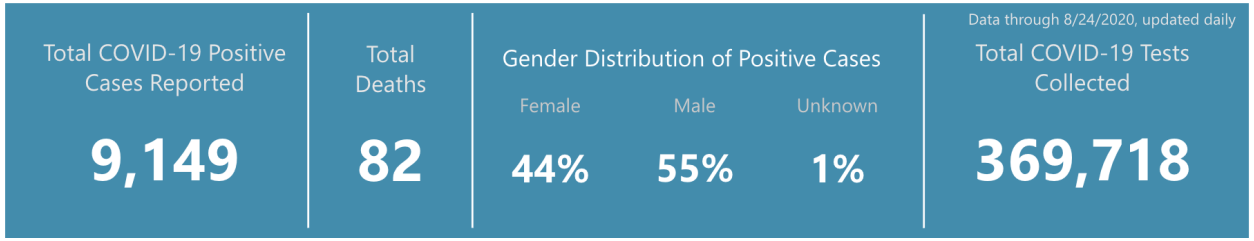
# Correlating COVID-19 Cases with Neighborhood Venues in San Francisco

**IBM Applied Data Science Capstone**

# San Francisco COVID-19 Statistics

- San Francisco has been one of the earliest COVID-19 responders
- Stay-at-home orders were issued on March 14, 2020.
- Currently, San Francisco is one of the large metropolitan areas in the US that has been keeping COVID-19 largely under control
- Compared to its population (~800,000) the number of cases and deaths are relatively low



# Project Aims

- This project aims to understand the relationship between confirmed COVID-19 cases and San Francisco neighborhoods
- Under the assumption that most individuals are infected outside of their home, we can consider each venue as a potential site of infection.
- We will analyze the relationship between the types and numbers of venues in a neighborhood and its cases.
- The results of this analysis would be invaluable for local policymaker. This will inform them in shaping re-opening policy for the city in order to maintain public safety while still stimulating the local economy.

# To correlate COVID-19 cases and venues, we will use two data sources: Four Square and DataSF

- Four Square is a location technology platform that provides crowdsourced data on on venues around a point of interest
  - Venue name
  - Venue address
  - Venue type
  - Venue tips
  - Venue photos
- DataSF provides public datasets to the city departments of San Francisco. The dataset we will be using, will detail:
  - Medical provider confirmed COVID-19 cases
  - Medical provider confirmed COVID-19 related deaths
  - Neighborhood population

