Segmenting and Clustering

Affordable Rental Project

in San Francisco

Capstone Project

Content Page

- 1 - Introduction: Business Problem
- 2 - Data
  - *2.1 - Source of Data*
  - *2.2 - Download and Explore Dataset*

1. Introduction

Business problem

- Let assume that your boss request you to have a business trip at San Francisco for 3 month with a limited budget.
- There would be a thousand rental house in San Francisco is waiting for you. There for your have a look to the listed affordable rental house published by Mayor's Office of Housing and Community Development Affordable Rental.
- You need to choose the right one which have an affordable payment and nearest the living facility that make you save more money from transportation and movement times.
- In this project, I will try to find a category of optimal affordable rental projects with better living facilities. This report will be targeted to individuals who want to travel to San Francisco for business or holiday with a limit budget.

2. Data

2.1 Source of Data

- Based on the business problem, factors that will influence visitor to choose are:
  1) Number of existing facilities around each project
  2) Type of existing facilities around each project

- Following data sources will be needed to generate the proper decision:
  1) Basic information (project name/location/ Year Affordability Began) of all the affordable rental programs, which can be get from open data website of San Francisco GOV (https://data.sfgov.org/)
  2) Number of existing facilities and their type and location in every neighborhood will be obtained using Foursquare API

2. Data

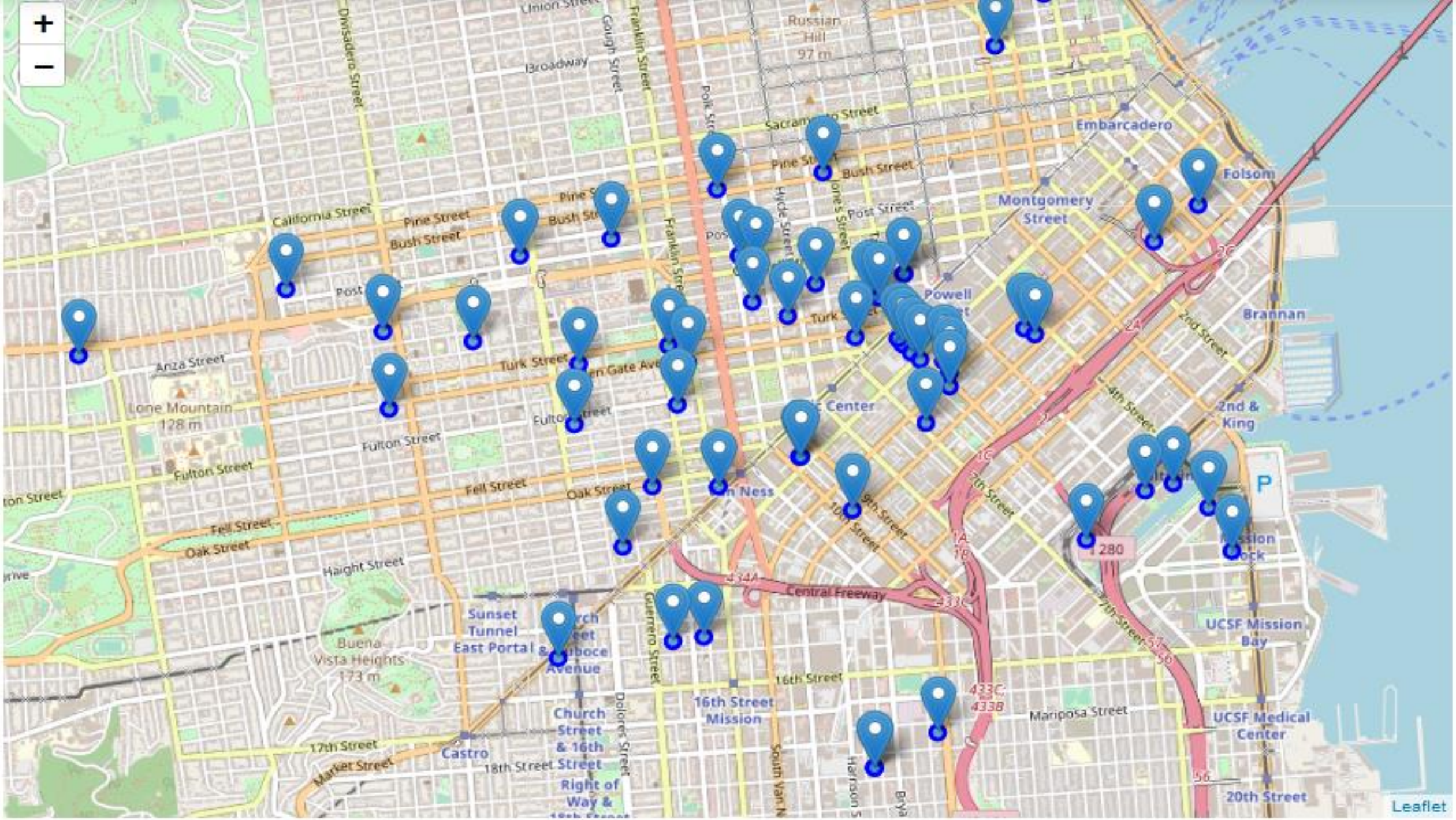
2.2 Download and Explore Dataset

- Datasets include some NaN value in Longitude and Latitude, we need to drop the rows which not clear Longitude and Latitude
- The house too old include many risk and low quality so we only choose the rental house which have constructed year after 2005
- Below is basic information data of Rental Project and Visualization using Folium to help use easy to observe the distribute on the San Francisco map

2. Data

2.2 Download and Explore Dataset

	project_name	project_address	year_building_constructed	street_name	street_number	neighborhood	planning_neighborhood	latitude	longitude
0	Alice Griffith - Phase 3B (Block 1B)	94124	2018	Arelious Walker	2500	Bayview Hunters Point	Bayview	37.719645	-122.384831
8	125 Mason Street	94102	2008	Mason	125	Tenderloin	Downtown/Civic Center	37.784805	-122.409744
10	Martin Luther King-Marcus Garvey Square Cooper...	94115	2011	Eddy	1680	Western Addition	Western Addition	37.781597	-122.434860
