

Family Friendly Neighborhoods in San Francisco, California.

Contents

Introduction	1
Data	1
Methodology.....	1
Results and observations	3
Cluster 0	3
Cluster 1	3
Cluster 2	4
Cluster 3	4
Conclusion.....	4

Introduction

San Francisco is a popular tourist destination. Also it is famous as a big financial center. Usually family people with children prefer to live outside the city. It will be great to find neighborhoods of San Francisco with 'good' work-life balance to increase San Francisco population diversity.

Objective: Data analysis from public sources to identify family friendly neighborhoods (more parks, sport activities, culture spots and less bars and urban development).

Data

- Neighborhoods of San Francisco (real estate site): <http://www.houseofkinoko.com/district-guide/>. San Francisco has 10 districts. Each neighborhood belongs one district.
- Venues of each neighborhoods: <https://api.foursquare.com>. Venues includes restaurants, parks, schools, culture spots, business locations which help us to define a type/style of neighborhoods.
- geopy library to find coordinates of each neighborhoods. Unfortunately, sometimes geopy finds coordinates not correctly. In this case we will find significant points for such neighborhoods using Google and use them as main point to find coordinates.

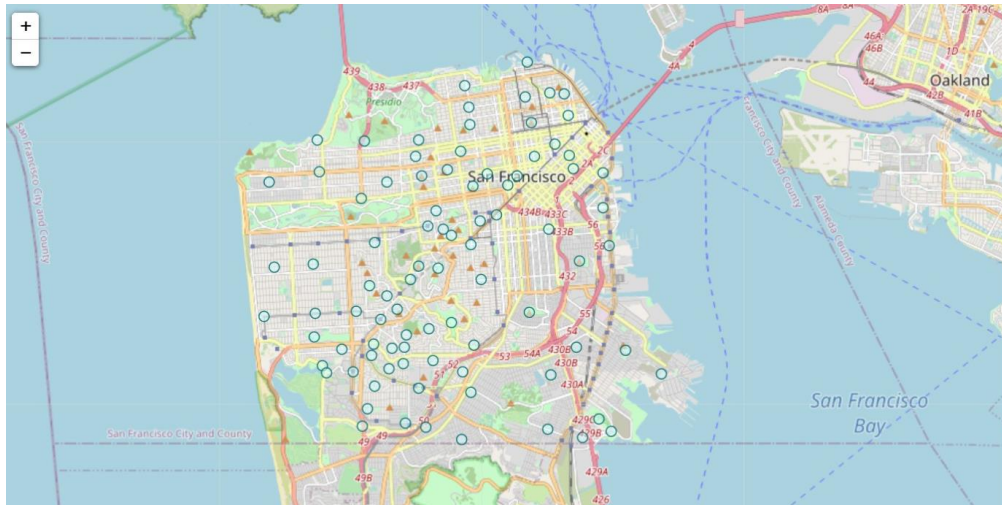
Methodology

1. Find all neighborhoods of San Francisco and their belonging to districts.

San Francisco has 10 districts and 88 neighborhoods.

District	Neighborhood
DISTRICT 1 – NORTHWEST	7
DISTRICT 10 – SOUTH EAST	12
DISTRICT 2 – CENTRAL WEST	7
DISTRICT 3 – SOUTHWEST	9
DISTRICT 4 – TWIN PEAKS WEST	15
DISTRICT 5 – CENTRAL	11
DISTRICT 6 – CENTRAL NORTH	6
DISTRICT 7 – NORTH	4
DISTRICT 8 – NORTHEAST	8
DISTRICT 8 – NORTHEAST, DISTRICT 9 – CENTRAL EAST	1
DISTRICT 9 – CENTRAL EAST	8

- Find coordinates of each neighborhood using geopy. In case geopy defines coordinates not correctly, find significant point of the neighborhood and use it to get coordinates.
- Visualization. San Francisco Map.



- Using Foursquare API get the top 100 venues around each neighborhood within a radius of 500 meters.

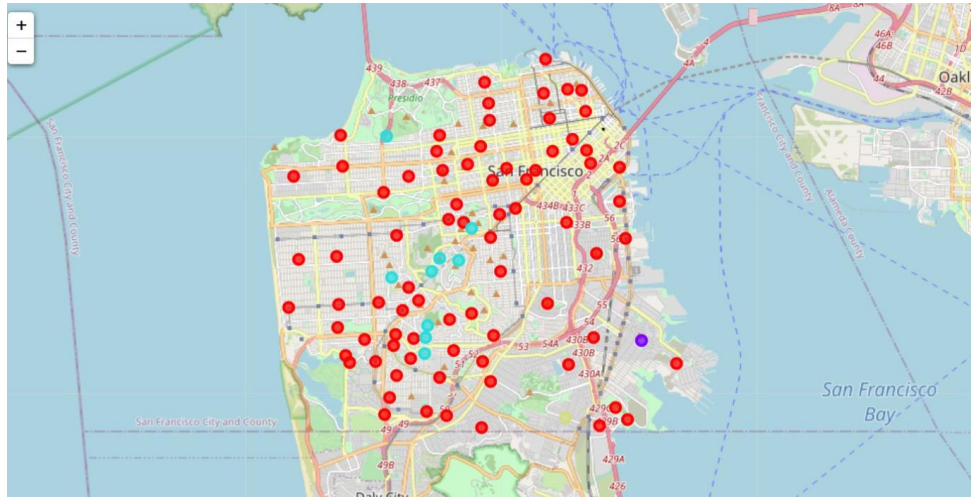
There are 3772 venues total. We will analyze neighborhoods based on categories of these venues. We have 335 unique categories. The most categories present restaurants, bars and shopping places. But we also have categories such as schools, playgrounds, parks which can be very helpful in defining family neighborhoods.

