

The Battle of Neighborhoods

IBM_Data_Science_Certificate_Capstone



CONTENTS

1

INTRODUCTION

2

DATA DESCRIPTION

3

METHODOLOGY

4

RESULTS

5

DISCUSSION

6

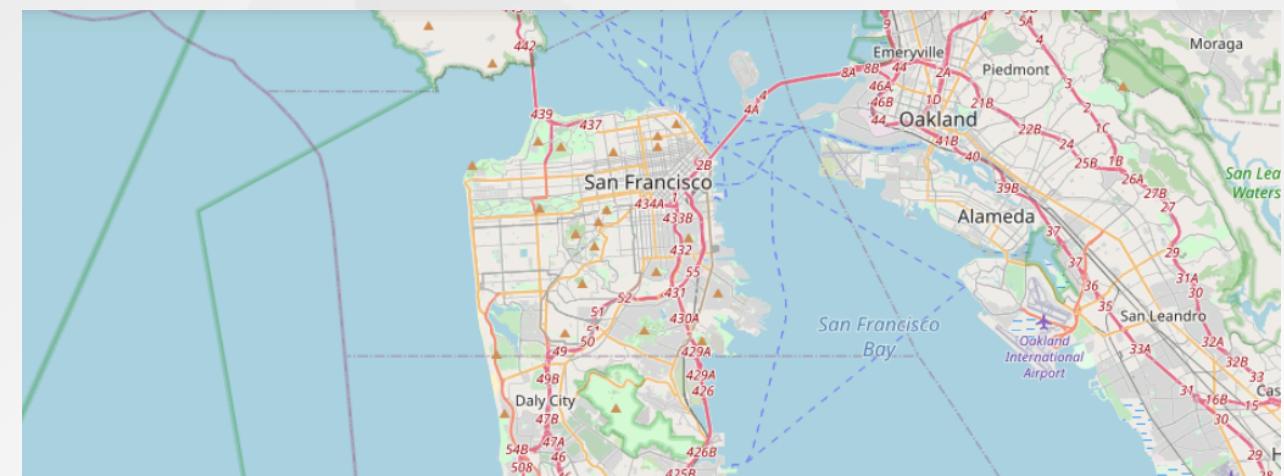
CONCLUSION



01

INTRODUCTION

San Francisco





02

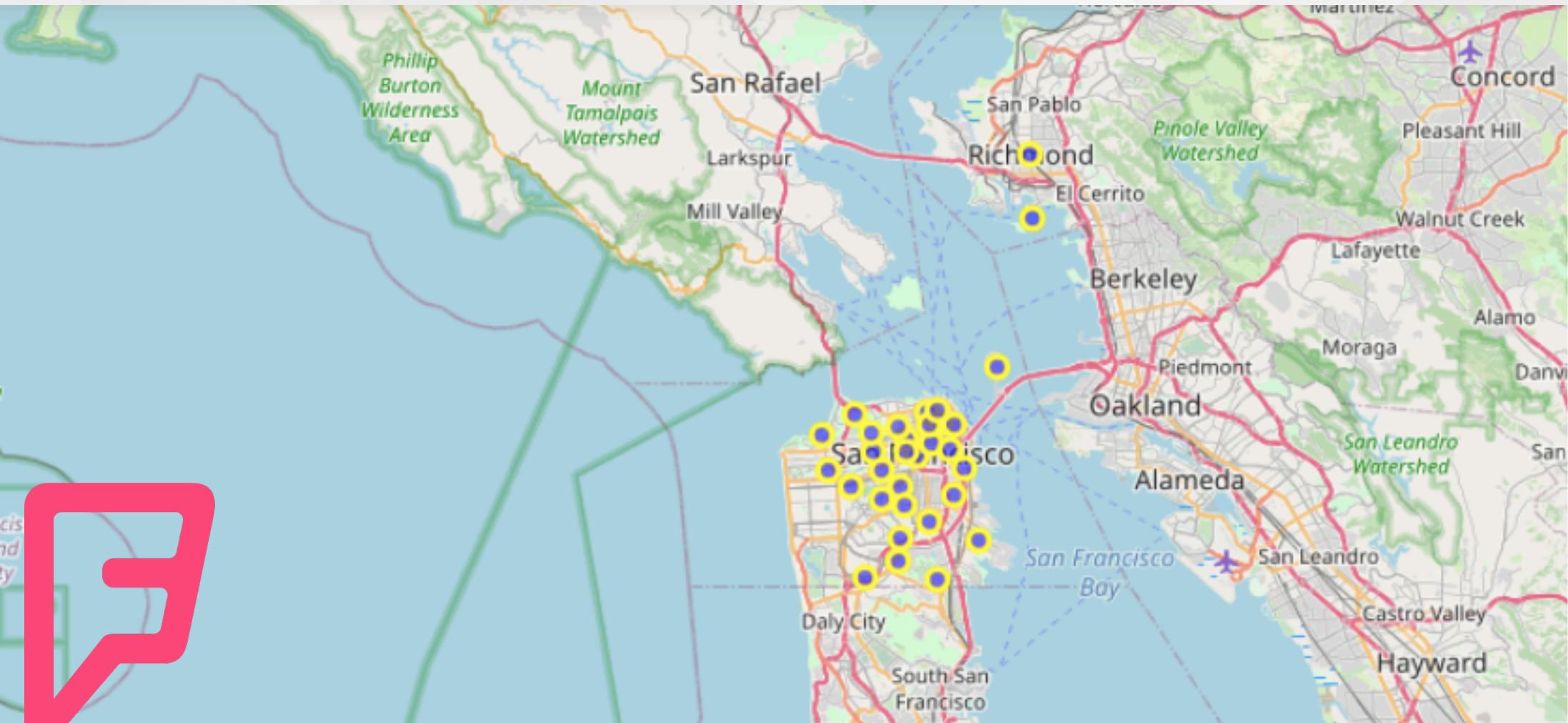
DATA DESCRIPTION



DATA



FOURSQUARE



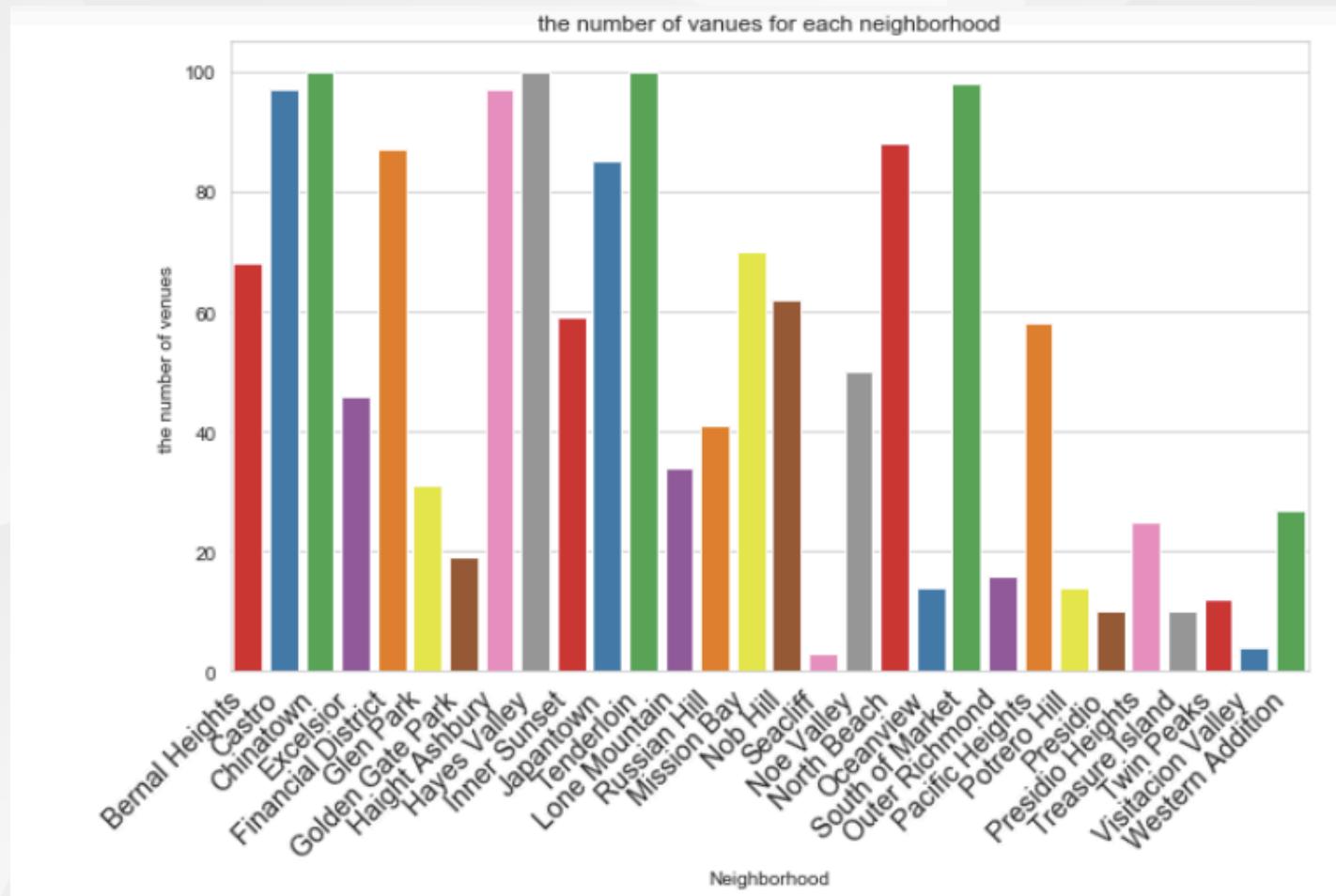
1. Bayview Hunters
2. Point Bernal
3. Heights Castro
4. Chinatown
5. Excelsior Financial
6. District Glen Park
7. Inner Richmond
8. Golden Gate Park
9. Haight Ashbury
10. Hayes Valley
11. Inner Sunset
12. Japantown
13. Tenderloin Lone
14. Mountain Russian
15. Hill Mission Bay
16. Nob Hill
17. Seacliff Noe Valley
18. North Beach
19. Oceanview
20. South of Market
21. Outer Richmond
22. Pacific Heights
23. Potrero
24. Hill Presidio
25. Presidio
26. Heights Treasure
27. Island Twin Peaks
28. Visitacion Valley
29. Western Addition