19	Richardson Apartments (Parcel G)	94102	2011	Fulton	365	Hayes Valley	Downtown/Civic Center	37.778492	-122.422958
20	1100 Ocean	94112	2015	Ocean	1100	West of Twin Peaks	West of Twin Peaks	37.725575	-122.454155
28	Friendship House	94103	2005	Julian	56	Mission	Mission	37.767296	-122.421402
31	2175 Market	94114	2013	Market	2175	Castro/Upper Market	Castro/Upper Market	37.766285	-122.429916
33	Hunters View Phase IIB (Block 10)	94124	2017	Middle Point	112	Bayview Hunters Point	Bayview	37.735226	-122.380759
36	Octavia Court	94102	2010	Octavia	261	Hayes Valley	Western Addition	37.774601	-122.424461
51	Edith Witt Senior Community	94103	2008	9th	66	South of Market	South of Market	37.776009	-122.415817



Russian Hill 97 m
Union Street
Broadway
Divisadero Street
Gough Street
Franklin Street
Polk Street
Sacramento Street
Pine Street
Bush Street
Post Street
Hyde Street
Jones Street
Embarcadero
Folsom
Montgomery Street
Powell
Brannan
2nd Street
California Street
Anza Street
Turk Street
Lone Mountain 128 m
Fulton Street
Fell Street
Oak Street
Haight Street
Van Ness
Center
2nd & King
4th Street
7th Street
9th Street
10th Street
Central Freeway
Sunset Tunnel East Portal
Buena Vista Heights 173 m
Guerrero Street
16th Street
16th Street Mission
Church Street & 16th Street
Dolores Street
Castro
17th Street
Market Street
18th Street
Right of Way &
South Van N
Harrison S
Mariposa Street
UCSF Mission Bay
UCSF Medical Center
20th Street
Leaflet

3. Methodology

1. The first step should be defined the business problem, we already have done that in Introduction
2. The second step should be downloaded the data and explore it, as we have done in Data
3. The Third step is to explore neighborhoods of each affordable housing projects in San Francisco with Foursquare API
4. The final step, cluster the all the affordable housing projects with K-means & visualize geographic details of each cluster

4. Analysis

4.1 Analyze Each Project

VENUES AROUND 'AFFORDABLE RENTAL PROJECTS'

	project_name	project_name Latitude	project_name Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Alice Griffith - Phase 3B (Block 1B)	37.719645	-122.384831	Double Rock	37.720106	-122.386265	Racetrack
1	Alice Griffith - Phase 3B (Block 1B)	37.719645	-122.384831	Candlestick Point Community Garden	37.720781	-122.385004	Garden
2	Alice Griffith - Phase 3B (Block 1B)	37.719645	-122.384831	Alice Griffith Community Garden	37.719263	-122.386818	Garden
3	Alice Griffith - Phase 3B (Block 1B)	37.719645	-122.384831	Gillman Playground	37.717453	-122.387888	Playground
4	Alice Griffith - Phase 3B (Block 1B)	37.719645	-122.384831	Candlestick RV Park	37.716071	-122.383111	Campground