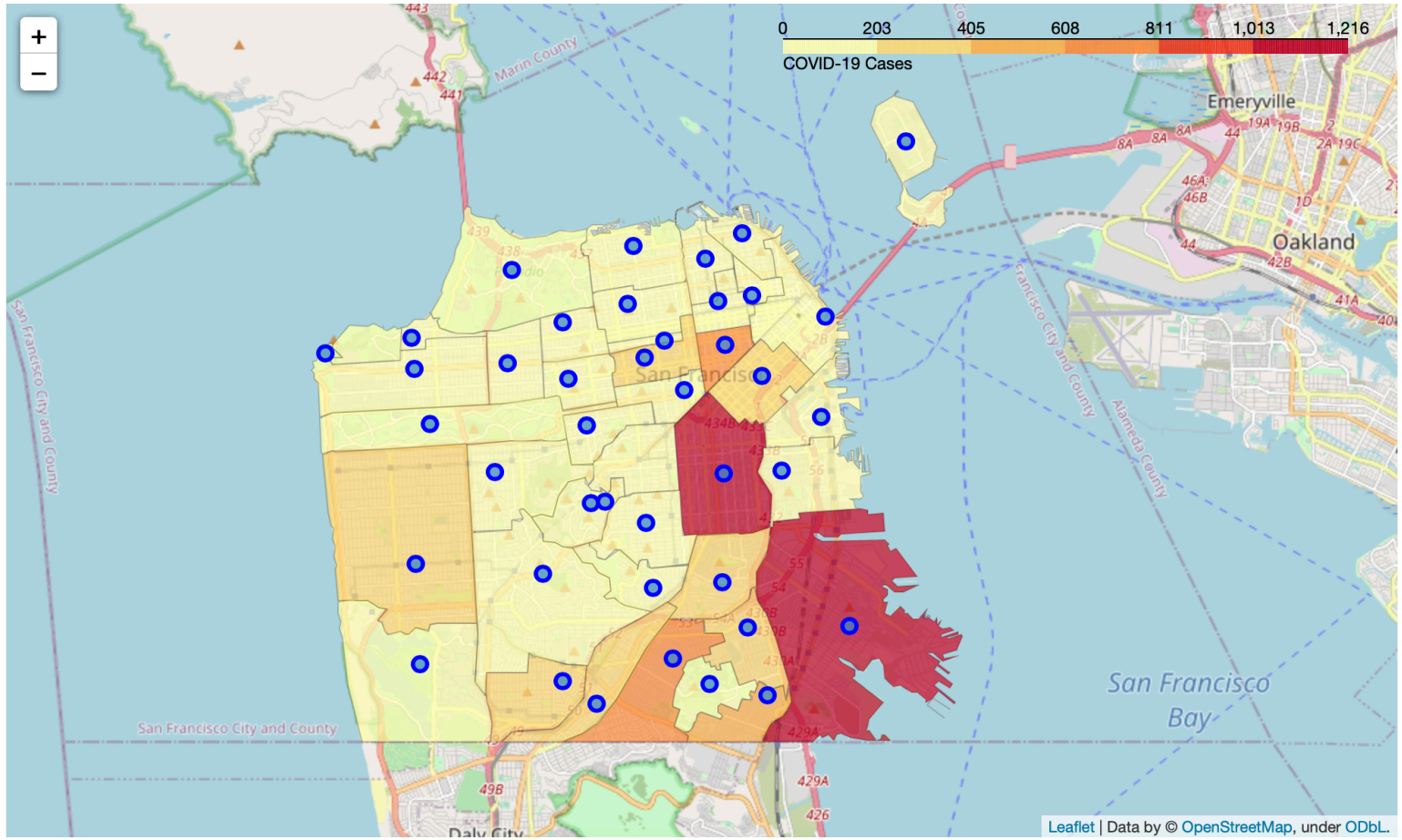
# Using Four Square API, we generated a venue profile for each neighborhood based on its top 10 most common venues

| 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 |
|---|---|---|---|---|---|---|---|---|---|---|
| Bayview Hunters Point | Food & Drink Shop | Gym | Grocery Store | Women's Store | Fast Food Restaurant | Ethiopian Restaurant | Event Space | Farm | Farmers Market | Filipino Restaurant |
| Bernal Heights | Trail | Bakery | Coffee Shop | Park | Italian Restaurant | Gourmet Shop | New American Restaurant | Bus Stop | Butcher | Café |
| Castro/Upper Market | Scenic Lookout | Trail | Tailor Shop | Hill | Café | Park | Reservoir | Women's Store | Farm | Ethiopian Restaurant |
| Chinatown | Chinese Restaurant | Coffee Shop | Bakery | Italian Restaurant | Cocktail Bar | Hotel | New American Restaurant | Men's Store | Burger Joint | Café |
| Excelsior | Scenic Lookout | Convenience Store | Lake | Park | Fast Food Restaurant | Escape Room | Ethiopian Restaurant | Event Space | Farm | Farmers Market |
| Financial District/South Beach | Coffee Shop | Food Truck | Café | Gym | Gym / Fitness Center | Scenic Lookout | Art Gallery | Salad Place | Sandwich Place | Spa |
| Glen Park | Trail | Park | Coffee Shop | Bubble Tea Shop | Gift Shop | Cheese Shop | Grocery Store | Gym | Salon / Barbershop | French Restaurant |
| Golden Gate Park | Park | Intersection | Bus Stop | Disc Golf | Playground | French Restaurant | Food Truck | Food & Drink Shop | Food | Flower Shop |
| Haight Ashbury | Boutique | Coffee Shop | Clothing Store | Shoe Store | Convenience Store | Bookstore | Breakfast Spot | Café | Pizza Place | Park |
| Hayes Valley | Wine Bar | Boutique | Clothing Store | Dessert Shop | New American Restaurant | Optical Shop | Café | Sushi Restaurant | French Restaurant | Pizza Place |
| Inner Richmond | Chinese Restaurant | Thai Restaurant | Sushi Restaurant | Japanese Restaurant | Asian Restaurant | Bakery | Vietnamese Restaurant | Korean Restaurant | Café | Dim Sum Restaurant |

