

Data Science Capstone: The Best and safest Neighborhood in San Francisco for Opening a Coffee Shop



Introduction

San Francisco is a cultural, commercial, and financial center in Northern California. San Francisco is the 16th most populous city in the United States, and the fourth most populous in California, with 881,549 residents as of 2019.

With its diverse culture , comes diverse alternatives business relationship with food items,beverages and cafe in safest Neighborhoods.

So as part of this project , we will list and visualize all options to create "businnes altenative safest" relations with cafe in San Francisco (CA).








Objective

As a business idea, the goal is to open a coffee shop in a safe and central place in san francisco. Based on data science and data provided by the police department along with current business locations, optimal recommendations should be indicated to a set of investors before opening stores.



Objectives

-  - Which all areas have more number of coffee shops or café ?
-  - Which all areas have less number of coffeeshops?
-  - Which districts have the neighborhoods safest ?
-  - Which districts Reveals The Most Unsafe Districts ?
-  - What places are have best place to stay if I want a coffee ?



Data

1- For Crime Rate in San Francisco Section [San Francisco Crime Data]

<https://data.sfgov.org/Public-Safety/Police-Department-Incident-Reports-2018-to-Present/wg3w-h783>

```
pd.read_csv('Police_Department_Incident_Reports__2018_to_Present2.csv')
```

2- For Data Set with locations of businesses [San Francisco Registered Business Data]

```
with urllib.request.urlopen("https://data.sfgov.org/resource/wg3w-h783.json") as  
url:francisco = json.loads(url.read().decode())
```



Approach



- A) We can pull two datasets:
 - a) list of every business registered in San Francisco from the last couple of decades from the data SF website.
 - b) a list of date crime in San francisco (since 2018 - present)
- B) Using FourSquare API to find all venues for each safe neighborhood.
- C) Filter out all venues that are nearby by locality.
- D) Using aggregative rating for each Coffee Shop to find the best places.
- E) Visualize the Ranking of neighborhoods using folium library(python)
- f) Apply Kmeans to find best candidates to open a new coffee shop.



Results about Crime in San Francisco



San Francisco Map

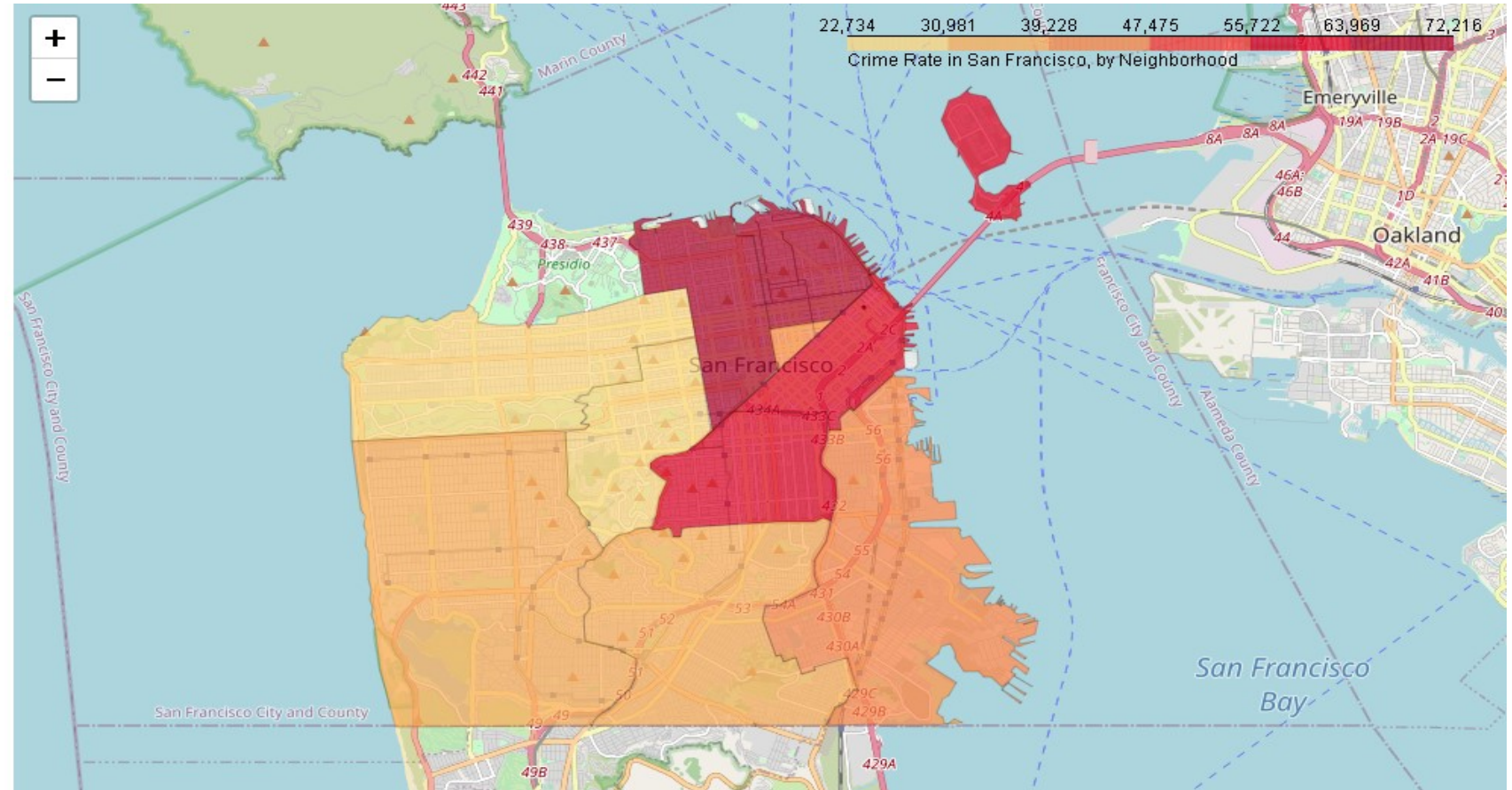


The above Map is is for the next references

Information about Crime

Table and graphic

	Neighborhood	Count
0	CENTRAL	71731
1	NORTHERN	65588
4	MISSION	62049
3	SOUTHERN	57937
5	TENDERLOIN	45731
9	BAYVIEW	42197
8	INGLESIDE	36626
7	TARAVAL	33280
6	RICHMOND	30103
2	PARK	23219

