122.447547 | 1      | Baseball Field        | Café                            | Breakfast Spot        | Bus Station           | Flower Shop                     | BBQ Joint                 | Asian Restaurant      | Light Rail Station    | Skate Park              |                        | E       |
| 4   | Bayview            | 37.728889 | -122.392500 | 1      | Bakery                | Southern / Soul Food Restaurant | Café                  | Mexican Restaurant    | Dumpling Restaurant             | Pharmacy                  | Light Rail Station    | Home Service          | Coffee Shop             |                        |         |
| ... | ...                | ...       | ...         | ...    | ...                   | ...                             | ...                   | ...                   | ...                             | ...                       | ...                   | ...                   | ...                     | ...                    |         |
| 101 | West Portal        | 37.741141 | -122.465634 | 1      | Italian Restaurant    | Wine Bar                        | Burger Joint          | Mexican Restaurant    | Chinese Restaurant              | Coffee Shop               | Pizza Place           | Yoga Studio           | Diner                   |                        | Br      |
| 102 | Western Addition   | 37.779559 | -122.429810 | 1      | Liquor Store          | Boutique                        | Grocery Store         | Park                  | Playground                      | Seafood Restaurant        | Farmers Market        | Theater               | Historic Site           |                        | Re      |
| 103 | Westwood Highlands | 37.725726 | -122.458199 | 1      | Asian Restaurant      | Yoga Studio                     | Chinese Restaurant    | Café                  | Pharmacy                        | Coffee Shop               | Bubble Tea Shop       | Mexican Restaurant    | Grocery Store           |                        |         |
| 104 | Westwood Park      | 37.725726 | -122.458199 | 1      | Asian Restaurant      | Yoga Studio                     | Chinese Restaurant    | Café                  | Pharmacy                        | Coffee Shop               | Bubble Tea Shop       | Mexican Restaurant    | Grocery Store           |                        |         |
| 105 | Butchertown        | 37.784827 | -122.727802 | 3      | Music Venue           | Park                            | Food                  | Falafel Restaurant    | Farmers Market                  | Fast Food Restaurant      | Filipino Restaurant   | Fish Market           | Flea Market             |                        |         |

106 rows × 14 columns

sf\_merged

## Part 4: Result and Discussion

### Cluster 1: Remote Peace

Cluster 1 contains the neighborhoods that are distant from business districts, and these neighborhoods have the most common venues such as parks and trail. Therefore, I would not recommend businesses to start in this cluster, but for people who want to get away from city life, they can visit neighborhoods in this cluster.

|    | Neighborhood      | 1st Most Common Venue | 2nd Most Common Venue | 3rd Most Common Venue | 4th Most Common Venue | 5th Most Common Venue | 6th Most Common Venue | 7th Most Common Venue | 8th Most Common Venue | 9th Most Common Venue | 10th Most Common Venue |
|----|-------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|------------------------|
| 14 | Clarendon Heights | Trail                 | Park                  | Art Gallery           | Wine Bar              | Garden                | Reservoir             | Monument / Landmark   | Playground            | Bus Stop              | Eye Doctor             |
| 31 | Forest Hill       | Japanese Restaurant   | Playground            | Park                  | Hotpot Restaurant     | French Restaurant     | Football Stadium      | Food Truck            | Eye Doctor            | Falafel Restaurant    | Farmers Market         |
| 46 | Laguna Honda      | Trail                 | Light Rail Station    | Jewelry Store         | Art Gallery           | Hotpot Restaurant     | Lake                  | French Restaurant     | Park                  | Event Space           | Bus Stop               |
| 58 | Merced Manor      | Trail                 | Light Rail Station    | Jewelry Store         | Art Gallery           | Hotpot Restaurant     | Lake                  | French Restaurant     | Park                  | Event Space           | Bus Stop               |
| 61 | Miraloma Park     | Bus Stop              | Jewelry Store         | Park                  | Trail                 | Gym                   | Mountain              | Monument / Landmark   | Farmers Market        | Fast Food Restaurant  | Filipino Restaurant    |
| 96 | Twin Peaks        | Trail                 | Scenic Lookout        | Hill                  | Bus Stop              | Bus Station           | Reservoir             | Food Truck            | Filipino Restaurant   | Eye Doctor            | Fountain               |