03

METHODOLOGY

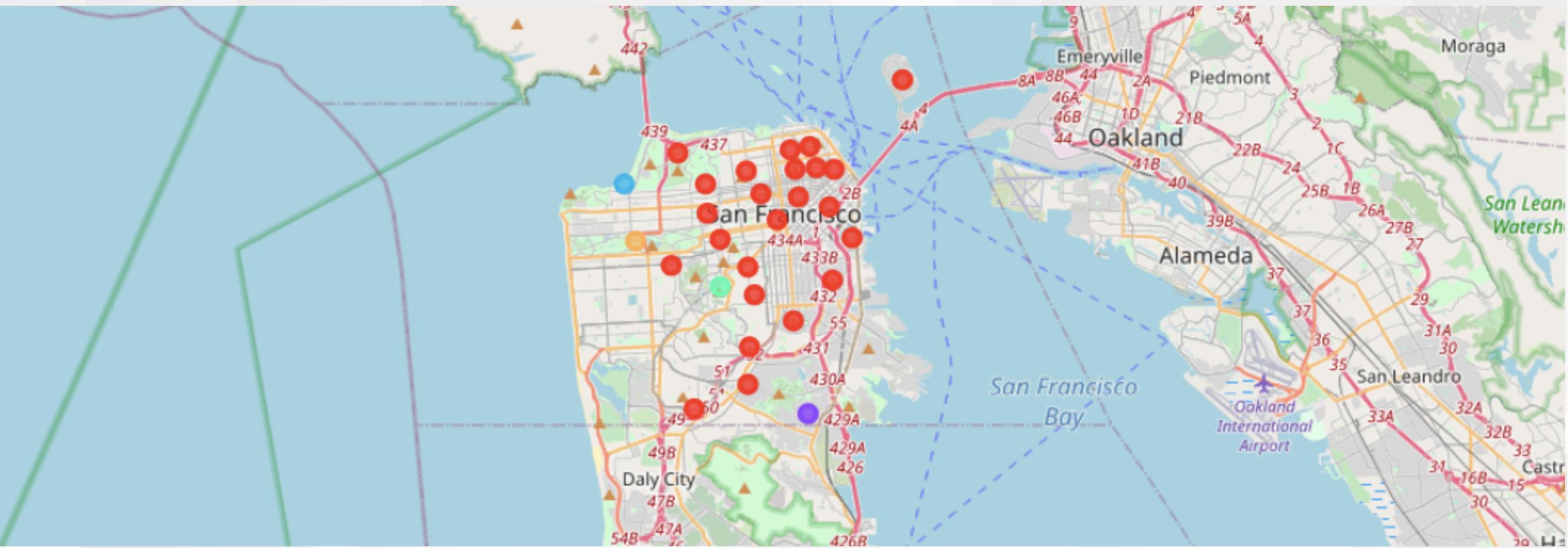
Use Foursquare API to filter out the venues in the radius of 500 meters (no more than 100) for each neighborhood.



