

Exploring San Francisco Neighborhoods

by
Sree Lakshmi Gudreddi

Business Problem

Which neighborhood should I choose to live in San Francisco?

How to choose?

Based on the life style i.e activities and interests.

Target Audience

1. Families with Children and Dog
2. Young and Urban Professionals
3. Artists

Data Description

Data Needed

San Francisco Neighborhood Data

- Borough Name

- Neighborhood Name

- Latitude

- Longitude

Information about Venues

- Venue Name

- Category Name

- Latitude

- Longitude

Data Sources

Wikipedia Page

https://en.wikipedia.org/wiki/List_of_neighborhoods_in_San_Francisco

Foursquare API

Data Obtained

5 Boroughs

60 Neighborhoods

246 Venue Categories

1155 Venues

Data Set

Families with Children and Dog

'Park', 'Garden', 'Trail', 'Library', 'Dance Studio', 'Music School', 'Skating Rink', 'Athletics & Sports', 'Soccer Field', 'Mini Golf', 'Bookstore', 'Church', 'Veterinarian', 'Pet Store', 'Shopping Mall', 'Supermarket'

Young and Urban Professionals

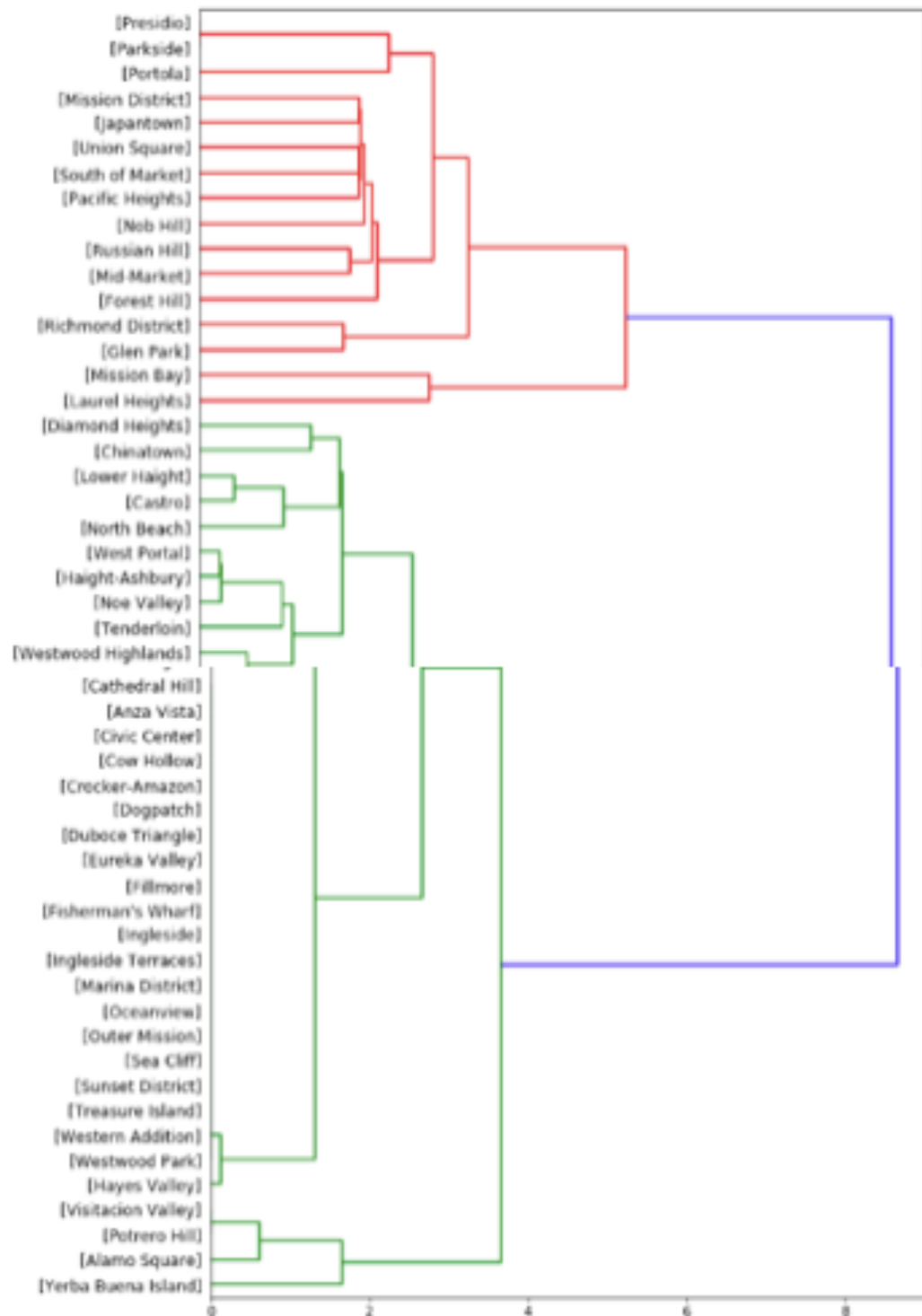
'Light Rail Station', 'Bus Stop', 'Gym / Fitness Center', 'Gym', 'Pilates Studio', 'Yoga Studio', 'ATM', 'Flower Shop', 'Boutique', 'Electronics Store', 'Paper / Office Supplies Store', 'Comic Shop', 'Beer Bar', 'Cocktail Bar', 'Karaoke Bar', 'Jazz Club', 'Nightclub'

