**Where to eat Italian food in San Francisco during the Coronavirus Pandemic**

**Jeremy Daggett**

**7/26/2020**

1. **Introduction**
   1. **Background**

With the COVID-19 pandemic, many businesses like offices, schools, colleges, etc. have been shut down forcing people to work/study from home. However, the food industry personnel do not have the opportunity to work from home and are required to continue with their daily work activities while adhering to the safety measures.

What’s concerning is that it can survive on different surfaces for quite a long time and once we come in contact with a contaminated surface, we can get infected too. This naturally makes us think twice before ordering food online or before visiting a restaurant to dine-in.

This would also change consumer behavior and would affect our decision while choosing any food outlet. Consumers would not only look at food outlets that serve good quality food at a good price but will also look at the hygiene rating and the area in which the food outlet is located in to ensure that safety and hygiene are not being compromised.

* 1. **Problem**

Going out to eat during the COVID-19 pandemic has become a challenging decision for many people. Many people are making their choices based upon more of safety than ever before. In order to reduce risk and thus provide more information for people to make better decisions, the gathering of different types of safety data points are needed: coronavirus counts, restaurant hygiene ratings, inspection scores and reviews.

* 1. **Interest**

Consumer behavior has and will change for the foreseeable future with respect to many choices people will need to make including deciding on food outlets. This behavior will drive food outlets themselves to make better choices and will lead to better data points for software to gather for consumers.

1. **Data Acquisition and Cleaning**
   1. **Data Sources**

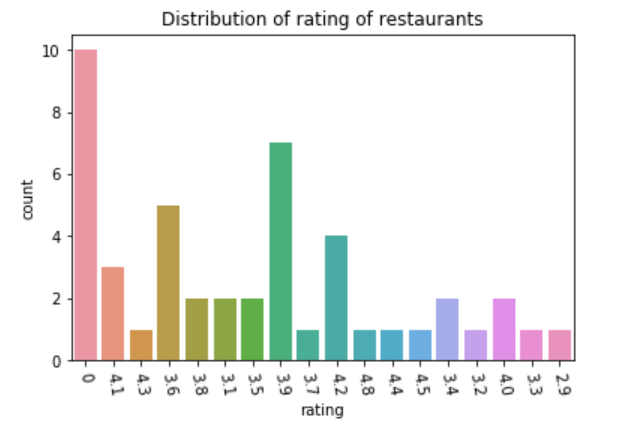
Data has been gathered from the San Francisco Government API for Covid-19 and Health Inspection Data, the Foursquare API and the Zomato API:

* [San Francisco Neighborhood Covid-19 Data](#Covid\_Data) - To get all the confirmed coronavirus cases in the different neighborhoods of San Francisco
* [San Francisco Government Restaurant Health Inspection Data]- Using the San Francisco's LIVES restaurant inspection data leverages the LIVES Flattened Schema (https://goo.gl/c3nNvr), which is based on LIVES version 2.0, cited on Yelp's website (http://www.yelp.com/healthscores).
* [FourSquare API]- Used the location coordinates of the districts we received from the Covid19 API and pass it as input to the FourSqaure API to retrieve 100 venues within 4 kms for each Neighborhood of San Francisco.
* [Zomato API] - Used the coordinates of the neighborhoods we receive from the FourSquare Venue Dataset, pass it to the Zomato API to retrieve Pricing for Two, Pricing Tier and Customer Rating. Note: Hygiene Ratings were NOT available for any of the restaurants.
  1. **Data Cleaning**

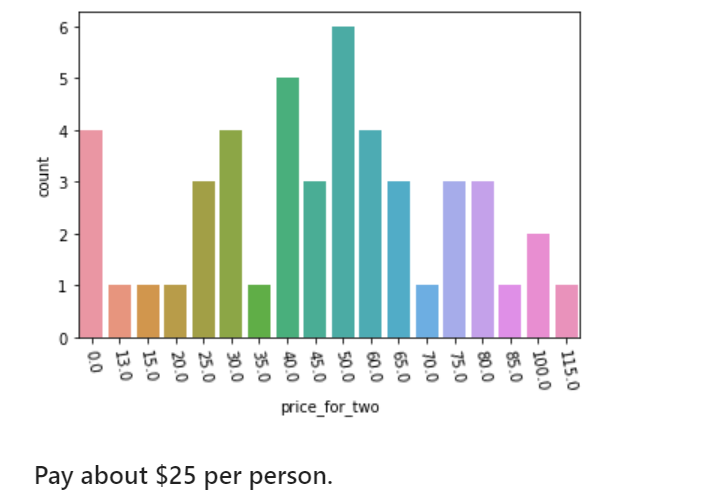
After each data set was downloaded into a Python Dataframe; the dataset was inspected and basic updates where necessary was performed (i.e. removing of duplicates).

