Applied Data Science Capstone

Final Project Report

By Ted Lin

May 28, 2020

INTRODUCTION

The current COVID-19 pandemic, also known as coronavirus pandemic, has been ravaging the world recently. As of this writing, over 5.6 million cases worldwide have been confirmed with approximately 355,000 death. In U.S., nearly 1.7 million cases have been confirmed with over 100,000 deaths[1].

Despite being a diverse metropolitan city with its economy supported by industries such as technologies, conventions, and tourism, , the city/county of San Francisco only has slightly over 2,400 confirmed cases with 40 deaths[2]. Early decisions by the city to close non-essential businesses and impose social distancing had played a major factor in its relatively low case counts.

This analysis attempts to see if any correlation can be found between COVID-19 case counts and Foursquare location data in San Francisco. Specifically, we will look at estimate COVID-19 case counts per 10K population by zip code, and attempt to correlate those with venue found within those zip codes from Foursquare.

If any correlation can be found, audiences such as city/county administrators, public health officials, and general public can expand on the finding to learn how to utilize such correlation to predict and prevent future outbreak of diseases and other public health issues.

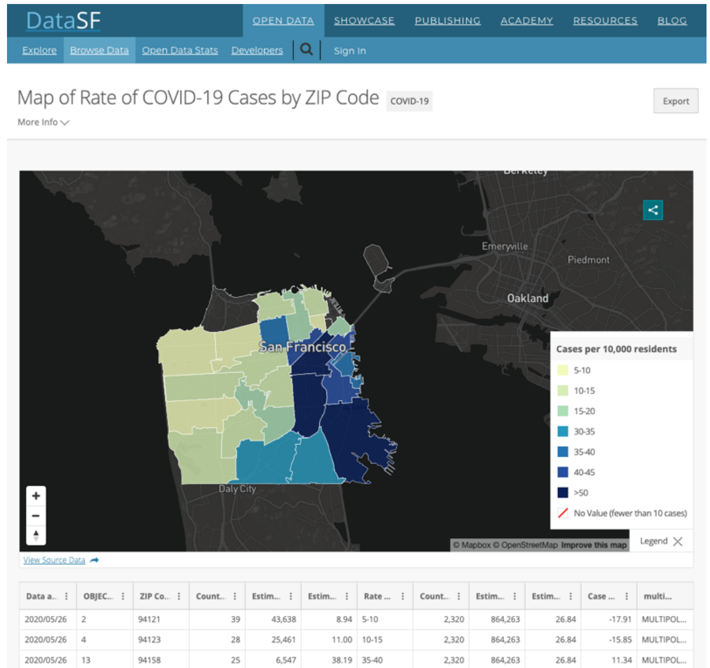
DATA

Data Sources

Data from this analysis mainly came from two main sources.

First source was "Rate of COVID-19 Cases by Census ZIP Code Tabulation Area" published by DataSF (https://datasf.org/). DataSF is part of City and County of San Francisco. A description of the data set, including how the data set is created, how it is updated, definition of data field, and preview of the data set can be found on it website (see link [here]()).

The website hosts a well-presented map visualization with data table (see image to the right; to see the actual website, see link here). Because this page is dynamically updated with embedded JavaScript, it was a hard to use Python web-scraping tool (BeautifulSoup) to gather the data table. Fortunately, the website provides links to download the data set in various format, including CSV.



The data sets included 27 rows and 12 columns. The 27 rows represented 27 zip codes in San Francisco, while the 12 columns contained data such as zip codes, count of confirmed cases (in a zip code), and count of San Francisco confirmed cases (for the entire city/county). We mainly utilized two columns of data:

- Zip code is the postal code assigned by United States Postal Services (USPS) and represents geographical boundaries.
- Rate groups are segmentation of estimated cases per 10K population; for example, if the estimated cases per 10K population is 8.54, it will be categorized as “5-10” in rate groups columns. The segmentation is in increment of 5 (“0-5”, “5-10”, …etc.) with any estimated case count per 10K population greater than 50 categorized in “>50” segment. (*Estimated* case count per 10K population is used in the data set as the population by zip code data is from 2017.)

The zip code data was augmented further by adding:

- Neighborhood name(s) – we web-scrapped San Francisco neighborhood names by zip code and append them to the data set.
- Geo-coordinates – latitude and longitude of each zip code was extracted using Python’s GeoPy library and appended to the data set.