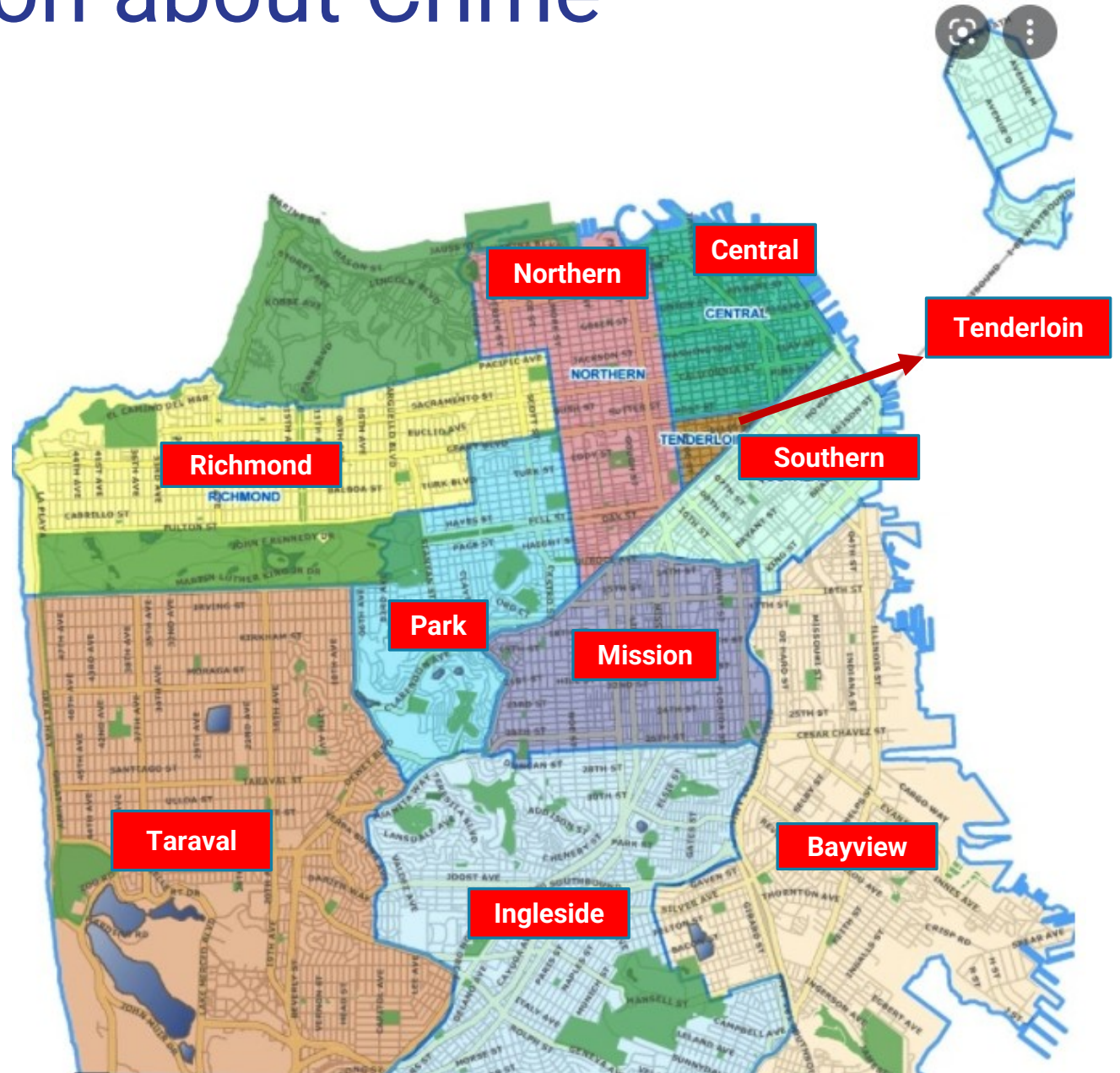


The above table has a classification between districts and neighborhoods. We can see the least safe areas according to the statistics.

superimposing the locations of the crimes onto the map, Notice how the most dangerous areas are Central, Northern, Mission and Southern. likewise, the safest are Richmond and Park. Consequently, in these last two areas we will concentrate our exploration.

Information about Crime

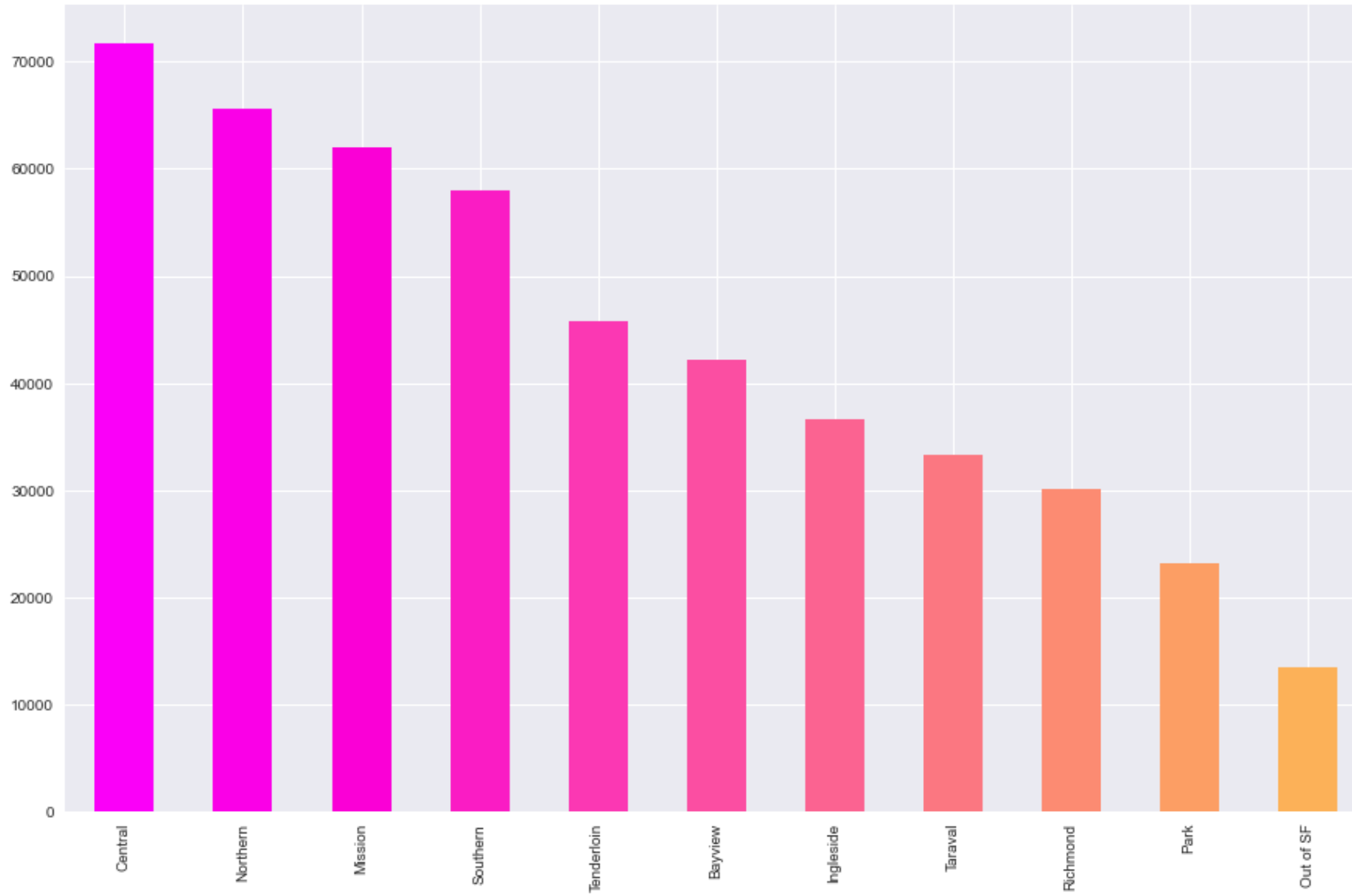
	Neighborhood	Count
0	CENTRAL	71731
1	NORTHERN	65588
4	MISSION	62049
3	SOUTHERN	57937
5	TENDERLOIN	45731
9	BAYVIEW	42197
8	INGLESIDE	36626
7	TARAVAL	33280
6	RICHMOND	30103
2	PARK	23219



The table has a more distinctive classification about safest and dangerous areas.