Once each data set was sound; the next step was to merge the datasets together and that meant to come up with uniform column names to merge on; make sure casing is not preventing valid merges

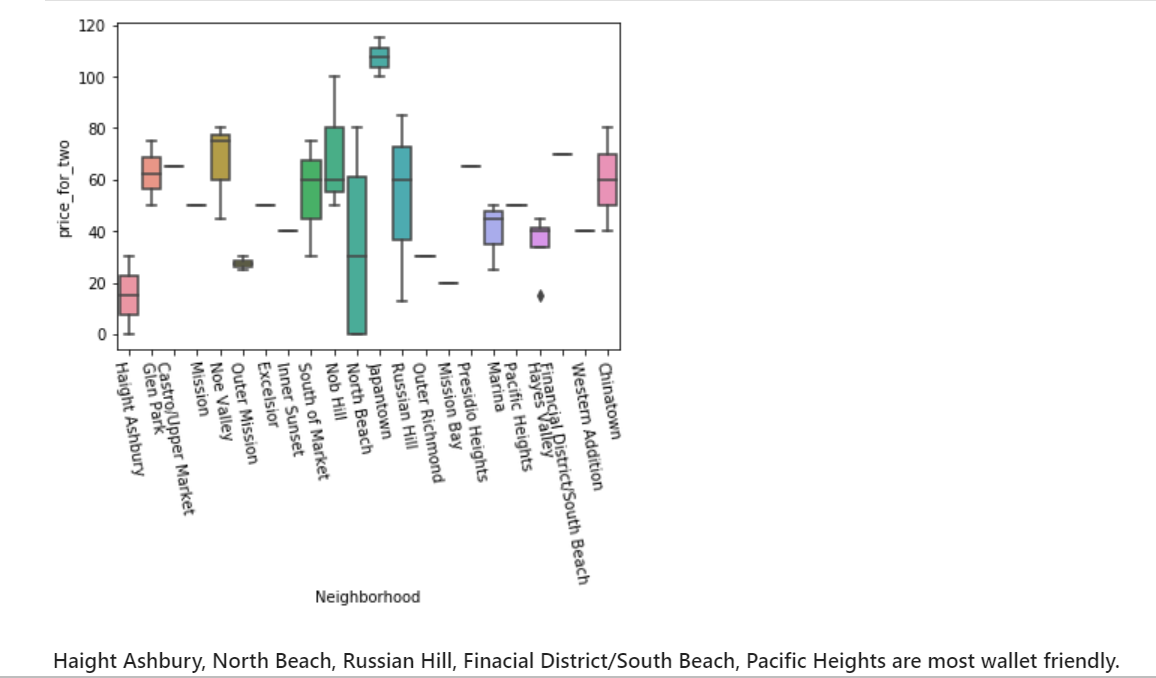
1. **Exploratory Data Analysis**
   1. **Distribution of Ratings**



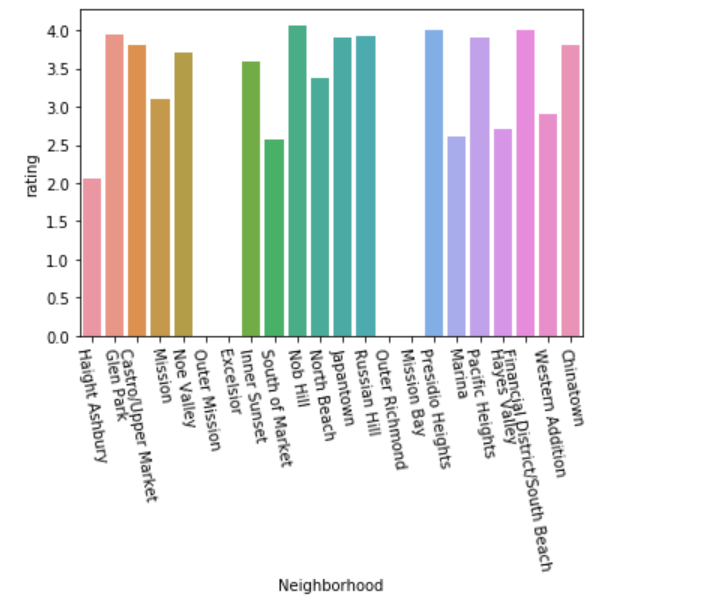
* 1. **Price Distributions**



* 1. **Price Distribution per Neighborhood**



* 1. **Ratings Distribution per Neighborhood**



1. **Clustering of Neighborhoods**
   1. **Cluster 1**

Cluster 1 has mean price of 49.08 dollars for two, avg. rating of 3.26, avg. coronavirus cases of 65.20. Haight Ashbury, Glen Park, Castro/Upper Market

* 1. **Cluster 2**

Cluster 2 has mean price of 50.00 dollars for two, avg. rating of 3.10, avg. coronavirus cases of 810.00. The Mission neighborhood was in this cluster.

* 1. **Cluster 3**

Cluster 2 has mean price of 45.00 dollars, avg. rating of 1.28, avg. coronavirus cases of 244.67. South of Market, Excelsior and Outer Mission are neighborhoods in this cluster.

1. **Conclusion**

The purpose of this project was to explore the different food outlets in the different neighborhoods of San Francisco keeping in mind rating and coronavirus cases. Data have been put together from the Foursquare API, the Zomato API and the San Francisco Government Data (Covid and Health Inspections).

After merging data from the different sources, we have then performed Exploratory Data Analysis to reveal a few interesting observations and we have also performed K Means clustering on the dataset based on rating, avg. price, location, coronavirus cases in the Neighborhood in which the outlet is present, whether it provides online delivery and the total reviews present for that food outlet.

The clustering exercise revealed that restaurants in San Francisco and some in Haight Ashbury, Chinatown (can you believe that one), North Beach, The Financial District, Glenn Park and the Castro neighborhoods are safe to order from and also provide the best quality food while restaurants in Mission and Tenderloin neighborhoods I would hesitate to venture into. Based on the visitor's choice of rating, price and other requirements, people can choose amongst different food outlets located in these three clusters.