Covid19 Rate & Venues Data Analysis of San Francisco

1. Introduction

1.1 Background

As of July 8, 2020, coronavirus disease 2019 (Covid-19) has been confirmed in more than 11,940,258 people worldwide, with 3,096,516 cases in the United States. WHO Coronavirus Updates shows that the observed case-fatality ratio is about 4.5% worldwide and 4.3% for the United States specifically [1]. In terms of factors that affect infection prevention and control, it interested me that analysis in Israel showed that poorer regions have more difficulties in reducing mobility [2], which is of huge importance for our fight with the epidemic [3].

San Francisco is the cultural, commercial, and financial center of Northern California, with 881,549 residents as of 2019. It covers an area of about 46.89 square miles (121.4 km^2) and has the highest salaries, disposable income, and median home prices in the world at $1.7 million, as well as the highest median rents [4]. Neighborhood grouping data was collected from San Francisco Government website and some data of Covid-19 cases summarized by zip code as well.

1.2 Problem

Under the same mobility restriction regulation within a city, one's actual mobility can be affected by his/her attitude, personality, personal needs and neighborhood type. First three factors are hard to estimate, so I mainly focus on the last. The following part used neighborhood venue data to indicate the neighborhood type. The above assumptions were made based on my own common sense and need to be testified in the future if possible.

The goal is to analyze the relationship between Covid-19 confirmation rate and neighborhood type clustered by venue types nearby neighborhood. As one's level of wealth may probably affect his/her mobility as well, I further used home value index to indicate the level of wealth, trying to improve the clustering and help with the cluster annotation.

2. Data acquisition and cleaning

2.1 Data sources

COVID-19 Cases and Deaths Summarized by ZIP Code Tabulation Area [5].
Planning Neighborhood Groups Map [6].
Neighborhood central coordinates scraped from google map.
Foursquare API to get the most common venues of given neighborhoods of San Francisco.
Home Value Index of San Francisco provided by Zillow [7]

2.2 Data cleansing and preprocessing

COVID-19 Cases and Deaths Summarized by ZIP Code Tabulation Area table that was last updated on Jul-02-2020 was downloaded and inserted into the notebook as a dataframe. NaNs in count and rate were replaced by 0. A new dataframe was created with only zip code and corresponding confirmation rate (Figure 1).

	ZIP_code	rate
0	94158	49.38
1	94131	24.75
2	94124	120.75
3	94107	50.86
4	94129	35.05

Figure 1

Planning Neighborhood Groups Map table that provides all neighborhoods name and their multipolygon coordinates was downloaded and inserted into the notebook as a dataframe. However, there are some problems using the mean coordinates as the neighborhood center. So, I decided to use the coordinates for each neighborhood provided by google. Neighborhood central coordinates were scraped from google map and inserted into the notebook as a dataframe with column neighborhood, latitude and longitude (Figure 2).

	Neighborhood	Latitude	Longitude
0	Seacliff	37.78590	-122.49070
1	Haight Ashbury	37.77050	-122.44860
2	Outer Mission	37.71570	-122.44580
3	Inner Sunset	37.76020	-122.47030
4	Downtown/Civic Center	37.78160	-122.41560

Figure 2

3. Methodology

3.1 Confirmation rate choropleth map

Choropleth map was created with folium python package from the dataframe with only zip code and corresponding confirmation rate mentioned above (Figure 3).