Distribution Crime

Districts vs Incidents



We can check the Districts safest are: Park and RichMond, because their crime rates are lower than the rest. in deed, if the area is safe, the business is more likely to be sAccording to the incident report from 2018 to date, the safest districts and neighborhoods are Park and Richmond (with rates 23219 and 30103). For its part. the most insecure

Conclusion about Crime

According to the incident report from 2018 to date, the safest districts and neighborhoods are **Park and Richmond** (with rates 23219 and 30103). For its part, the most insecure are Central an Northern with rates 71731 and 65588.

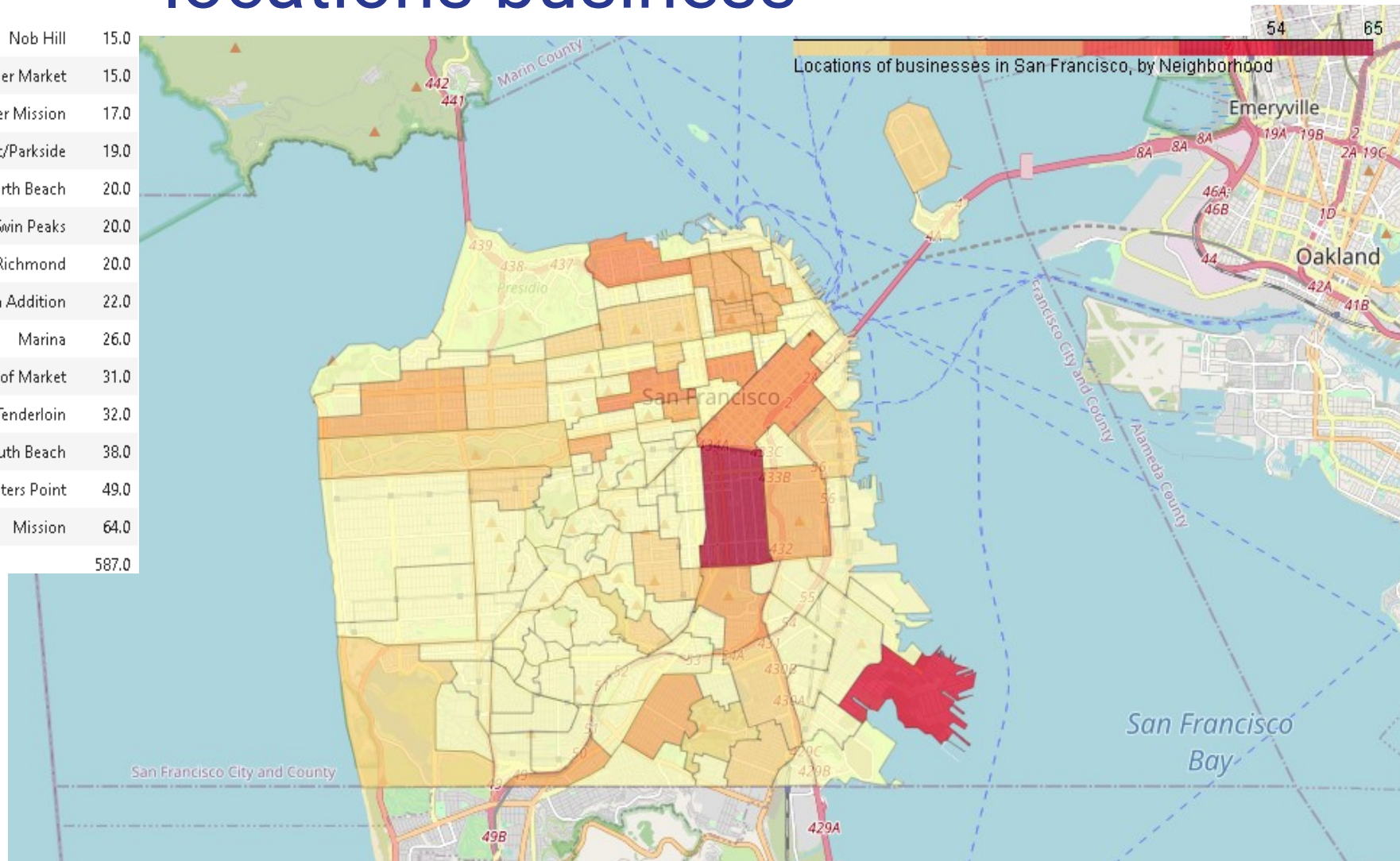
In fact, with these statistics the ideal places to have a "coffee shop" are: Park and Richmond for safety aspects