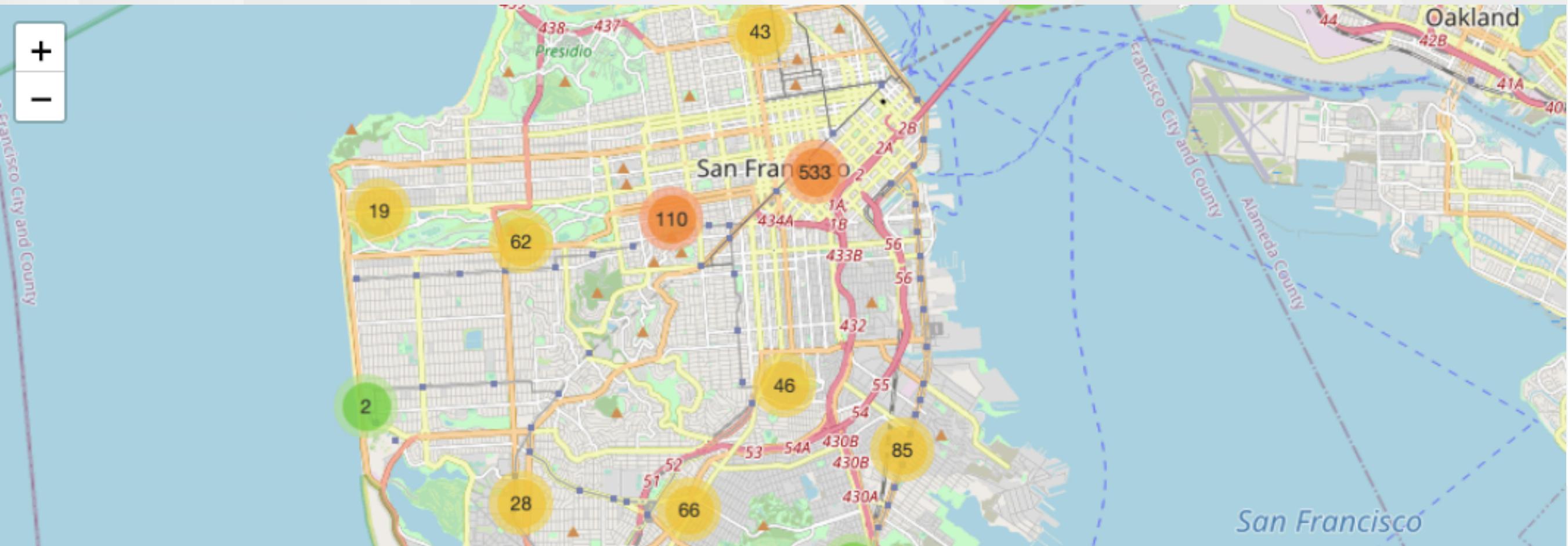
Analyze each neighborhood through looking at the categories of venues.

	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	Bernal Heights	Coffee Shop	Mexican Restaurant	Italian Restaurant	Playground	Gourmet Shop	Grocery Store	Food Truck	Cocktail Bar	Trail	Bakery
1	Castro	Gay Bar	Coffee Shop	Thai Restaurant	Indian Restaurant	Pet Store	Yoga Studio	New American Restaurant	Mediterranean Restaurant	Deli / Bodega	Convenience Store
2	Chinatown	Italian Restaurant	Chinese Restaurant	Coffee Shop	New American Restaurant	Cocktail Bar	Bakery	Hotel	Tea Room	Bubble Tea Shop	Szechuan Restaurant
3	Excelsior	Mexican Restaurant	Bakery	Vietnamese Restaurant	Chinese Restaurant	Deli / Bodega	Pizza Place	Latin American Restaurant	Pharmacy	Sandwich Place	Bank
4	Financial District	Coffee Shop	Gym	Food Truck	Men's Store	Café	Gym / Fitness Center	Restaurant	Park	New American Restaurant	Japanese Restaurant

Use K-means cluster machine learning algorithm to cluster the neighborhoods



Use folium to visualize the number of incidents for different areas





04

RESULTS

We clustered the neighborhoods into five different clusters. The biggest cluster is cluster 1 (red). In this cluster, we can see the most common venues are restaurants, bars, gyms etc.