4. Analysis

4.1 Analyze Each Project

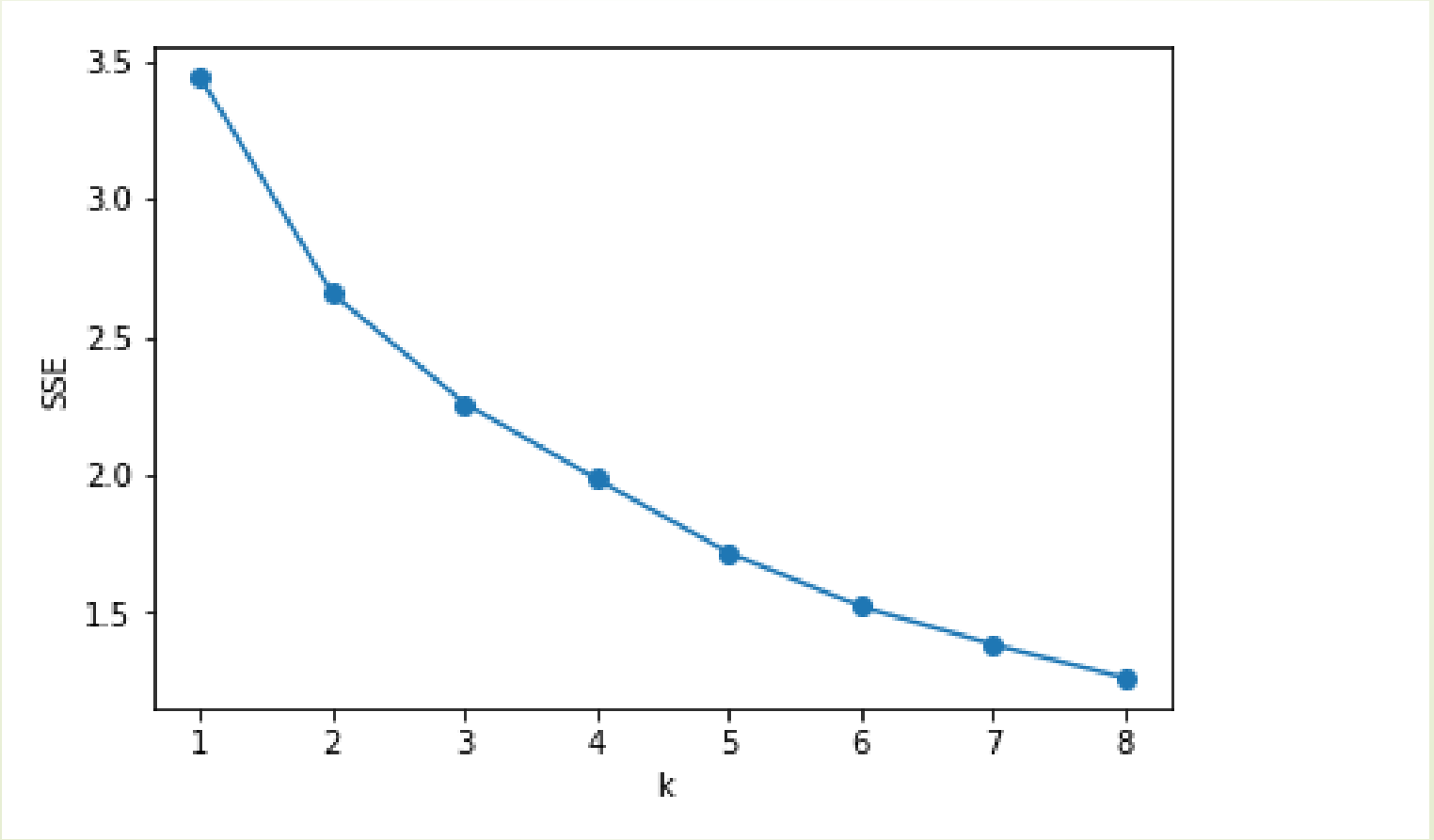
MOST COMMON VENUES AROUND EACH PROJECT

	project_name	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	10th & Mission Family Housing	Coffee Shop	Cocktail Bar	Beer Bar	Theater	Café	Art Gallery	Mexican Restaurant	Gay Bar	Street Food Gathering	Performing Arts Venue
1	1100 Ocean	Liquor Store	Poke Place	Grocery Store	Asian Restaurant	Chinese Restaurant	Pharmacy	Sandwich Place	Bubble Tea Shop	Burger Joint	Light Rail Station
2	1180 Fourth Street	Food Truck	Coffee Shop	Park	Pier	Street Food Gathering	Gym	Soccer Field	Pizza Place	Outdoor Sculpture	Sports Bar
3	125 Mason Street	Theater	Cosmetics Shop	Clothing Store	Hotel	Coffee Shop	Music Venue	Women's Store	Speakeasy	Toy / Game Store	Art Gallery
4	149 Mason Street Apartments	Theater	Cosmetics Shop	Clothing Store	Hotel	Coffee Shop	Music Venue	Women's Store	Speakeasy	Toy / Game Store	Art Gallery