The second source of data was Foursquare. We used API codes covered in earlier modules in this class. 7 columns of data were extracted from the JSON file downloaded by the API codes; however, we mainly focused on the two columns:

- Neighborhood – neighborhood name will represent zip code.
- Venue Category – category of venue found within the zip code’s geographical area.

Data Date/Timing

For this analysis, we downloaded data from both sources on May 27, 2020. San Francisco COVID-19 case count by zip code data set reflected its case count data as of May 26, 2020, as the website updates daily with prior date's case counts.

METHODOLOGY

Data Analysis Tools

We utilized Python programing language in Jupyter Notebook. In addition, the following Python libraries were used to assist in data extraction, analysis, and modeling:

- Data manipulation and mathematical calculation
  - pandas (https://pandas.pydata.org/)
  - numpy (https://numpy.org/)

- Data importing, web scraping, and geo-encoding
  - requests (https://requests.readthedocs.io/en/master/)
  - beautifulsoup (https://www.crummy.com/software/BeautifulSoup/)
  - wget (https://pypi.org/project/wget/)
  - geopy (https://geopy.readthedocs.io/en/stable/#)

- Data normalization, machine learning, and data model evaluation
  - scikit-learn (https://scikit-learn.org/stable/)

- Data visualization
  - matplotlib (https://matplotlib.org/)
  - folium (https://python-visualization.github.io/folium/)

Jupyter Notebook containing Python scripts for this analysis is available on GitHub repository: https://github.com/tedlin1/Coursera_Capstone/blob/master/Capstone_Project_Final.ipynb

Data Cleansing & Wrangling

More often than not, main data sets imported from sources are not ready for analysis and modeling. They may contain too much data, not enough parameters, or missing data elements. Therefore, initial data cleansing and wrangling is often necessary to prepare and format data sets for analysis and modeling steps.

Below is a quick summary of data cleansing and wrangling steps taken for this analysis. For detailed cleansing and wrangling techniques and progressions, please see codes in GitHub repository (link available in Data Analysis Tools section above).

As mentioned above, data set containing San Francisco COVID-19 case counts by zip code was downloaded from DataSF in CSV format. It contained 27 rows and 12 columns. A preview of the data set downloaded into a pandas dataframe is shown here:

	Data as of	OBJECTID	ZIP Code	Count of Confirmed Cases	Estimated 2017 ACS Population	Estimated Rate of Cases per 10k	Rate Groups	Count of San Francisco Confirmed Cases	Estimated 2017 ACS San Francisco Population	Estimated Rate of San Francisco Cases per 10k	Case Rate Difference from San Francisco	multipolygon
0	2020/05/26	2	94121	39.0	43638	8.94	5-10	2320	864263	26.84	-17.91	MULTIPOLYGON (((-122.48542599984555 37.7898249...
1	2020/05/26	4	94123	28.0	25461	11.00	10-15	2320	864263	26.84	-15.85	MULTIPOLYGON (((-122.45005999994794 37.8024729...
2	2020/05/26	13	94158	25.0	6547	38.19	35-40	2320	864263	26.84	11.34	MULTIPOLYGON (((-122.3836959998312 37.75470099...
3	2020/05/26	18	94107	122.0	29920	40.78	40-45	2320	864263	26.84	13.93	MULTIPOLYGON (((-122.38530302568738

The following steps were taken to clean and prepare this data set:

- Eliminated rows containing no COVID-19 case counts (as they were determined to be insignificant)
- Eliminated certain columns that were not needed for analysis
- Augmented integer segmentation for COVID-19 case count groups
- Added neighborhood names for each zip code
- Added geo-coordinates (latitude and longitude) for each zip code

After cleaning and formatting, data frame for San Francisco COVID-19 case counts by zip code contained 23 rows and 10 columns. A preview is shown here:

	Data as of	ZIP Code	Count of Confirmed Cases	Estimated 2017 ACS Population	Estimated Rate of Cases per 10k	Rate Groups	group_id	Neighborhood	Latitude	Longitude
0	2020/05/26	94121	39.0	43638	8.94	5-10	1.0	Outter Richmond	37.778591	-122.492289
1	2020/05/26	94123	28.0	25461	11.00	10-15	2.0	Marina District/Cow Hollow	37.801901	-122.430807
2	2020/05/26	94158	25.0	6547	38.19	35-40	7.0	Mission Bay	37.769982	-122.386828
3	2020/05/26	94107	122.0	29920	40.78	40-45	8.0	Portrero Hill	37.782740	-122.392789
4	2020/05/26	94118	39.0	41417	9.42	5-10	1.0	Richmond District	37.775515	-122.457818
5	2020/05/26	94114	40.0	34561	11.57	10-15	2.0	Castro	37.761403	-122.435242

Using the list of geo-coordinates by neighborhood from the data set above, we downloaded venue data (with limit of 100 venues within approximately 1 mile circumference) from Foursquare API. The initial data download contained 1,769 rows and 7 columns. A preview of that data set downloaded into a pandas dataframe is shown here:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Outter Richmond	37.778591	-122.492289	Pacific Cafe	37.779782	-122.494428	Seafood Restaurant
1	Outter Richmond	37.778591	-122.492289	Kufu-ya Japanese Restaurant	37.779641	-122.494581	Japanese Restaurant
2	Outter Richmond	37.778591	-122.492289	Pagan	37.781520	-122.493383	Burmese Restaurant
3	Outter Richmond	37.778591	-122.492289	Cassava	37.775722	-122.496702	New American Restaurant

The following steps were taken to clean and prepare this data set:

- Transformed data into a table with venue categories as columns and neighborhoods as rows
- Grouped neighborhoods and recalculated each venue categories as statistical mean for each neighborhood

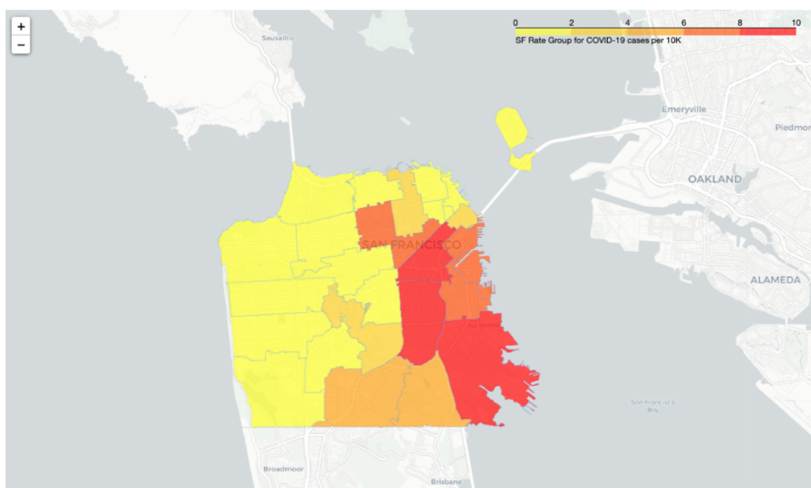
The resulting data set contained 23 rows and 276 columns. A preview is shown here:

	Neighborhood	ATM	Accessories Store	Adult Boutique	African Restaurant	Alternative Healer	American Restaurant	Antique Shop	Arcade	Arepa Restaurant	...	Video Store	Vietnamese Restaurant	Vineyard	Wagashi Place	Whisk Ba
0	Bayview	0.000000	0.00	0.00	0.00	0.00	0.095238	0.000000	0.00	0.00	...	0.00	0.000000	0.00	0.00	0.0
1	Castro	0.000000	0.00	0.00	0.00	0.00	0.010000	0.000000	0.01	0.00	...	0.00	0.000000	0.00	0.00	0.0
2	Chinatown	0.000000	0.00	0.00	0.00	0.00	0.020000	0.000000	0.00	0.00	...	0.00	0.010000	0.00	0.00	0.0
3	Cole Valley/Height District	0.000000	0.02	0.00	0.00	0.00	0.010000	0.000000	0.01	0.00	...	0.01	0.010000	0.00	0.00	0.0
4	Embarcadero South	0.000000	0.00	0.00	0.00	0.01	0.020000	0.000000	0.00	0.00	...	0.00	0.010000	0.00	0.01	0.0

Exploratory Analysis

Because the San Francisco COVID-19 case counts by zip code data set was small (23 rows by 10 columns), it could be quickly browsed and all data were self-explanatory. Therefore, no further data manipulation was needed to explore the data further.