### Cluster 2: City Business

This cluster contains the most neighborhoods. Neighborhoods within this clusters are abundant with bars, coffee shops and restaurants. I would suggest a new business to start in these neighborhoods, since they have good business vibes, but new business should conduct research to address imminent competition from other shops. For those who enjoy city life and convenience, this cluster provides the best lifestyles for them.

|     | Neighborhood       | 1st Most Common Venue | 2nd Most Common Venue           | 3rd Most Common Venue | 4th Most Common Venue | 5th Most Common Venue           | 6th Most Common Venue     | 7th Most Common Venue | 8th Most Common Venue | 9th Most Common Venue   | 10th Most Common Venue |
|-----|--------------------|-----------------------|---------------------------------|-----------------------|-----------------------|---------------------------------|---------------------------|-----------------------|-----------------------|-------------------------|------------------------|
| 0   | Alamo Square       | Bar                   | Hotel                           | Café                  | Record Shop           | Sushi Restaurant                | Liquor Store              | BBQ Joint             | Seafood Restaurant    | Park                    | Wine Bar               |
| 1   | Anza Vista         | Café                  | Cosmetics Shop                  | Tunnel                | Burger Joint          | Southern / Soul Food Restaurant | Big Box Store             | Mexican Restaurant    | Grocery Store         | Health & Beauty Service | Liquor Store           |
| 2   | Ashbury Heights    | Café                  | Bank                            | Coffee Shop           | Dog Run               | Supermarket                     | Middle Eastern Restaurant | Mexican Restaurant    | Outdoor Sculpture     | Sculpture Garden        | Massage Studio         |
| 3   | Balboa Park        | Baseball Field        | Café                            | Breakfast Spot        | Bus Station           | Flower Shop                     | BBQ Joint                 | Asian Restaurant      | Light Rail Station    | Skate Park              | Bus Stop               |
| 4   | Bayview            | Bakery                | Southern / Soul Food Restaurant | Café                  | Mexican Restaurant    | Dumpling Restaurant             | Pharmacy                  | Light Rail Station    | Home Service          | Coffee Shop             | Piercing Parlor        |
| ... | ...                | ...                   | ...                             | ...                   | ...                   | ...                             | ...                       | ...                   | ...                   | ...                     | ...                    |
| 100 | Vista del Mar      | Coffee Shop           | Café                            | Hotel                 | Park                  | Wine Bar                        | Theater                   | Cocktail Bar          | Gym                   | Sushi Restaurant        | Juice Bar              |
| 101 | West Portal        | Italian Restaurant    | Wine Bar                        | Burger Joint          | Mexican Restaurant    | Chinese Restaurant              | Coffee Shop               | Pizza Place           | Yoga Studio           | Diner                   | Bookstore              |
| 102 | Western Addition   | Liquor Store          | Boutique                        | Grocery Store         | Park                  | Playground                      | Seafood Restaurant        | Farmers Market        | Theater               | Historic Site           | German Restaurant      |
| 103 | Westwood Highlands | Asian Restaurant      | Yoga Studio                     | Chinese Restaurant    | Café                  | Pharmacy                        | Coffee Shop               | Bubble Tea Shop       | Mexican Restaurant    | Grocery Store           | Bank                   |
| 104 | Westwood Park      | Asian Restaurant      | Yoga Studio                     | Chinese Restaurant    | Café                  | Pharmacy                        | Coffee Shop               | Bubble Tea Shop       | Mexican Restaurant    | Grocery Store           | Bank                   |