Results about locations business in Neighborhoods



Map about counting Neighborhoods and locations business

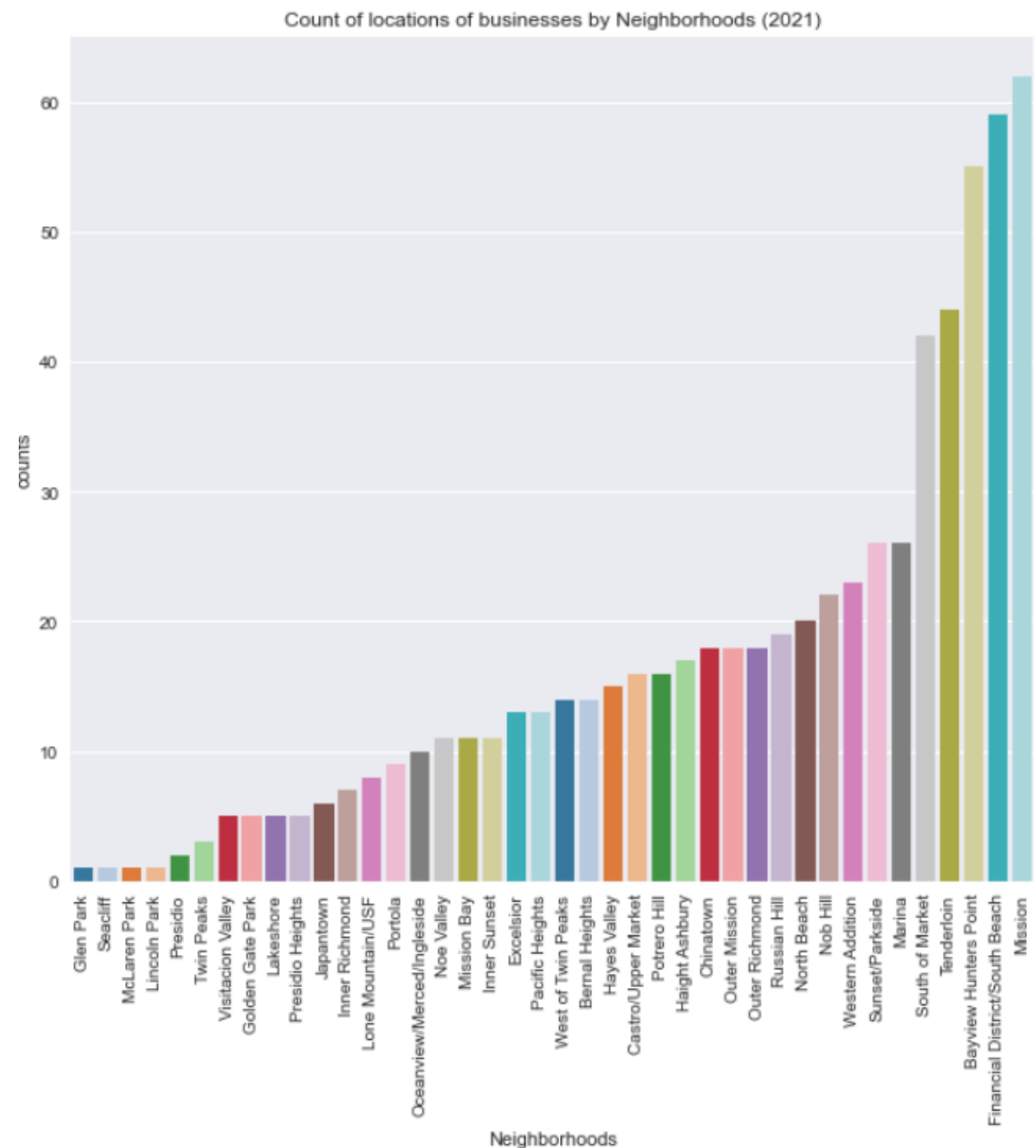
Neighborhood		
29	Presidi	
35	Treasure Island	
14	Lincoln Park	
17	McLaren Park	
36	Twin Peaks	
6	Glen Park	
23	Oceanview/Merced/Ingleside	
30	Presidio Heights	
27	Portola	
37	Visitacion Valley	
12	Japantown	
7	Golden Gate Park	
13	Lakeshore	
15	Lone Mountain/USF	
26	Pacific Heights	
19	Mission Bay	
11	Inner Sunset	10.0
10	Inner Richmond	11.0
21	Noe Valley	11.0
9	Hayes Valley	12.0
31	Russian Hill	13.0
1	Bernal Heights	13.0
8	Haight Ashbury	14.0
28	Potrero Hill	14.0
4	Excelsior	14.0
3	Chinatown	14.0
20	Nob Hill	15.0
2	Castro/Upper Market	15.0
24	Outer Mission	17.0
33	Sunset/Parkside	19.0
22	North Beach	20.0
38	West of Twin Peaks	20.0
25	Outer Richmond	20.0
39	Western Addition	22.0
16	Marina	26.0
32	South of Market	31.0
34	Tenderloin	32.0
5	Financial District/South Beach	38.0
0	Bayview Hunters Point	49.0
18	Mission	64.0
Total		587.0



The above table has counting about Neighborhoods and business locations in san francisco
More red means more quantity, while yellow means less



Count of locations of businesses and neighborhoods



The above table has counting about Neighborhoods and business locations in san francisco

Conclusion about locations of business and neighborhoods

Despite there being more businesses in Mission, financial District and Bayview. we can appreciate that they are very unsafe areas. consequently, we are going to approach the data from security aspects.

Analyzing on the notebook Part 8.6 and Part 7.6 takes the Safest Neighborhoods by Park and Richmond Districts.

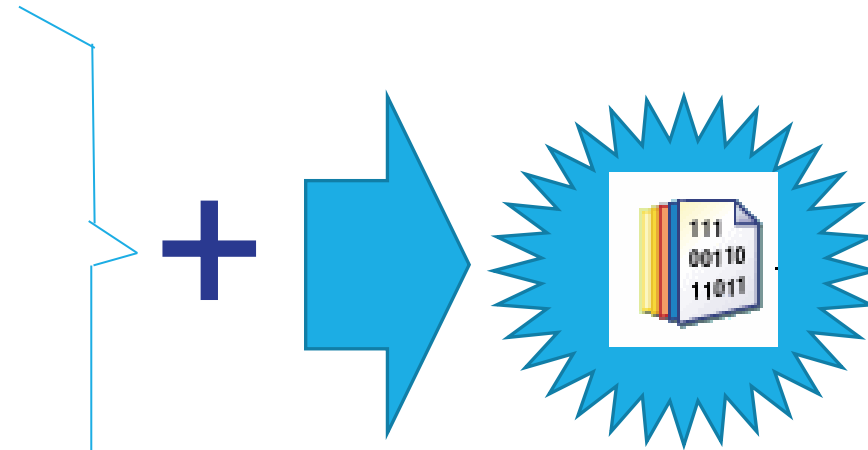
Conclusion about Final Dataset

Followed, we will only show results from a single dataset with Park and Richmond data

1- Dataset For Crime Rate in San Francisco



2-Data Set with locations of businesses
[San Francisco Registered Business Data]



Dataset Safest locations of
business in San Francisco

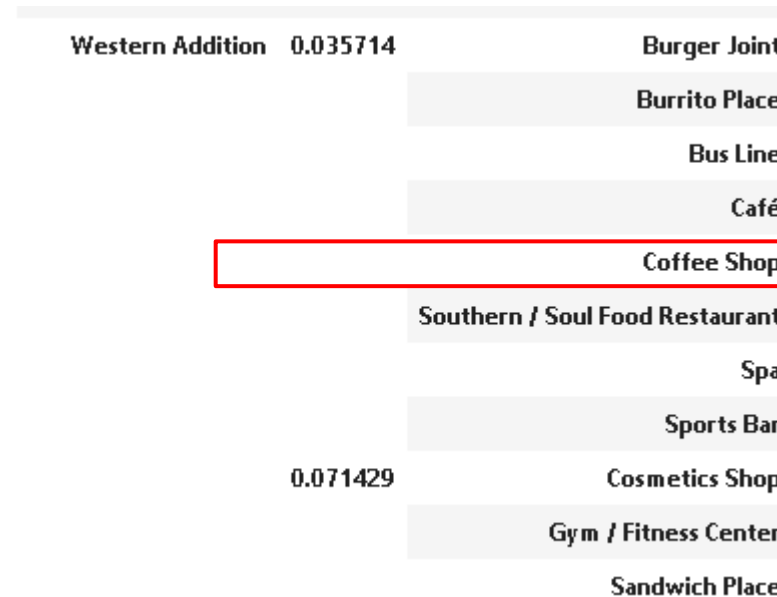
Top 10 Venues X neighborhoods filtered by Park and Richmond Districts - – part [1 -3]