Besides CSV format, DataSF also provides the same data set in other formats including GEOJSON. We downloaded the GEOJSON data format, and utilized folium library in Python to provide a geo-visualization of the COVID-19 case count grouping by zip code (see image on the right). You can also access an interactive map on my personal website here.



Since the venue category data set from Foursquare was much larger, more in-depth exploratory analysis was needed. First, we explored the initial data set downloaded from Foursquare to see if we can spot any data pattern or anomaly; then, we took the cleansed data show in section above to further analyze and determine data patterns and insights.

As mentioned above, data set from initial download contained 1,769 rows by 7 columns. Grouping the data by neighborhood, we could determine venue category count for each neighborhood. First few rows of this data is shown here:

Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Bayview	21	21	21	21	21	21
Castro	100	100	100	100	100	100
Chinatown	100	100	100	100	100	100
Cole Valley/Height District	100	100	100	100	100	100

Though most neighborhoods had 100 venue categories, some neighborhoods only have a small count of venue categories. As shown above, Bayview had 21 venue categories.

We can also see two other neighborhoods had small quantities of venue categories:

Portola:

Portola	11	11	11	11	11	11

Sunset District:

Sunset District	15	15	15	15	15	15

This may be due to the following factors:

- Though San Francisco is a small city geographically, some neighborhoods are more densely packed than others. Even though we used approximately one-mile radius to capture venue data, more dispersedly populated neighborhoods may find less venues.
- Foursquare users tend to be more affluent, and may not frequent and rate venues in neighborhoods that are economically depressed. Therefore, those areas may show less venue data.

After the data cleansing step, the data set resulted in 23 rows and 276 columns. This was still a fairly large data set to review; therefore, we refined the data down to top 10 venue categories and bottom 10 venue categories by neighborhood.

First few rows of the top 10 venue categories is shown here:

	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	Light Rail Station	Mountain	Park	Café	American Restaurant	Latin American Restaurant	Sandwich Place	Bus Station	Food Truck	Burger Joint
1	Castro	Coffee Shop	Park	Gay Bar	Thai Restaurant	Playground	New American Restaurant	Indian Restaurant	Yoga Studio	Juice Bar	Deli / Bodega
2	Chinatown	Hotel	Italian Restaurant	Coffee Shop	Café	Speakeasy	Gym / Fitness Center	Bar	Boutique	Breakfast Spot	Bubble Tea Shop
3	Cole Valley/Height District	Coffee Shop	Park	Boutique	Café	Bookstore	Garden	Thrift / Vintage Store	Clothing Store	Supermarket	Middle Eastern Restaurant
4	Embarcadero South	Coffee Shop	Gym	Burger Joint	Art Gallery	Gym / Fitness Center	Museum	Café	Mediterranean Restaurant	Hotel	Food Truck

You can see the bottom 10 venue categories by viewing the Jupyter Notebook in GitHub repository.

These 2 ranking data sets still did not show us much pattern or insight. Therefore, we further ranked the appearance of each venue categories in top and bottom 10 venues; another word, we ranked venue categories by how many times they showed up in “1st Most Common Venue” column in descending order, and repeated this for all other columns. For 10 most common venues, here was the result (showing top 10 frequently-appeared venue categories):