4. Analysis

4.2 Cluster Project

BEST K WITH **ELBOW CRITERION**



4. Analysis

4.2 Cluster Project

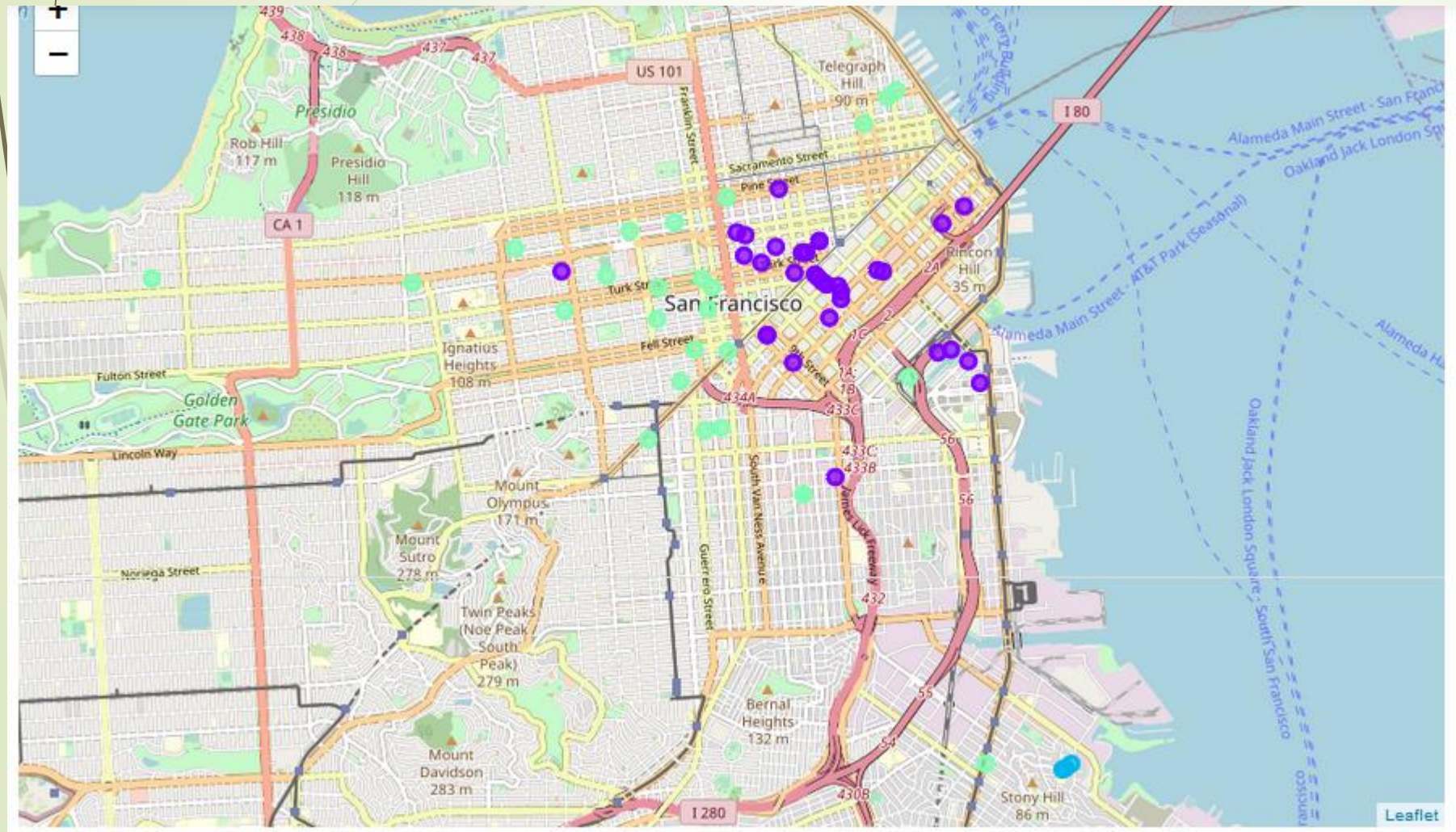
MERGED TABLE WITH CLUSTER LABELS

neighborhood	planning_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 Com V
Bayview Hunters Point	Bayview	37.719645	-122.384831	0	Garden	Playground	Racetrack	Park	Football Stadium	Campground	Furniture / Home Store	S
Tenderloin	Downtown/Civic Center	37.784805	-122.409744	1	Theater	Cosmetics Shop	Clothing Store	Hotel	Coffee Shop	Music Venue	Women's Store	Speak
Western Addition	Western Addition	37.781597	-122.434860	3	Indian Restaurant	Jazz Club	Ice Cream Shop	New American Restaurant	Tea Room	Lounge	Hobby Shop	Ra Resta
Hayes Valley	Downtown/Civic Center	37.778492	-122.422958	3	Boutique	Café	French Restaurant	Clothing Store	Sushi Restaurant	Wine Bar	Furniture / Home Store	O
West of Twin Peaks	West of Twin Peaks	37.725575	-122.454155	3	Liquor Store	Poke Place	Grocery Store	Asian Restaurant	Chinese Restaurant	Pharmacy	Sandwich Place	Bu Tea

4. Analysis

4.2 Cluster Project

Geographic details of each cluster



5. Results and Discussion

1. Cluster 1 contains 3 affordable rental projects, top 10 Most Common Venue mainly contains Playground/ Racetrack /Park /Football Stadium /Campground /Furniture / Home Store /Yoga Studio /Donut Shop /Dive Shop, if the individual would like to have a fresh air and some quiet place for relax, these areas will be quite good.
2. Cluster 2 contains 33 affordable rental projects, this cluster contains most projects, top 10 Most Common Venue mainly contains Coffee Shop /Cocktail Bar /Nightclub /Gay Bar /Food Truck /Café /Motorcycle Shop /Thai Restaurant /Art Gallery /Furniture / Home Store etc.., almost cover every aspect of our daily life, obviously, it will be very convenient to live in these areas.
3. Cluster 3 contains 3 affordable rental projects, top 10 Most Common Venue mainly contains Park /Skate Park /Bookstore /Brewery /Event Space /Donut Shop /Dumpling Restaurant /Electronics Store /English Restaurant /Ethiopian Restaurant, it seems like these are community areas, living facilities are too simple here, but it may be good for those who work here.
