Battle of the Neighborhoods
Visiting a restaurant in San Francisco during the COVID 19 Pandemic

Matias Garib
July 27th, 2020

1. Introduction

1.1 Background

The COVID-19 pandemic has had a tremendous impact on our way of life. Today, hygiene has become a fundamental aspect in households, businesses and especially public places. Appropriate measures such as washing your hands, using a facemask and practicing social distancing have proved to be the most effective ways of avoiding the spread of the disease. While many day-to-day activities have been affected by the virus, going out to dinner is for many the most sorely missed thing to do. As restaurants start to re-open, families, foodies, couples and anyone going out for a meal will have to choose wisely where to go in order to avoid getting infected. Although hygiene has always been of utmost importance in restaurants, this matter is now even more relevant, and it would be very useful for people to know which places have the highest and lowest hygiene standards when choosing where to go.

1.2 Problem

People want to start going out and visiting restaurants, but they want to visit places with the best hygiene practices. The questions we want to answer, for the city of San Francisco, are: which are the cleanest restaurants in each neighborhood? Which are the safest neighborhoods to go out to eat?

2. Data

2.1 Sources

The first dataset to be used consists of a GeoJSON file with the names and boundaries of 41 San Francisco neighborhoods. It is based on the 2010 census and can be accessed here (DataSF, 2019). A view of the map is showed below.

Secondly, Foursquare APIs will be used to explore each neighborhood's restaurants. Foursquare contains massive datasets of accurate location data and works as social media as well. The main API that will be used is the Explore API, which returns popular spots around a certain location.

Finally, the set of data used to perform the proposed restaurant hygiene analysis is provided by the City of San Francisco and is the result of the Health Department's inspection program. Based on the restaurants retrieve from the Foursquare API we will determine the health risks for each one, cluster them and map them. The data was last updated July 27th, 2020 and can be accessed here. Some of the relevant data we will use include:

- business_name: Common name of the business.
- business_address: Street address of the business
- inspection_date: Date of the inspection in YYYYMMDD format
- inspection_score: Inspection score on a 0-100 scale. 100 is the highest score.
- inspection_type: String representing the type of inspection. Must be one of: initial, routine, followup, complaint
- violation_description: One-line description of the violation. 200-character max.
- risk_category: Low, medium or high risk depending on inspection score.

References