91 rows × 11 columns

### Cluster 3: Remote Convenient

This cluster contains neighborhoods that are distant to business district, but it still provides some level of convenience. For those who want to live close to parks, meanwhile they do not drive to far for some essential stores, this cluster of neighborhoods could be a good option for them. New business can start here if they do not want to face intense level of competitions.

|    | Neighborhood   | 1st Most Common Venue | 2nd Most Common Venue | 3rd Most Common Venue | 4th Most Common Venue | 5th Most Common Venue | 6th Most Common Venue | 7th Most Common Venue | 8th Most Common Venue | 9th Most Common Venue | 10th Most Common Venue |
|----|----------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|------------------------|
| 38 | India Basin    | Home Service          | Park                  | Women's Store         | Flower Shop           | Falafel Restaurant    | Farmers Market        | Fast Food Restaurant  | Filipino Restaurant   | Fish Market           | Flea Market            |
| 55 | Lower Nob Hill | Park                  | Convenience Store     | Trail                 | Road                  | Dog Run               | Scenic Lookout        | Shoe Store            | Food                  | Monument / Landmark   | Szechuan Restaurant    |
| 66 | Mount Davidson | Park                  | Playground            | Bus Line              | Monument / Landmark   | Tree                  | Falafel Restaurant    | Farmers Market        | Fast Food Restaurant  | Filipino Restaurant   | Fish Market            |
| 79 | Potrero Hill   | Grocery Store         | Park                  | Hill                  | Japanese Restaurant   | Gym / Fitness Center  | Deli / Bodega         | Cosmetics Shop        | Convenience Store     | Café                  | Liquor Store           |
| 86 | Silver Terrace | Grocery Store         | Park                  | Athletics & Sports    | Dessert Shop          | Soccer Field          | Women's Store         | Flea Market           | Farmers Market        | Fast Food Restaurant  | Filipino Restaurant    |