Artists

'History Museum', 'Art Museum', 'Art Gallery', 'Public Art', 'Sculpture Garden', 'Antique Shop', 'Monument / Landmark', 'Outdoor Sculpture', 'Historic Site', 'Harbor / Marina', 'Beach', 'Mountain', 'Hobby Shop', 'Arts & Crafts Store'

Data Analysis

Hierarchical Clustering: Families with Children and Dog



Clustered based on proximity to Venues

Leafs denote the neighborhoods

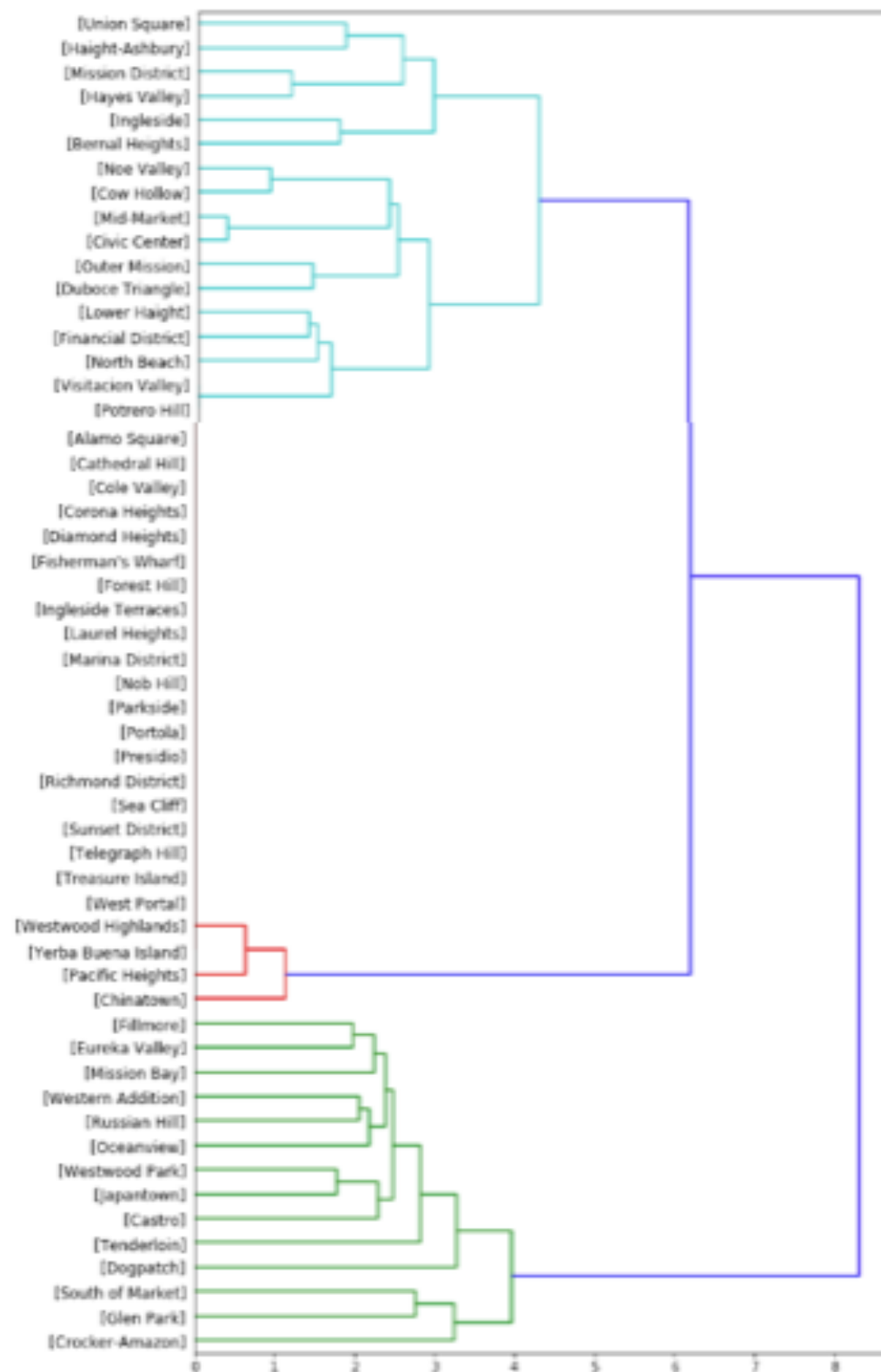
Branches denote the distance to venues

Divided into 4 clusters

Results

Target Audience	Cluster 1	Cluster 2	Cluster 3	Cluster 4
Families with Children and Dogs	Mission District Hayes Valley Haight-Ashbury Outer Mission Dogpatch Ingleside Bernal Heights Noe Valley Cow Valley Castro Fillmore Eureka Valley Russian Hill Mission Bay Oceanview Crocker-Amazon Tenderloin Union Square South of Market Japantown Glen Park	North Beach Financial District Visitation Valley Potrero Hill Lower Haight Duboce Triangle Mid-Market Civic Center	Pacific Heights Chinatown Alamo Square	Western Addition Hayes Valley