Each category will be treated as one feature. We will apply one-hot encoding for each venue in each cluster. After that we will group rows by neighborhood and by taking the mean of the frequency of occurrence of each category

5. Use unsupervised learning - KMeans Clustering with $k=4$ to divide neighborhoods on clusters.

Results and observations

As a result, we have:



Cluster 0

Total: 77 neighborhoods

Example Cluster 0 neighborhoods:

	Neighborhood	District	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	DISTRICT 6 – CENTRAL NORTH	Bar	BBQ Joint	Hotel	Seafood Restaurant	Ethiopian Restaurant	Pizza Place	Record Shop	Bakery	Food Truck	Liquor Store
1	Anza Vista	DISTRICT 6 – CENTRAL NORTH	Café	Coffee Shop	Health & Beauty Service	Bus Station	Shop & Service	Sandwich Place	Southern / Soul Food Restaurant	Big Box Store	Donut Shop	Arts & Crafts Store
2	Ashbury Heights	DISTRICT 5 – CENTRAL	Breakfast Spot	Convenience Store	Scenic Lookout	Coffee Shop	Park	Hill	Plaza	Bus Station	Snack Place	Café
3	Balboa Terrace	DISTRICT 4 – TWIN PEAKS WEST	Bubble Tea Shop	ATM	Bakery	Japanese Restaurant	Korean Restaurant	Light Rail Station	Gym	Fountain	Park	Pharmacy
		DISTRICT 10										

The most neighborhoods of Cluster 0 have a lot of bars, restaurants, shops. if you are young, you don't have a family, you work and spend time with friends, most neighborhoods of the city will suit you. Also, such environment is very friendly for tourists.

Cluster 1

Total: 1 neighborhood

	Neighborhood	District	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
4	Bayview	DISTRICT 10 – SOUTH EAST	Health & Beauty Service									

This cluster has only one neighborhood Bayview. Bayview is the most isolated neighborhood. It's very industrial.

Cluster 2

Total: 9 neighborhoods

	Neighborhood	District	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
11	Clarendon Heights	DISTRICT 5 – CENTRAL	Trail	Playground	Monument / Landmark	Reservoir						
13	Corona Heights	DISTRICT 5 – CENTRAL	Park	Trail	Museum	Scenic Lookout	Tennis Court	Hill	Monument / Landmark	Grocery Store	Dog Run	Shoe Store
25	Golden Gate Heights	DISTRICT 2 – CENTRAL WEST	Trail	Park	Playground	Tennis Court	Scenic Lookout					
37	Lake Mountain	DISTRICT 1 – NORTHWEST	Trail	Intersection	Dog Run	Tennis Court	Park	Tunnel	Art Gallery	Scenic Lookout	Bus Station	Playground
46	Midtown Terrace	DISTRICT 4 – TWIN PEAKS WEST	Trail	Light Rail Station	Bus Stop	Playground	Lake	Monument / Landmark				
72	Sherwood Forest	DISTRICT 4 – TWIN PEAKS WEST	Trail	Bus Stop	Tree	Park	Monument / Landmark					
80	Twin Peaks	DISTRICT 5 – CENTRAL	Trail	Scenic Lookout	Bus Station	Reservoir	Garden	Tailor Shop	Hill	Bus Stop		
85	Westwood Highlands	DISTRICT 4 – TWIN PEAKS WEST	Bus Line	Scenic Lookout								
86	Westwood Park	DISTRICT 4 – TWIN PEAKS WEST	Gym / Fitness Center	Park	Bus Line	Scenic Lookout						

The neighborhoods of this cluster have a lot of parks, trails, sport activities and good transportation.

Cluster 3

Only 1 neighborhood.

	Neighborhood	District	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
82	Visitation Valley	DISTRICT 10 – SOUTH EAST	Vietnamese Restaurant									

This cluster is pretty similar to cluster 1. The neighborhood is isolated but it's more residential than commercial.

Conclusion

As we can see, the cluster 2 is the most beneficial for families. It contains many parks and sports spots. Most of the neighborhoods in this cluster is located in the city center (the most neighborhoods belong DISTRICT 4 and DISTRICT 5), in close proximity to cultural sites and restaurants/shopping.

Unfortunately, the data does not have enough information about schools and childcare. For future improvements we can use additional public sources (for example, <https://www.cde.ca.gov/ds/si/ds/pubschls.asp>)

Finally, the neighborhoods of the cluster 2 is a good combination for living with children in San Francisco.