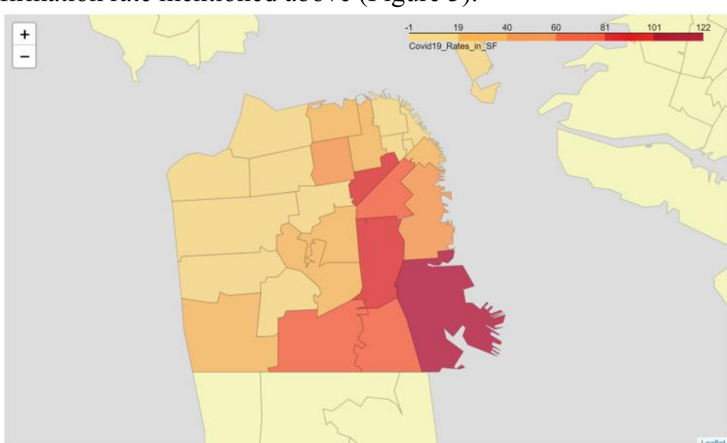

Figure 3

3.2 10 Most common venue types nearby

Explore API was used to obtain nearby venues information from Foursquare with neighborhood central coordinates of San Francisco. Venues number for each neighborhood was limited by 100. Radius of 1km were chosen because some neighborhoods have less then 5 venues in the radius of 500m.

To select the 10 most common venues type for each neighborhood, I first created neighborhood-categories one hot table and then grouped the rows by neighborhood. The mean of the frequency of occurrence of each category was calculated to replace the value in the one hot table. After sorting the categories of each row in descending order, I concatenated it with neighborhood names to construct the final table (Figure 4).

	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	Bayview	Southern / Soul Food Restaurant	Mexican Restaurant	Park	Light Rail Station	Bakery	Thrift / Vintage Store	Baseball Field	Theater	Grocery Store	Chinese Restaurant
1	Bernal Heights	Mexican Restaurant	Coffee Shop	Playground	Café	Cocktail Bar	Grocery Store	Italian Restaurant	Park	Bakery	Yoga Studio
2	Castro/Upper Market	Coffee Shop	Gay Bar	New American Restaurant	Seafood Restaurant	Wine Bar	Gym	Indian Restaurant	Park	Pizza Place	Dog Run
3	Chinatown	Coffee Shop	Pizza Place	Men's Store	Hotel	Café	Chinese Restaurant	Cocktail Bar	Park	New American Restaurant	Yoga Studio
4	Crocker Amazon	Mexican Restaurant	Pizza Place	Vietnamese Restaurant	Latin American Restaurant	Playground	Liquor Store	Sandwich Place	Grocery Store	Hot Dog Joint	Motel

Figure 4

3.3 Clustering models

3.3.1 First model and its problem

K-means clustering models can be used to divide neighborhoods into different groups. The first model was only based on the 10 most common venue types nearby the neighborhoods. To select the best k from 1 to 37, I tried to minimize the inertia or within-cluster sum-of-squares but found that inertia vs number of k was a monotonic decreasing function. Therefore, the number of k = 7 was chosen for it gave the largest decrease by adding one cluster, starting from sum of squared distance < 0.5. Each neighborhood was added to the confirmation rate choropleth map as circle marker.

The problem was that some clusters were very similar in terms of the 10 most common venues with could hardly explain the confirmation rate choropleth map.

3.3.2 Solution to the problem

I added one more feature to the data as home value index, based on the inference that poorer regions have more difficulties in reducing mobility [2] and will thus have a higher chance of infection. A set of new clustering models were built following the same steps as above. The result showed that adding home value index did not change the clustering result, but gave us some new understanding of the clusters.

4. Results

There were 37 neighborhoods and 323 unique venue categories.

4.1 Cluster Annotation

Clusters, corresponding neighborhoods, their 10 most common venues and home value index were shown below (Figure 5-11). Cluster annotation was written as table name.

	index	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	Home value index
0	20	Bayview	Southern / Soul Food Restaurant	Mexican Restaurant	Park	Light Rail Station	Bakery	Thrift / Vintage Store	Baseball Field	Theater	Grocery Store	Chinese Restaurant	1032682
1	21	Visitacion Valley	Light Rail Station	Park	Breakfast Spot	Bakery	Café	Garden	Coffee Shop	Marijuana Dispensary	Donut Shop	Art Gallery	971176

Figure 5. Cluster 0 has a relatively low home value index.

