Evaluation of SF areas of COVID-19 risk

1. Introduction

**Coronavirus disease 2019 (COVID-19)** is an infectious disease caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). SARS-CoV-2, is a positive-sense single-stranded RNA virus and quite contagious in humans. Once infected, people might develop symptoms including fever, cough, shortness of breath, diarrhea, sore throat, and abdominal pain. While the majority of cases result in mild symptoms, some progress to pneumonia and multi-organ failure.

The on-going pandemic of COVID-19 that has been designated a Public Health Emergency of International Concern on 30 January 2020 by the World Health Organization (WHO), and a pandemic on 11 March 2020. As to date, over 3 millions of people over the world have been infected, with 1 million of these cases in the United States. Thousands people lost their lives, while millions of people have to follow social distancing or stay-at-home orders to protect themselves.

It is widely believed that COVID-19 is not going away soon. So the modifications to Stay-at-home Order must be guided by health risk and a commitment to equity. San Francisco is the one of the most populous city in the United States. The denser the city, the more easily disease can spread. Residents in larger cities definitely have to be more careful and engage in more intense social distancing. In this project, I would like to use data sets obtained from DataSF.org to examine different areas in San Francisoco. I will consider the COVID-19 confirmed **case numbers, population, health care availability, as well as venue type and density** to identify the most risky areas in San Francisco. I hope this can be an easy guide for people to choose where to go for essential businesses, such as grocery shopping. This information might also be helpful to small business owners (e.g. restaurant owners) to make decisions about whether it is too risky to still continue businesses based on their locations.

1. Data

**DataSF.org** provides a lot of open data sets about San Francisco, while Four Square API is resourceful of venue information. We will use the following data sources for this project: 1) Rate of COVID-19 cases by census ZIP code from DataSF.org, which contains data updated as of 04/25/2020 about confirmed COVID-19 cases in different regions of San Francisco. This file also provides population in different areas. 2) Health care facilities information from DataSF.org, which summarizes all the health care facilities available in San Francisco. 3) Information about venues in this region from the Four Square API.

Data downloaded were converted to data frames, cleaned, and converted to other relevant information if needed, as follows.

First, I collected the **confirmed cases of COVID-19 and population in regions with different zip codes. This one is straightforward from the csv file. Coordinates for each zip codes were then obtained by geolocator.**

Next, I cleaned the health care facilities file. This data set included only longitude and latitude values for each facility, but did not indicate the corresponding neighborhood or zip code. So the coordinates were converted to zip code by geopy in order to join the other tables later. I also **identified health care facilities more relevant to fight against COVID-19, i.e. excluding facilities such as drug treatment facilities or sexual health clinics**. One hot coding was then used to collect different types of health care facilities and then summarized for each zip code.

Venue information for each zip code area were collected from Four Square API and sorted. Using one hot coding to examine different categories and group them by zip code. The frequency of occurrence of each category was calculated and the top 10 venue categories for each zip code region were displayed in a new data frame.

These steps left us with data frames that could be joined by zip codes and contain information about the COVID-19 confirmed **case number/population, health care availability, as well as top venue categories.**