For the other clusters, there are only one or two neighborhoods fitting into. Those neighborhoods are first far away from the downtown SF, and have many parks, trails, or fields.

Let's look at the incidents in SF area. We can see that the closer to SF downtown, the more incidents we can see. The tourist spots such as the Golden Gate Park, the beaches are relatively safer. The southeast area out of SF downtown have fewer incidents as well.



05

DISCUSSION

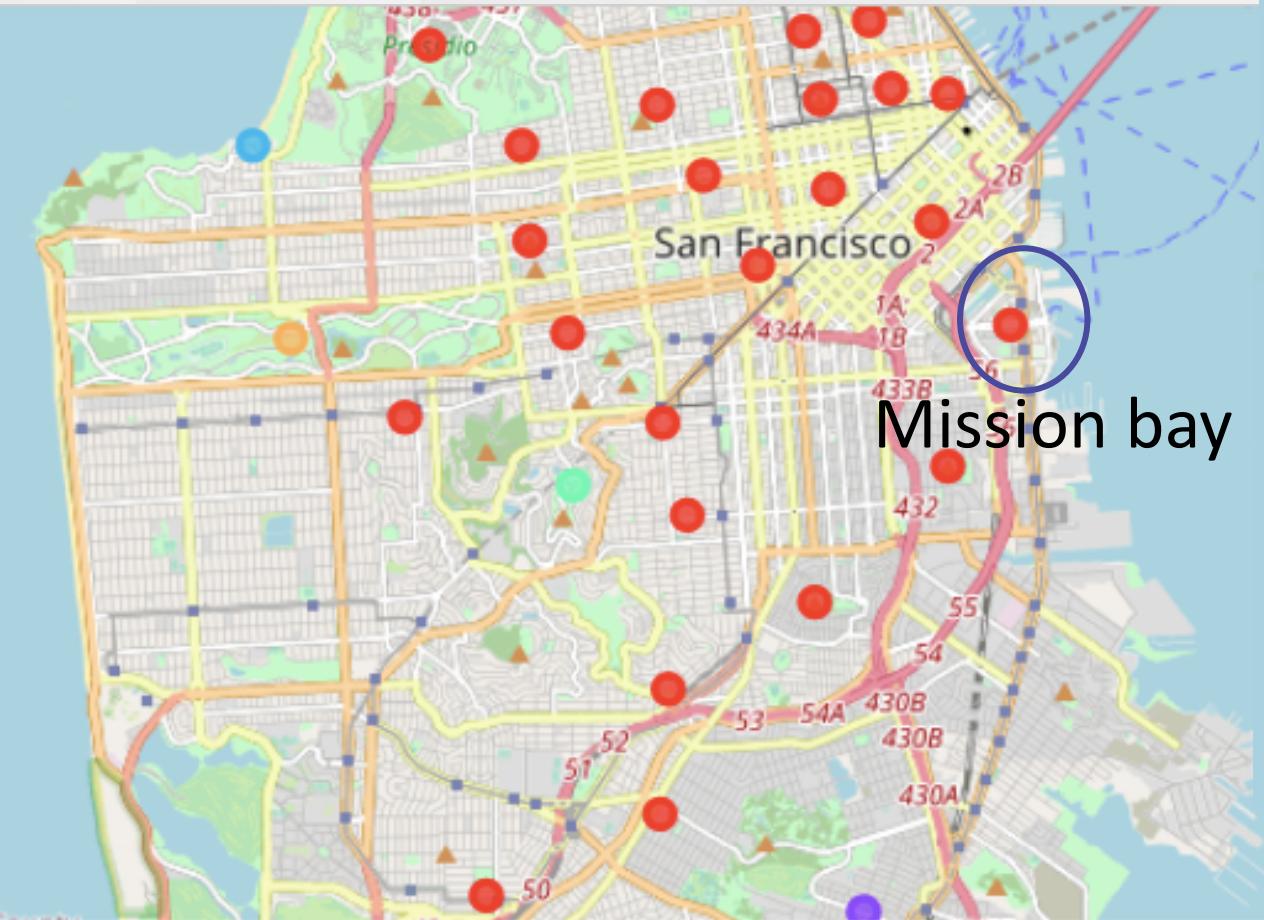


We can see from the clustering map, the majority of the neighborhoods are good for living because we can find a lot of venues such as coffee shops, various kinds of restaurants, and gyms and yoga studios. But as far as the safety's concern, the southeast of the sf downtown, such as the neighborhood Mission bay, could be a better choice for living.



06

CONCLUSION



Mission bay