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	Home value index
Haight Ashbury	Park	Café	Coffee Shop	Bookstore	Mexican Restaurant	Boutique	Breakfast Spot	Playground	Shoe Store	Pizza Place	1722314
Downtown/Civic Center	Coffee Shop	Marijuana Dispensary	Theater	Bakery	Vietnamese Restaurant	Cocktail Bar	Beer Bar	Thai Restaurant	Music Venue	Sandwich Place	930025
Lakeshore	Lake	Golf Course	Sandwich Place	Gym	Park	Performing Arts Venue	Candy Store	Tennis Court	Thai Restaurant	Gym Pool	389712
Russian Hill	Park	Italian Restaurant	Bar	Coffee Shop	Pizza Place	Playground	Sushi Restaurant	Ice Cream Shop	Café	National Park	1838217
Noe Valley	Coffee Shop	Café	Gift Shop	Trail	Park	Pizza Place	Playground	Bookstore	Sushi Restaurant	Bakery	2180031
Parkside	Seafood Restaurant	Coffee Shop	Ice Cream Shop	Bakery	Café	Tour Provider	Park	American Restaurant	Italian Restaurant	Trail	1462524
Financial District	Coffee Shop	Food Truck	Bookstore	New American Restaurant	Seafood Restaurant	Japanese Restaurant	Men's Store	Gym	Wine Bar	Park	1131762
Mission	Mexican Restaurant	Café	New American Restaurant	Coffee Shop	Bar	Music Venue	Cocktail Bar	Yoga Studio	Bookstore	Boxing Gym	1506592
West of Twin Peaks	Coffee Shop	Park	Burger Joint	Pizza Place	Pharmacy	Italian Restaurant	Wine Bar	Cosmetics Shop	Grocery Store	Pub	1506592
Marina	Gym / Fitness Center	Park	French Restaurant	Sandwich Place	Wine Bar	Mexican Restaurant	Coffee Shop	Bookstore	Juice Bar	Thai Restaurant	2402845
Pacific Heights	Gym / Fitness Center	Italian Restaurant	French Restaurant	Sandwich Place	Park	Cosmetics Shop	Bakery	Wine Bar	Ice Cream Shop	Salon / Barbershop	2014414
Presidio Heights	Coffee Shop	Trail	Furniture / Home Store	Bakery	American Restaurant	Cosmetics Shop	Bank	Italian Restaurant	Golf Course	Breakfast Spot	4948331
South of Market	Coffee Shop	Vietnamese Restaurant	Bakery	Pizza Place	Marijuana Dispensary	Music Venue	Sandwich Place	Bar	Gym / Fitness Center	Art Gallery	1067121
Potrero Hill	Café	Park	Coffee Shop	Art Gallery	Deli / Bodega	Mexican Restaurant	Sandwich Place	Cocktail Bar	Wine Shop	New American Restaurant	1514888
Castro/Upper Market	Coffee Shop	Gay Bar	New American Restaurant	Seafood Restaurant	Wine Bar	Gym	Indian Restaurant	Park	Pizza Place	Dog Run	1514888
Bernal Heights	Mexican Restaurant	Coffee Shop	Playground	Café	Cocktail Bar	Grocery Store	Italian Restaurant	Park	Bakery	Yoga Studio	1587365
Nob Hill	Italian Restaurant	Wine Bar	Sushi Restaurant	Pet Store	Pizza Place	Bar	Vietnamese Restaurant	Grocery Store	Coffee Shop	Steakhouse	1544584
Chinatown	Coffee Shop	Pizza Place	Men's Store	Hotel	Café	Chinese Restaurant	Cocktail Bar	Park	New American Restaurant	Yoga Studio	1623693
North Beach	Seafood Restaurant	Coffee Shop	Pizza Place	Bakery	Park	Tour Provider	American Restaurant	Trail	Ice Cream Shop	Café	1376459
Western Addition	Bakery	Tea Room	Ice Cream Shop	Sushi Restaurant	Park	Gift Shop	Grocery Store	Ramen Restaurant	Seafood Restaurant	Bookstore	1332515

Figure 6. Cluster 1 has mix-cultured neighborhoods with many parks and bars, with a relatively high average home value index 167973.6. Most restaurants are Mexican, Italian and New American style.