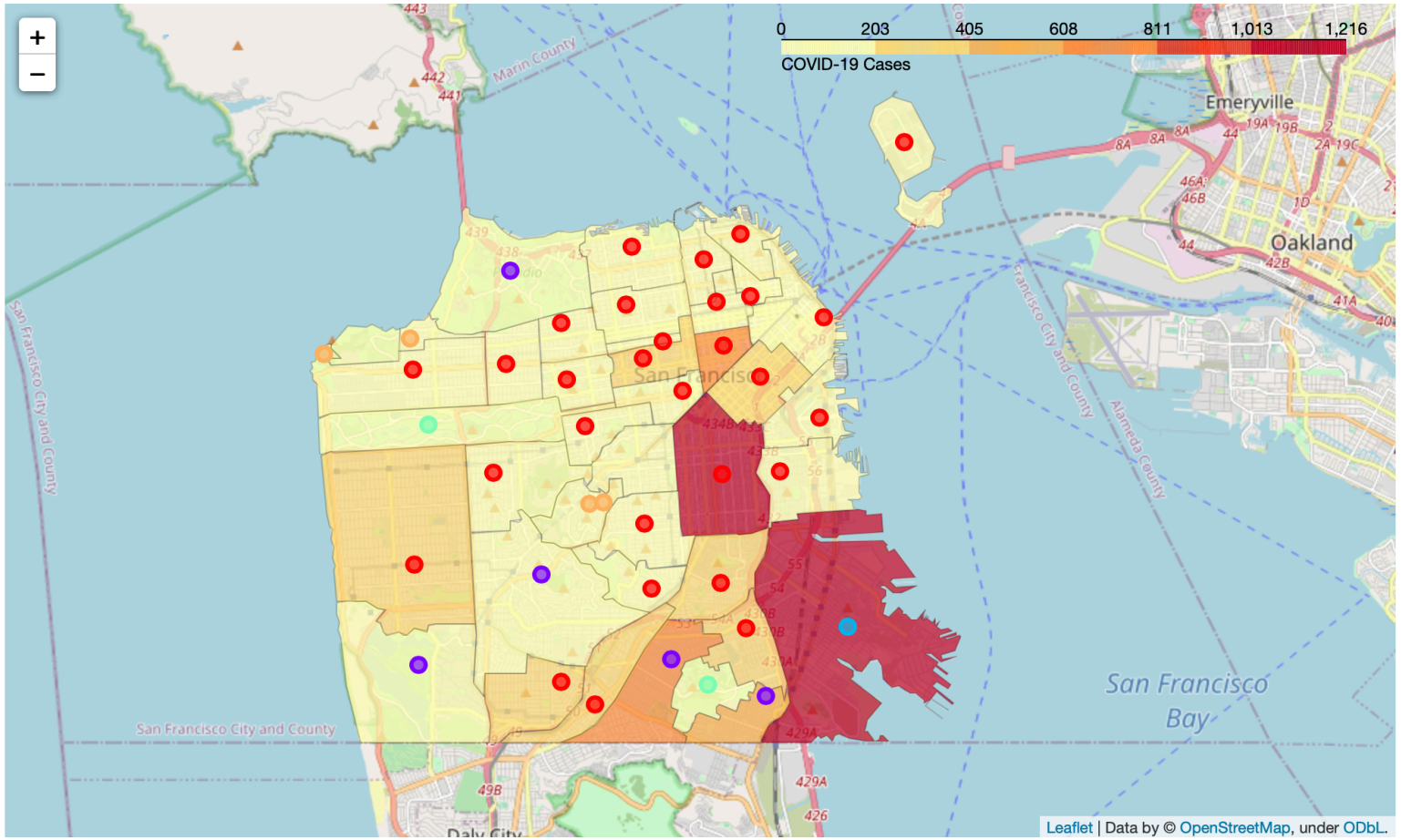
# Using the dataset from DataSF, we obtained the number of confirmed COVID-19 cases in San Francisco by neighborhood