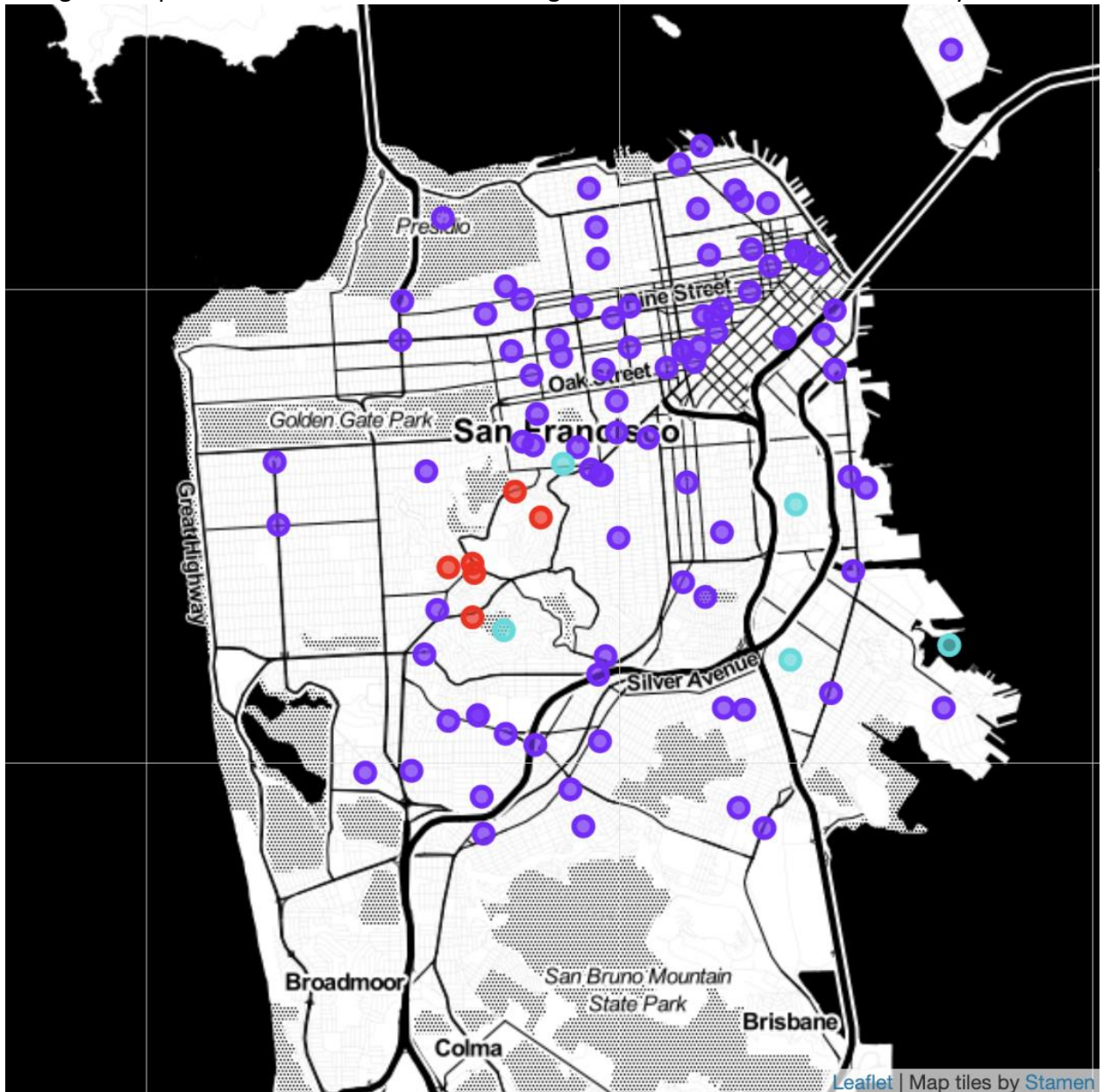
### Cluster 4: Fun Mix

This clusters of neighborhoods are made of leisure venues such as music venue, park and markets. For those who want to live close to a mixed range of leisure venues, this cluster is a good choice. New business can also consider this cluster of neighborhoods as a potential choice.

|     | Neighborhood    | 1st Most Common Venue | 2nd Most Common Venue | 3rd Most Common Venue | 4th Most Common Venue | 5th Most Common Venue | 6th Most Common Venue | 7th Most Common Venue | 8th Most Common Venue | 9th Most Common Venue | 10th Most Common Venue |
|-----|-----------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|------------------------|
| 10  | Cayuga Terrace  | Music Venue           | Park                  | Food                  | Falafel Restaurant    | Farmers Market        | Fast Food Restaurant  | Filipino Restaurant   | Fish Market           | Flea Market           | Flower Shop            |
| 59  | Midtown Terrace | Music Venue           | Park                  | Food                  | Falafel Restaurant    | Farmers Market        | Fast Food Restaurant  | Filipino Restaurant   | Fish Market           | Flea Market           | Flower Shop            |
| 60  | Mid-Market      | Music Venue           | Park                  | Food                  | Falafel Restaurant    | Farmers Market        | Fast Food Restaurant  | Filipino Restaurant   | Fish Market           | Flea Market           | Flower Shop            |
| 105 | Butchertown     | Music Venue           | Park                  | Food                  | Falafel Restaurant    | Farmers Market        | Fast Food Restaurant  | Filipino Restaurant   | Fish Market           | Flea Market           | Flower Shop            |



Following is a map the concludes the clusters of neighborhoods in the San Francisco City:



### Part 5: Conclusion

Based on the API and neighborhood data, I used the K-means clusters to divide the neighborhoods in San Francisco into four clusters, which are Remote Peace, City Business, Remote Convenient and Mix Fun. Businesses and city resident can use this report as a resource to choose the neighborhoods according to their need.