	index	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	Home value index
0	28	Twin Peaks	Trail	Park	Scenic Lookout	Playground	Yoga Studio	Hill	Grocery Store	Garden	Monument / Landmark	Mountain	0

Figure 7. Cluster 2 is a mountain spot

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	Home value index
Seacliff	Café	Chinese Restaurant	Coffee Shop	Sushi Restaurant	Korean Restaurant	Seafood Restaurant	Thai Restaurant	Vietnamese Restaurant	Pizza Place	Scenic Lookout	4203253
Inner Sunset	Vietnamese Restaurant	Bakery	Sushi Restaurant	Coffee Shop	Garden	Pizza Place	Deli / Bodega	Thai Restaurant	Sandwich Place	Ice Cream Shop	1775304
Outer Richmond	Café	Chinese Restaurant	Vietnamese Restaurant	Bakery	Grocery Store	Pizza Place	Sushi Restaurant	Korean Restaurant	Japanese Restaurant	Playground	1638102
Ocean View	Grocery Store	Café	Intersection	Bakery	Playground	Chinese Restaurant	Convenience Store	Liquor Store	Light Rail Station	Garden	1085615
Inner Richmond	Bakery	Japanese Restaurant	Chinese Restaurant	Korean Restaurant	Sushi Restaurant	Asian Restaurant	Coffee Shop	Burmese Restaurant	Thai Restaurant	Vietnamese Restaurant	2139254
Glen Park	Park	Pizza Place	Café	Trail	Breakfast Spot	Convenience Store	Chinese Restaurant	Coffee Shop	Bakery	Mexican Restaurant	1736050
Presidio	Trail	Café	Park	Food Truck	Tunnel	General Entertainment	Museum	Scenic Lookout	Baseball Field	Beach	2549839
Outer Sunset	Chinese Restaurant	Sandwich Place	Bubble Tea Shop	Dumpling Restaurant	Café	Pharmacy	Japanese Restaurant	Korean Restaurant	Bakery	Coffee Shop	1426495
Golden Gate Park	Chinese Restaurant	Park	Bubble Tea Shop	Vietnamese Restaurant	Bakery	Grocery Store	Dumpling Restaurant	Lake	Playground	Thrift / Vintage Store	0

Figure 8. Cluster 3 has many Asian restaurants, especially Chinese-style restaurants and a relatively high average home value index 2069239.0

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	Home value index
Outer Mission	Mexican Restaurant	Pizza Place	Latin American Restaurant	Vietnamese Restaurant	Sandwich Place	Liquor Store	Hot Dog Joint	Bus Station	Gastropub	Gas Station	1073887
Crocker Amazon	Mexican Restaurant	Pizza Place	Vietnamese Restaurant	Latin American Restaurant	Playground	Liquor Store	Sandwich Place	Grocery Store	Hot Dog Joint	Motel	1136531
Excelsior	Mexican Restaurant	Latin American Restaurant	Deli / Bodega	Convenience Store	Grocery Store	Restaurant	Pizza Place	Chinese Restaurant	Park	Vietnamese Restaurant	1157135

Figure 9. Cluster 4 has many Mexican, Latin American, Vietnamese restaurants, and a relatively low average home value index 1122517.7

	index	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	Home value index
0	5	Diamond Heights	Park	Playground	Trail	Theater	Pharmacy	Coffee Shop	Scenic Lookout	Grocery Store	Bus Station	Tennis Court	1427401

Figure 10. Cluster 5 has many parks and playgrounds and a relatively high home value index 1427401

	index	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	Home value index
0	9	Treasure Island/YBI	Food Truck	Athletics & Sports	Bus Station	Music Venue	Baseball Field	Harbor / Marina	Island	Rugby Pitch	Park	Flea Market	0

Figure 11. Cluster 6 is an island.

4.2 Choropleth map

Each neighborhood was added to the confirmation rate choropleth map as circle marker (Figure 12). Clusters were marked with different colors. Popup labels were added as well, which would show the cluster name when the marker was clicked on.