Neighborhood	Value	Venue
Castro/Upper Market	0.018605	Asian Restaurant
		Dog Run
		Men's Store
		Playground
		Sandwich Place
	0.027907	Gym
		Mexican Restaurant
	0.032558	Grocery Store
	0.037209	Coffee Shop
	0.046512	Park
Golden Gate Park	0.051163	Gay Bar
		Bubble Tea Shop
		Bus Station
		Grocery Store
		Pizza Place
	0.023474	Science Museum
		Trail
		Playground
		Coffee Shop
		Lake
Haight Ashbury	0.022026	Ice Cream Shop
		Bookstore
		Shoe Store
		Breakfast Spot
		Pizza Place
	0.030837	Clothing Store
		Café
		Park
		Thrift / Vintage Store
		Coffee Shop
Hayes Valley	0.055066	Boutique
		Bakery
		Chinese Restaurant
		Clothing Store
		Dive Bar
	0.034483	Mexican Restaurant
		Park
		Pet Store
		Cocktail Bar
		Coffee Shop
Inner Richmond	0.018212	Asian Restaurant
		Café
		Burmese Restaurant
		Park
		Thai Restaurant
	0.028146	Sushi Restaurant
		Vietnamese Restaurant
		Korean Restaurant
		Japanese Restaurant
		Bakery
Inner Sunset	0.067881	Chinese Restaurant
		Breakfast Spot
		Park
		Sandwich Place
		Thai Restaurant
	0.025735	Ice Cream Shop
		Pizza Place
		Science Museum
		Bakery
		Sushi Restaurant
Park	0.040441	Garden
	0.047794	Coffee Shop

Park and Richmond Districts – part [2 -3]

		Lone Mountain/USF	0.015625	Pet Store	Presidio Heights	0.019231	Clothing Store	
Japantown	0.020833	Church	0.019531	Convenience Store	0.024038	Bakery		
		Nightclub		Liquor Store		Bank		
	0.041667	Bakery		Thrift / Vintage Store		Chinese Restaurant		
		Cosmetics Shop		Bank		New American Restaurant		
		Gym / Fitness Center	0.023438	Bus Station	Wine Bar			
		Salon / Barbershop		Mexican Restaurant	0.028846	Furniture / Home Store		
		Sandwich Place	0.027344	Boutique	Gym			
		Spa		Cosmetics Shop	Spa			
		Sports Bar	0.062500	Café	0.033654	Coffee Shop		
		Sushi Restaurant		Coffee Shop	0.043269	American Restaurant		
	0.062500	Chinese Restaurant						
Lincoln Park	0.000000	Event Space	Outer Richmond	0.021546	Thai Restaurant	Twin Peaks	0.000000	Egyptian Restaurant
	0.090909	Art Museum	0.024081	Coffee Shop			Falafel Restaurant	
		Bus Stop		0.026616	Pizza Place	0.043478	American Restaurant	
		Cafeteria		Vietnamese Restaurant			Coffee Shop	
		Café	0.027883	Bus Stop			Reservoir	
		Golf Course		Korean Restaurant			Tennis Court	
		Monument / Landmark	0.036755	Bakery			Thai Restaurant	
		Park	0.039290	Grocery Store	0.086957	Hill		
		Sculpture Garden		Sushi Restaurant		Park		
		Trail	0.057034	Chinese Restaurant	0.260870	Scenic Lookout		
	0.181818	Pharmacy	0.058302	Café	0.347826	Trail		

Top 10 Venues X neighborhoods filtered by Park and Richmond Districts – part [3 -3]

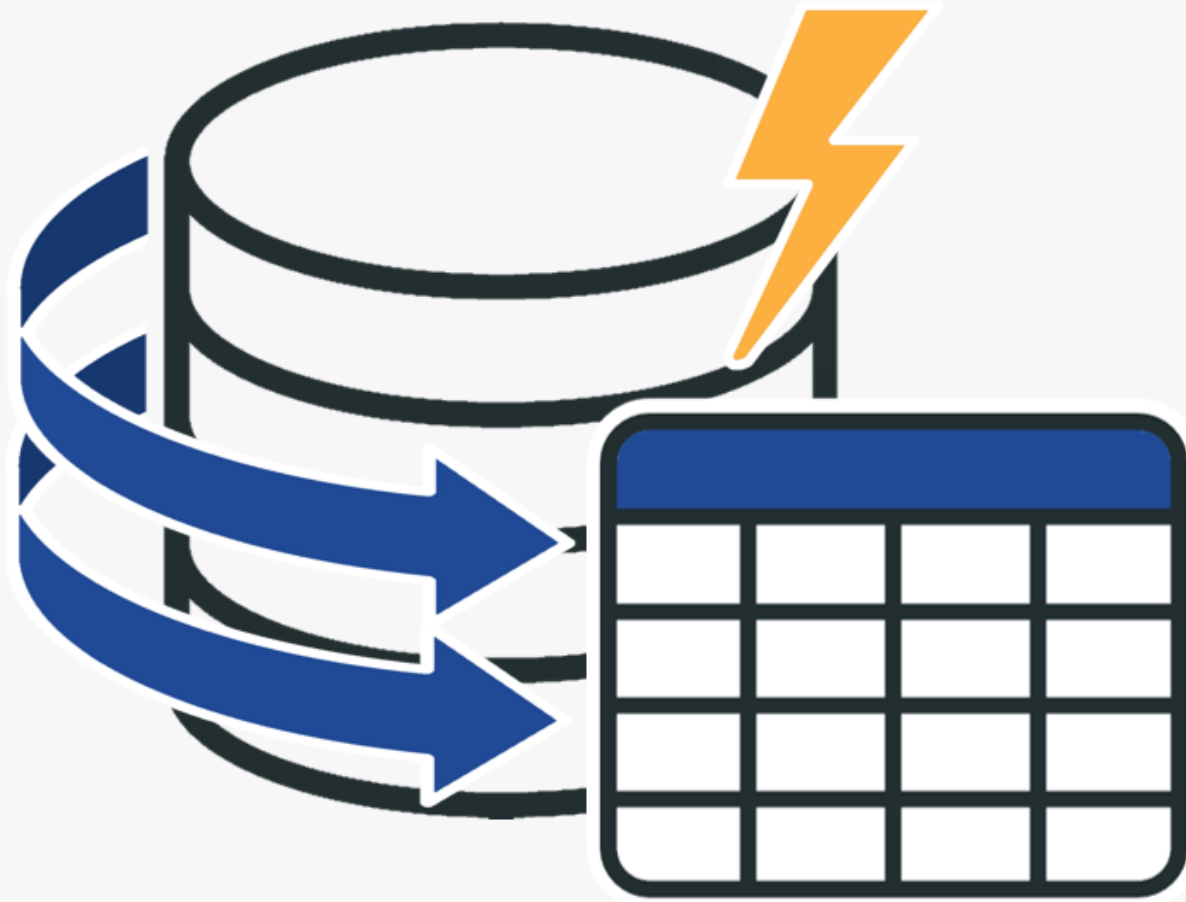


Analyzing the top 10 reports by neighborhood can be determined the coffee shop is a frequent and common business in the majority the venues.