	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	Coffee Shop	Park	Coffee Shop	Café	Playground	Bakery	Sandwich Place	Mediterranean Restaurant	Sandwich Place	Burger Joint
1	Chinese Restaurant	Café	Gym	Coffee Shop	New American Restaurant	Café	Pizza Place	Bus Station	Chinese Restaurant	Deli / Bodega
2	Park	Coffee Shop	Bus Station	Art Gallery	Park	Light Rail Station	Bar	Yoga Studio	Event Space	Mexican Restaurant
3	Italian Restaurant	Pizza Place	Italian Restaurant	French Restaurant	French Restaurant	Bar	Spa	Pizza Place	Supermarket	Art Gallery
4	Trail	Gym	Garden	Gay Bar	Bus Stop	Garden	Deli / Bodega	Gas Station	Juice Bar	Pizza Place
5	Hotel	Bar	Gay Bar	Dance Studio	Wine Bar	Pharmacy	Pharmacy	Indian Restaurant	Dance Studio	Food Truck
6	Bakery	Mountain	Boutique	Baseball Field	Gym / Fitness Center	Flower Shop	Indian Restaurant	Italian Restaurant	Sushi Restaurant	Middle Eastern Restaurant
7	Cocktail Bar	Theater	Vietnamese Restaurant	Baseball Stadium	Sushi Restaurant	Convenience Store	Convenience Store	Monument / Landmark	Breakfast Spot	Dance Studio
8	Food Truck	Gym / Fitness Center	Trail	Park	Chinese Restaurant	Coffee Shop	Gym / Fitness Center	Deli / Bodega	Cantonese Restaurant	Spa
9	Gym / Fitness Center	Pool	Yoga Studio	Gym / Fitness Center	Grocery Store	Scenic Lookout	Cocktail Bar	Scenic Lookout	Spa	Bubble Tea Shop

We could see that “Coffee Shop” appeared consistently in top rankings; “Bar” and “Gym/Fitness Center” also appeared frequently.

The result for 10 least common venues is shown here:

	1st Least Common Venue	2nd Least Common Venue	3rd Least Common Venue	4th Least Common Venue	5th Least Common Venue	6th Least Common Venue	7th Least Common Venue	8th Least Common Venue	9th Least Common Venue	10th Least Common Venue
0	ATM	Modern European Restaurant	Motel	Music School	Music Store	Music Venue	Music Venue	Motorcycle Shop	Moving Target	Music Store
1	Yoga Studio	Moving Target	Mobile Phone Shop	Mountain	Motel	Moroccan Restaurant	Motel	National Park	National Park	New American Restaurant
2	None	Movie Theater	Museum	Museum	Movie Theater	Music Store	Mountain	Movie Theater	Music School	Community
3	None	Miscellaneous Shop	Moving Target	Modern European Restaurant	Music School	Motorcycle Shop	Nail Salon	Nail Salon	Mountain	Movie Theater
4	None	Lake	Martial Arts Dojo	Martial Arts Dojo	Monument / Landmark	National Park	Mini Golf	Moroccan Restaurant	Community	Nail Salon
5	None	Market	Latin American Restaurant	Motel	Nail Salon	Mattress Store	Middle Eastern Restaurant	Mini Golf	Martial Arts Dojo	Mini Golf
6	None	Juice Bar	Mini Golf	Massage Studio	Sports Club	Mountain	Community	Market	Moroccan Restaurant	Market
7	None	Massage Studio	Lingerie Store	Music Store	Middle Eastern Restaurant	Museum	Movie Theater	Mac & Cheese Joint	Middle Eastern Restaurant	Massage Studio
8	None	Irish Pub	Music Store	Moroccan Restaurant	Lingerie Store	Video Store	Sports Bar	Community	Newsstand	Motel
9	None	Music School	Moroccan Restaurant	Music Venue	Massage Studio	Jewelry Store	Moving Target	Music Store	Lingerie Store	Newsstand

“ATM” and “Karaoke Bar” were the only categories in the 1st Lease Common Venue ranking; I suspect that these 2 categories showed up very infrequently in the entire original downloaded from Foursquare and therefore were outliers. Other than those two categories, categories such as “Motel” and “Moroccan Restaurant” appeared on this grid somewhat frequently.

Inferential Statistical Testing/Machine Learning

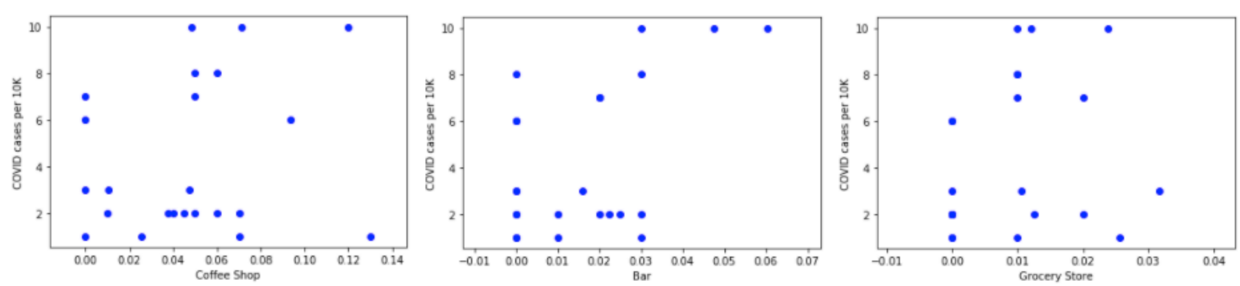
The main goal for this analysis is to determine if any correlation between COVID-19 case counts and Foursquare location data in San Francisco; therefore, statistical testing methods deployed should be those that measure correlation between data sets. For inferential statistical testing, we used two data modeling techniques: simple linear regression and K-Nearest Neighbors.