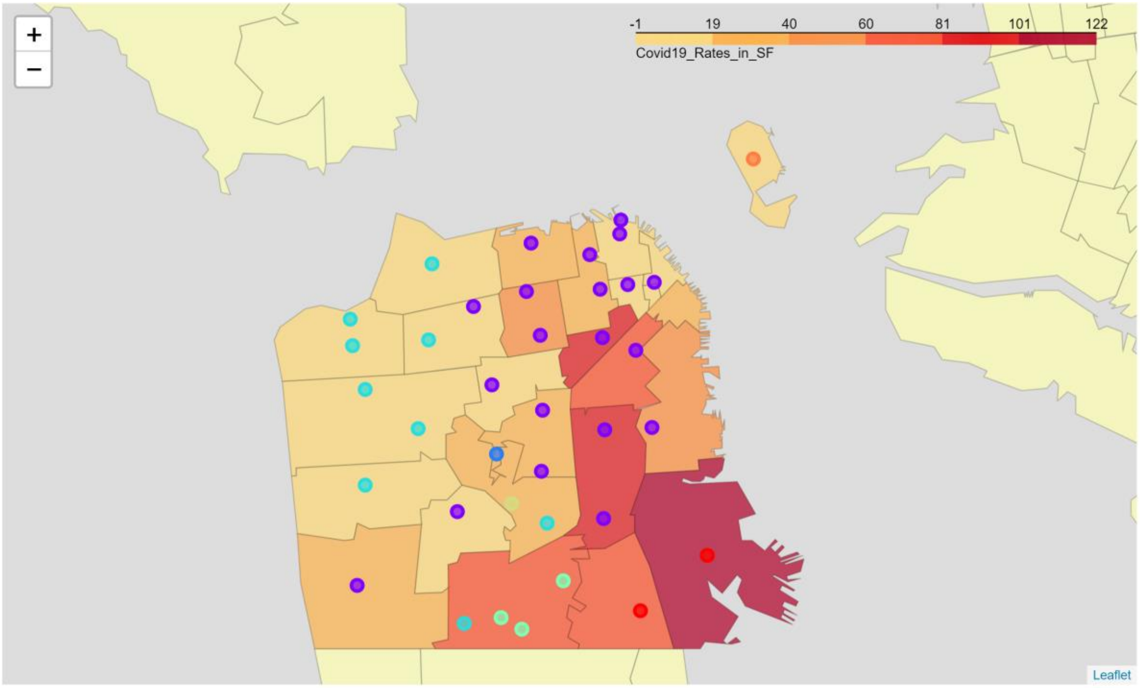

Figure 12 – Neighborhood clustering and covid19 confirmation rate choropleth map

5. Conclusion

In this study, I analyzed the relationship between neighborhood's covid19 confirmation rate and their nearby venue type and home value index data. I selected the 10 most common nearby venue types for building the clustering model and added the home value index to assist cluster annotation. Results showed that Asian neighborhood cluster had a relatively low covid19 confirmation rate while European, American, and Mexican neighborhood clusters can be further divided with the home value index, where a higher average home value index indicated a lower confirmation rate.

I hope further measures could be taken to help those neighborhood clusters with high confirmation rate and latent high risk.

6. References

[1] WHO coronavirus updates
[2] Human mobility and poverty as key factors in strategies against COVID-19
Matan Yechezkel, Amit Weiss, Idan Rejwan, Edan Shahmoon, Shahaf Ben Gal, Dan Yamin
medRxiv 2020.06.04.20112417; doi: https://doi.org/10.1101/2020.06.04.20112417
[3] Engle, Samuel and Stromme, John and Zhou, Anson, Staying at Home: Mobility Effects of COVID-19 (April 3, 2020). Available at SSRN: http://dx.doi.org/10.2139/ssrn.3565703
[4] San Francisco — Wikipedia
[5] https://data.sfgov.org/COVID-19/COVID-19-Cases-and-Deaths-Summarized-by-ZIP-Code-T/tef6-3vsw
[6] https://data.sfgov.org/Geographic-Locations-and-Boundaries/Planning-Neighborhood-Groups-Map/iacs-ws63
[7] https://www.zillow.com/san-francisco-ca/home-values/