Clustering

K means



Clustering

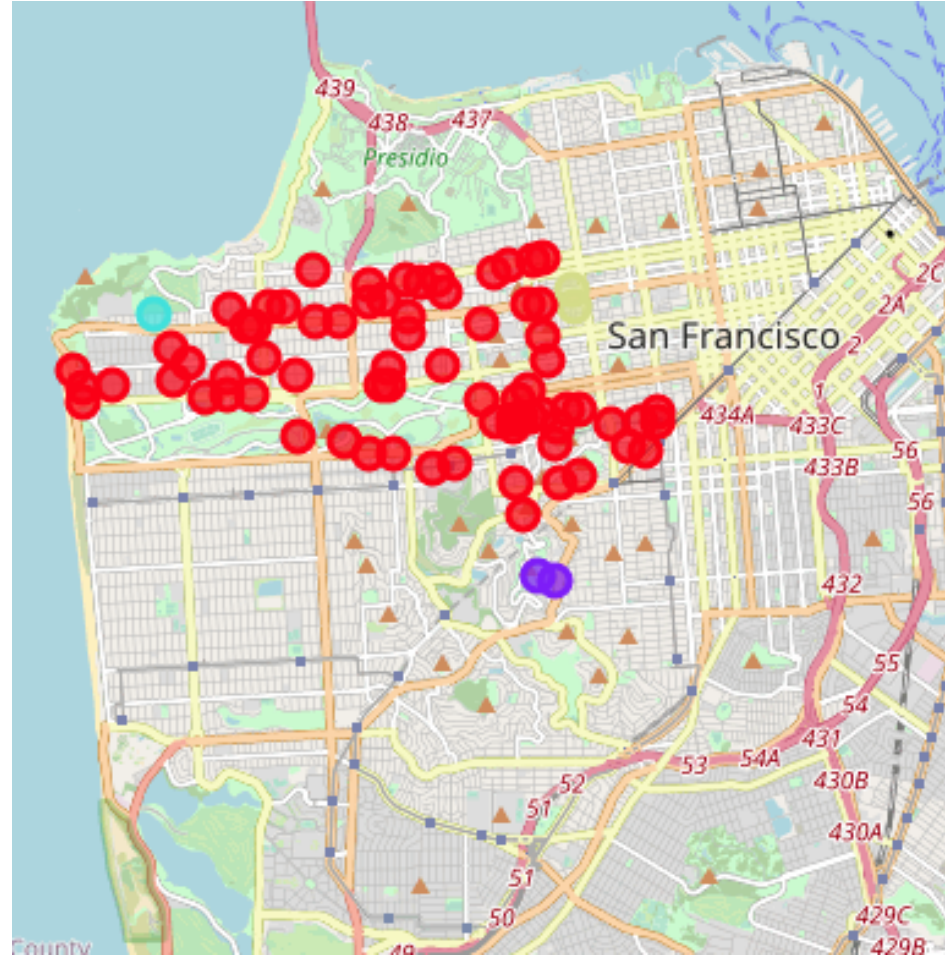
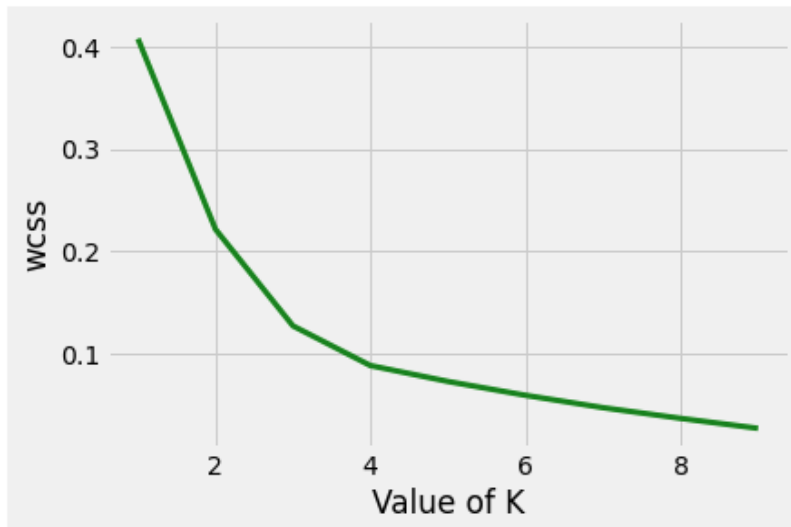
```
# Set dataframe for clustering, we only need the categories from ou
fran_grouped_clustering = fran_grouped.drop('Neighborhood', 1)

# Create empty array for within cluster sum of squares
wcss = []
for i in range(1, 10):
    KM = KMeans(n_clusters = i, max_iter = 500)
    KM.fit(fran_grouped_clustering)

    # Calculate within cluster sum of squares using function KM.inert
    wcss.append(KM.inertia_)

# Plot the cost against K values

style.use("fivethirtyeight")
plt.plot(range(1, 10), wcss, color = 'g', linewidth = '3')
plt.xlabel("Value of K")
plt.ylabel("wcss")
plt.show()
```



In cluster analysis, the elbow method shows a good candidate $k=4$



Cluster 0

According to the results for the first group (labels=0)
The most popular categories are:

- Boutique
- Coffee Shop
- Thrift / Vintage Store
- Clothing Store
- Chinese Rest.
- Garden
- Park
- Café
- Bakery