## Neighborhood COVID-19 Data

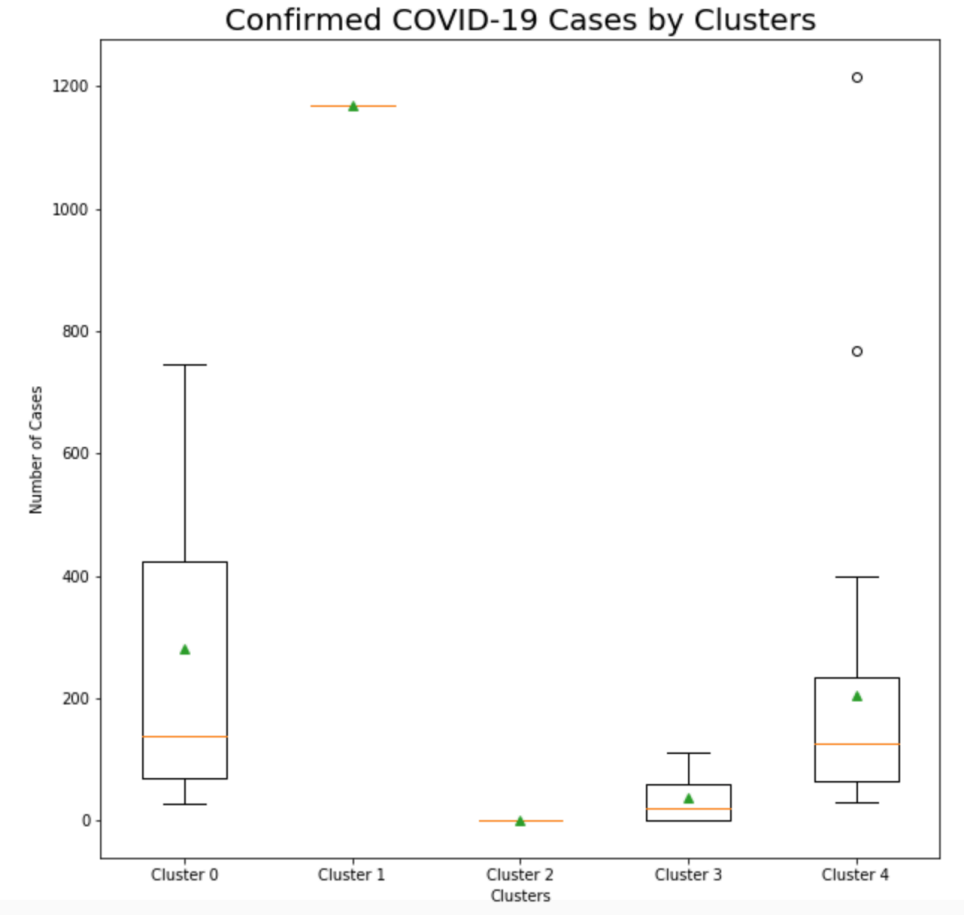
| Neighborhood | Resident Population | Cases |
|---|---|---|
| Mission | 59,639 | 1224 |
| Bayview Hunters Point | 37,394 | 1186 |
| Tenderloin | 29,588 | 775 |
| Excelsior | 40,701 | 754 |
| Visitacion Valley | 19,005 | 429 |
| Outer Mission | 24,853 | 401 |
| Bernal Heights | 25,858 | 311 |
| Oceanview/Merced/Ingleside | 28,217 | 292 |
| South of Market | 21,771 | 282 |
| Portola | 16,563 | 253 |
| Western Addition | 22,638 | 237 |
| Sunset/Parkside | 82,410 | 225 |
| Potrero Hill | 14,209 | 200 |
| Outer Richmond | 45,891 | 157 |



# Next, we cluster the neighborhoods by venues and superimposed it onto the heat map of COVID-19 cases to identify any correlation



Based on the distribution of COVID-19 cases per cluster, there does not seem to be a correlation between neighborhood venues and number of cases




## Limitations

- This analysis does not account for the movement of SF resident. This means that an infection occurring elsewhere but recorded for a SF neighborhood

- Another limitation is that the gradual re-opening and the state of the venues in each neighborhood is not factored in