Hierarchical Clustering: Young and Urban Professionals



Clustered based on proximity to Venues

Leafs denote the neighborhoods

Branches denote the distance to venues

Divided into 3 clusters

Hierarchical Clustering: Artists



Clustered based on proximity to Venues

Leafs denote the neighborhoods

Branches denote the distance to the venues

Divided into 3 clusters with one outlier

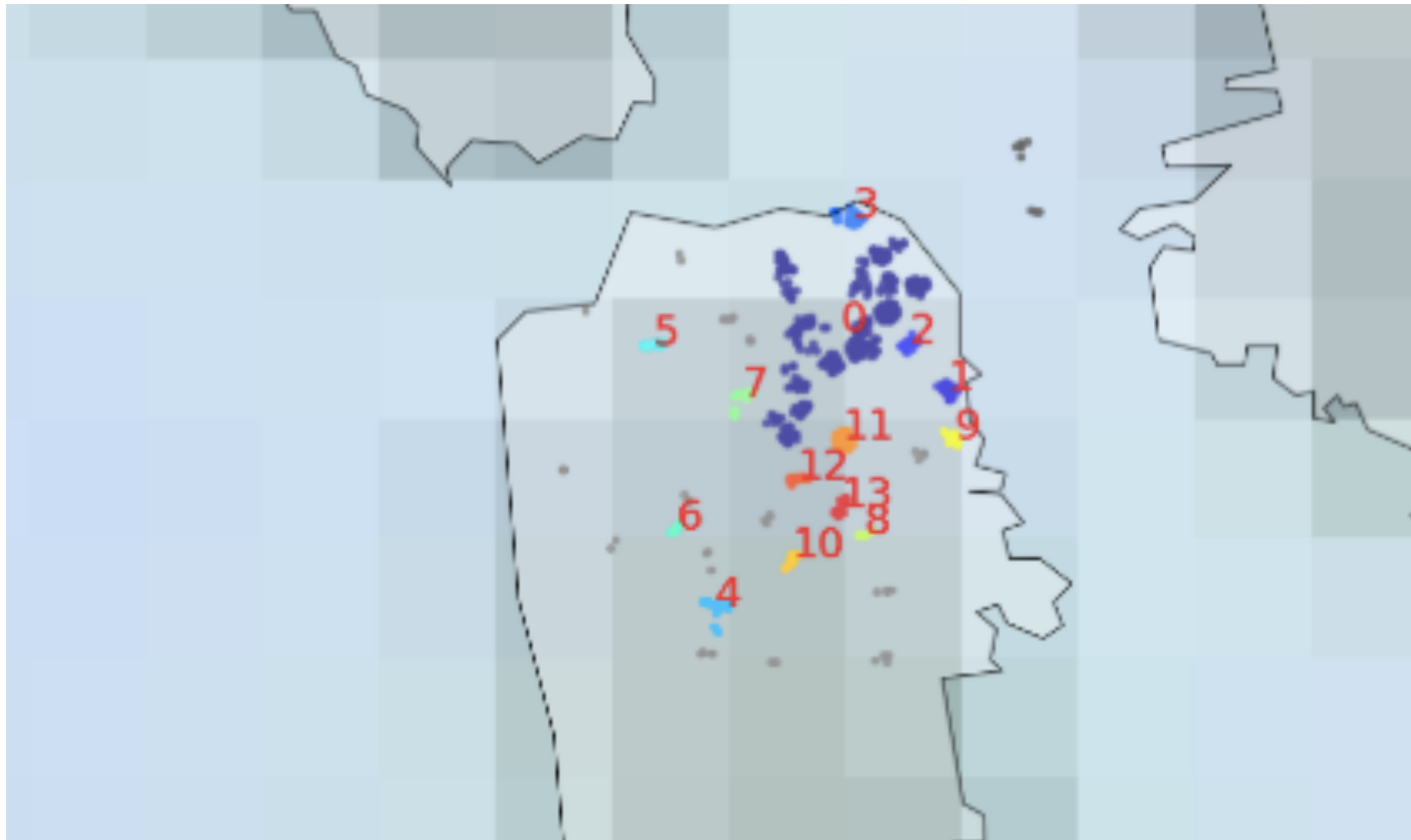
Results

Target Audience	Cluster 1	Cluster 2	Cluster 3	Cluster 4
Young and Urban professionals	Union Square Haight-Ashbury Mission District Hayes Valley Ingleside Bernal Heights Noe Valley Cow Hollow Mid-Market Civic Center Outer Mission Duboce Triangle Lower Haight Financial District North Beach Visitacion Valley	Westwood Highlands Yerba Buena Island Pacific Heights Chinatown	Fillmore Eureka Valley Mission Bay Western Addition Russian Hill Oceanview Westwood Park Japantown Castro Tenderloin Dogpatch South of Market Glen Park Crocker-Amazon	
Artists	Ingleside Terrace Civic Center Nob Hill Sea Cliff Telegraph Hill Marina District Visitacion Valley Mid-Market Corona Heights Richmond District North Beach Mission District South of Market China Town Eureka Valley	Castro Alamo Square Duboce Triangle Noe Valley Cow Hollow	Pacific Heights Dogpatch Haight-Ashbury Fisherman's Wharf	