4. Cluster 4 contains 32 affordable rental projects, top 10 Most Common Venue mainly contains Art Coffee Shop /Gay Bar /Gym /Sushi Restaurant /Grocery Store /Asian Restaurant /Sandwich Place /Café /Pet Store, it look like this cluster is a food areas which suitable for teenage. Their can try new food here and explore new culture such as Gay Bar or some Art Coffee Shop
5. Cluster 5 contains 1 affordable rental projects, this cluster contains most projects, top 10 Most Common Venue mainly contains Playground /Mountain Park /Yoga Studio /English Restaurant /Doctor's Office /Dog Run /Donut Shop /Dumpling Restaurant /Electronics Store etc.., living facilities are too simple here, but it may be good for those who work here.

5. Conclusion

1. Purpose of this project is try to find a category of optimal affordable Rental projects with better living facilities. Target to individuals who want to travel to San Francisco for business or holiday with a limit budget. As it is widely believed that a most of residential areas should be equipped with a range of living facilities, such as restaurants/gyms/markets/hospitals etc...
2. The Foursquare location data was leveraged to compare each project to provide reliable suggestions for individuals who want to choose some place with better living facilities.
3. With unsupervised learning K-means algorithm, all the affordable Rental projects were clustered in to 5 categories, the advantages of each category was expressed to help individuals choose their best one
4. This Project simply processed the Rental affordable housing programs data, and cluster them into 5 categories based one the living facilities data, the results can only help individuals choose the place they want to rent with affordable price and better living facilities. Further analysis can be done base on these 5 clusters, which can help provide more detail information to clarify the advantages of each category.