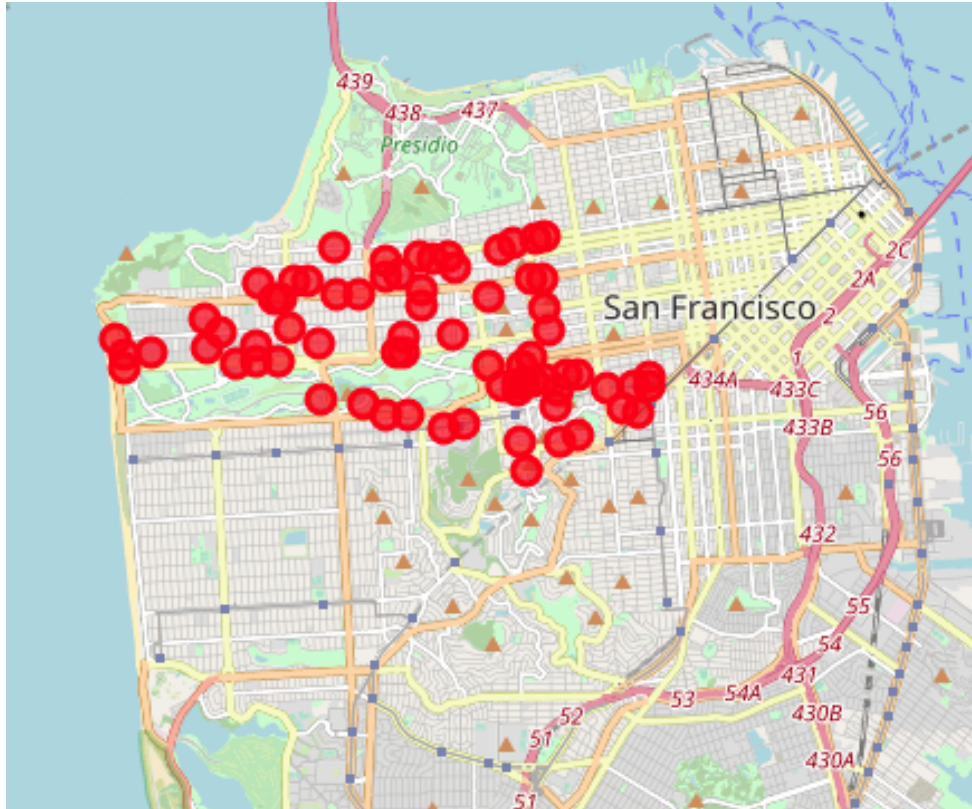
	Neighborhood	Latitude	Longitude	police_district	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Haight Ashbury	37.762519	-122.449296	Park	0	Boutique	Coffee Shop	Thrift / Vintage Store	Park	Café
1	Haight Ashbury	37.768729	-122.449454	Park	0	Boutique	Coffee Shop	Thrift / Vintage Store	Park	Café
2	Haight Ashbury	37.768944	-122.443356	Park	0	Boutique	Coffee Shop	Thrift / Vintage Store	Park	Café
3	Haight Ashbury	37.769307	-122.431570	Park	0	Boutique	Coffee Shop	Thrift / Vintage Store	Park	Café
4	Haight Ashbury	37.770210	-122.445345	Park	0	Boutique	Coffee Shop	Thrift / Vintage Store	Park	Café
5	Inner Sunset	37.764272	-122.460958	Park	0	Coffee Shop	Garden	Sushi Restaurant	Bakery	Science Museum
6	Inner Sunset	37.759091	-122.448080	Park	0	Coffee Shop	Garden	Sushi Restaurant	Bakery	Science Museum
7	Lone Mountain/USF	37.778837	-122.445347	Park	0	Café	Coffee Shop	Cosmetics Shop	Boutique	Bank
8	Lone Mountain/USF	37.772790	-122.447497	Park	0	Café	Coffee Shop	Cosmetics Shop	Boutique	Bank
9	Haight Ashbury	37.769058	-122.435794	Park	0	Boutique	Coffee Shop	Thrift / Vintage Store	Park	Café

It is recommended to open a coffee shop



Cluster 0

Graphic and Table



	Neighborhood	Latitude	Longitude	police_district	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Haight Ashbury	37.762519	-122.449296	Park	0	Boutique	Coffee Shop	Thrift / Vintage Store	Park	Café
1	Haight Ashbury	37.768729	-122.449454	Park	0	Boutique	Coffee Shop	Thrift / Vintage Store	Park	Café
2	Haight Ashbury	37.768944	-122.443356	Park	0	Boutique	Coffee Shop	Thrift / Vintage Store	Park	Café
3	Haight Ashbury	37.769307	-122.431570	Park	0	Boutique	Coffee Shop	Thrift / Vintage Store	Park	Café
5	Haight Ashbury	37.770210	-122.445345	Park	0	Boutique	Coffee Shop	Thrift / Vintage Store	Park	Café
6	Inner Sunset	37.764272	-122.460958	Park	0	Coffee Shop	Garden	Sushi Restaurant	Bakery	Science Museum
7	Inner Sunset	37.759091	-122.448080	Park	0	Coffee Shop	Garden	Sushi Restaurant	Bakery	Science Museum
8	Lone Mountain/USF	37.778837	-122.445347	Park	0	Café	Coffee Shop	Cosmetics Shop	Boutique	Bank
9	Lone Mountain/USF	37.772790	-122.447497	Park	0	Café	Coffee Shop	Cosmetics Shop	Boutique	Bank
10	Haight Ashbury	37.769058	-122.435794	Park	0	Boutique	Coffee Shop	Thrift / Vintage Store	Park	Café
11	Castro/Upper Market	37.762553	-122.442983	Park	0	Gay Bar	Park	Coffee Shop	Grocery Store	Gym
12	Castro/Upper Market	37.765996	-122.431038	Park	0	Gay Bar	Park	Coffee Shop	Grocery Store	Gym
13	Lone Mountain/USF	37.776017	-122.444780	Park	0	Café	Coffee Shop	Cosmetics Shop	Boutique	Bank

It is recommended to open a coffee shop



Cluster 1

According to labels=1 The most popular categories are:

- Trail
- Scenic Lookout Park
- Park
- Hill
- Tennis Court

1]:

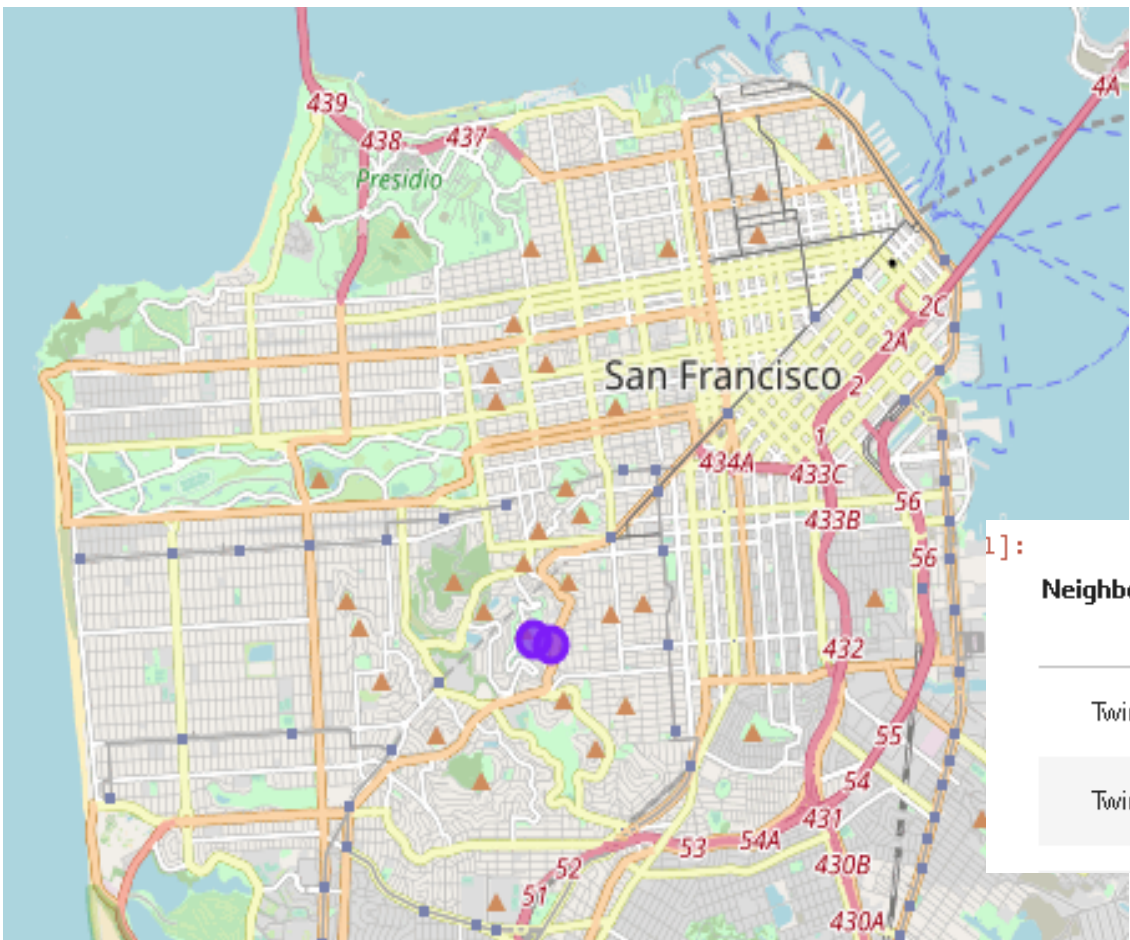
Neighborhood	Latitude	Longitude	police_district	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
Twin Peaks	37.751799	-122.443548	Park	1	Trail	Scenic Lookout	Park	Hill	Tennis Court
Twin Peaks	37.752352	-122.445984	Park	1	Trail	Scenic Lookout	Park	Hill	Tennis Court

It is not recommended to open a coffee shop



Cluster 1

Graphic and Table



1]:

Neighborhood	Latitude	Longitude	police_district	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
Twin Peaks	37.751799	-122.443548	Park	1	Trail	Scenic Lookout	Park	Hill	Tennis Court
Twin Peaks	37.752352	-122.445984	Park	1	Trail	Scenic Lookout	Park	Hill	Tennis Court

It is not recommended to open a coffee shop



Cluster 2

According to labels=2 The most popular categories are:

- Cosmetic Shop
- Gym / Fitness Center
- Sandwich Place
- Deli / Bodega
- Park
- Sports Bar
- Yoga Studio
- Chinese Restaurant

Neighborhood	Latitude	Longitude	Police District	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
Western Addition	37.782239	-122.440963	Park	2	Cosmetics Shop	Gym / Fitness Center	Sandwich Place	Deli / Bodega	Park
Japantown	37.784107	-122.441349	Richmond	2	Chinese Restaurant	Yoga Studio	Gym / Fitness Center	Salon / Barbershop	Sports Bar

It is not recommended to open a coffee shop

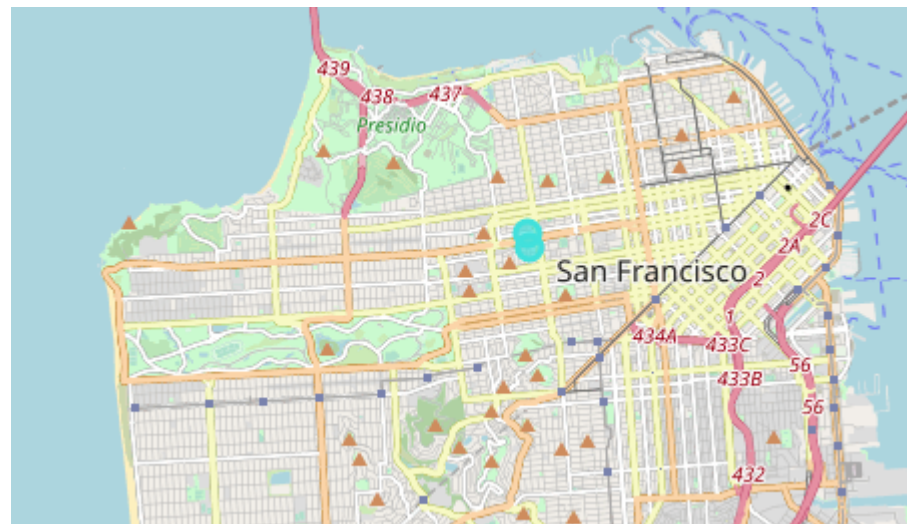


Cluster 2

Graphic and Table

According to labels=2 The most popular categories are:

- Cosmetic Shop
- Gym / Fitness Center
- Sandwich Place
- Deli / Bodega
- Park
- Sports Bar
- Yoga Studio
- Chinese Restaurant



It is not recommended to open a coffee shop



Cluster 3

According to labels=3 The most popular categories are:

- Pharmacy
- Park
- Gold Course
- Café
- Cafeteria
- Bus Stop

	Neighborhood	Latitude	Longitude	police_district	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Lincoln Park	37.781327	-122.499871	Richmond	3	Pharmacy	Park	Golf Course	Café	Cafeteria

It is recommended to open a coffee shop



Cluster 3

Graphic and table

According to labels=3 The most popular categories are:

- Pharmacy
- Park
- Gold Course
- Café
- Cafeteria
- Bus Stop



It is recommended to open a coffee shop

Answers to Initial Questions



- ❑ Which all areas have more number of coffee shops?

According to reports are: **Lincoln Park, Castro/Upper Market, Haight Ashbury, Hayes Valley, Inner Sunset, Lone Mountain, Inner Richmond, Outer Richmond, Presidio Heights , Western Addition**

- ❑ Which all areas have less number of coffeeshops?

According to reports are: **Inner Richmond, JapanTown, Twin Peaks**

- ❑ Which districts have the neighborhoods safest ?

According to reports are: **Park and Richmond**

- ❑ Which districts Neighborhoods The Most Unsafe Districts ?

According to reports are: **Marina, South of Market, Tenderloin, District/ south Beachark, BayView and misión.**



Conclusion

According to the final clustering applying modeling with k-means and elbow=4, it is recommended to open a **Coffee Shop** on the venues corresponding to **Lincoln Park** (corresponding to Richmond) and / or **Lincoln Park, Castro/Upper Market, Haight Ashbury, Hayes Valley, Inner Sunset, Lone Mountain, Inner Richmond, Outer Richmond, Presidio Heights, Western Addition** (corresponding to Park).