DBSCAN Clustering: Venue Location

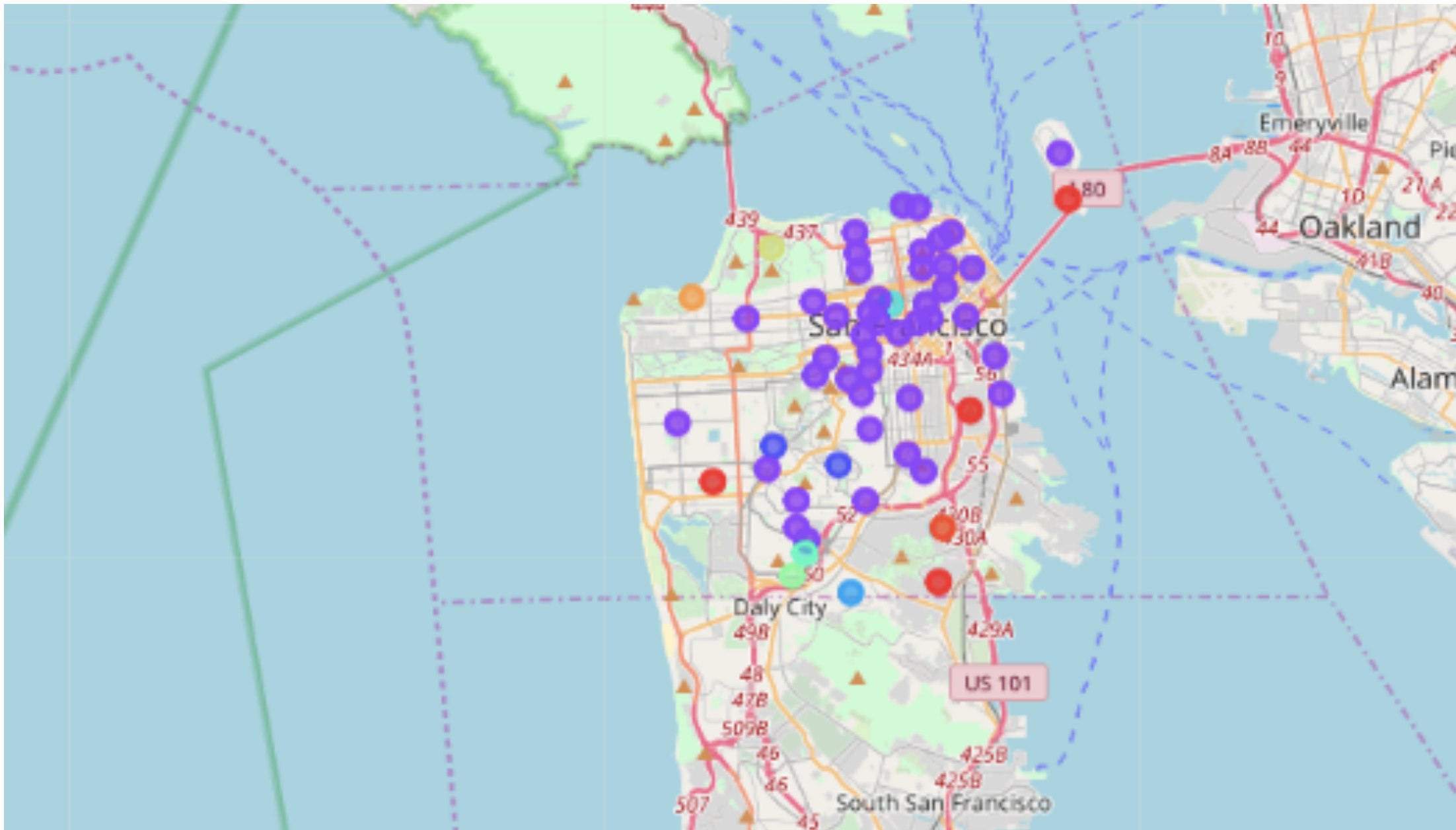
Clusters are formed based on density of Venues located within a set radius

Divided into 14 clusters with 1 outlier



K-Means Clustering: Venue Categories

Number of clusters $k = 10$



K-Means Cluster Information

Cluster ID	Neighborhood	Categories
0	4	Island/Park
1	43	Coffee shop/ Park
2	2	Trail
3	1	Light Trail Station
4	1	Cupcake Shop
5	1	Playground
6	1	Park
7	1	Stables
8	1	Beach
9	1	Bus Station

Number of clusters k=10

Observations

Geocodes obtained for few neighborhoods were incorrect

Clusters are not of the same size which shows that the venues are concentrated in few neighborhoods

Same can be shown when finding k using the elbow method

Neighborhoods are grouped into one large closet and other small clusters.

Clustering helps target audience to identify the neighborhoods with closer proximity to their interested venues.

Conclusion

Neighborhoods are identified based on the following:

- Venue Proximity

 - Focused on each group of target audience

 - Used Hierarchical clustering method

- Venue Density

 - Based on venue's geo spacial location

 - Used DBSCAN clustering method

- Venue Category

 - Based on the type of venue categories

 - Used K-Means clustering method