For each of the modeling techniques, we randomly selected 75% of data set as training data and the rest as testing data. We also used 2-D plots to help us visualize correlation and modeling results.

ANALYSIS RESULTS

A. Simple Linear Regression

First, we used scatter plots to quickly visualize if any linear relationship between venue categories and COVID-19 case counts can be spotted. Using “Coffee Shop”, “Bar”, and “Grocery Store” as independent variable, we generated the following three plots.



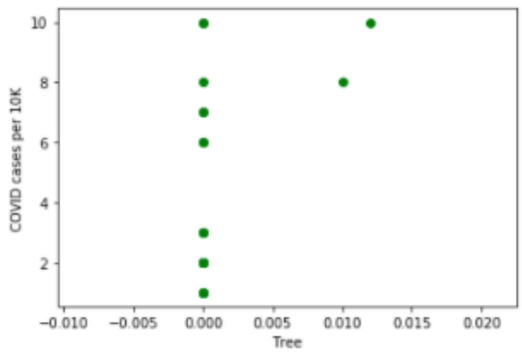
These plots did not show any linear correlations (positive or negative). We wanted to see if any other category showed correlation. However, with more than 270 categories, it was impossible to manually go through them. Therefore, codes using Python and its scikit-learn libraries were written to: 1)split data into train/test sets; 2)generate a linear regression machine learning model for each category; 3)train the model with train data set; 4)calculate Mean Absolute Error, Mean Squared Error (MSE), and R-Squared (R2-score) from model’s prediction of test data set, and; 5)store results of all categories in a dataframe. The data was then sorted by R2-score in descending order, shown here:

	Venue	Mean absolute error	MSE	R2-score
169	Mountain	3.500000	15.744792	0.104296
160	Middle Eastern Restaurant	2.976458	10.978351	0.072928
221	Scenic Lookout	4.042901	19.744010	0.072554
5	American Restaurant	3.812349	21.394497	0.070947
191	Pedestrian Plaza	2.866667	8.906667	0.000000
190	Parking	2.958333	9.420573	0.000000
108	General Entertainment	3.000000	9.666667	0.000000

The highest positive R2-score was 0.104 for “Mountain”, indicating that there was no category with meaningful positive correlation to COVID-19 case counts.

We proceeded by looking for any category with inverse correlation by searching out R2-score between -1.2 and -0.8, and the one resulting category is shown below:

	Venue	Mean absolute error	MSE	R2-score
262	Tree	2.206979	5.803123	-0.960189



Any reasonable person would question if “Tree” could really correlate with a respiratory epidemic. To examine this further, we plotted “Tree” venue category data against COVID-19 case counts, shown on the right.

The graph definitely did not substantiate inverse correlation between “Tree” and COVID-19 case counts. There were 22 neighborhoods (or zip codes) in the data set, but only 2 neighborhoods had “Tree” in their venue categories, meaning a majority of neighborhoods did not have “Tree” in their venue rankings.

Does this mean majority of neighborhoods in San Francisco do not have trees? No, it is fairly obvious that most people do not rate trees online, so it is reasonable to expect a lack of ranking data for categories similar to “Tree”.

B. K-Nearest Neighbors (KNN)

First, we determined the most frequently ranked venue categories by sorting the Foursquare data set by frequency of appearance, and put them in a list. A preview of resulting data is shown on the right.

```
Coffee Shop             93
Park                    65
Café                    55
Pizza Place             44
Gym                     34
Bakery                  33
Italian Restaurant      33
Gym / Fitness Center    32
Bar                     32
Sandwich Place          27
```

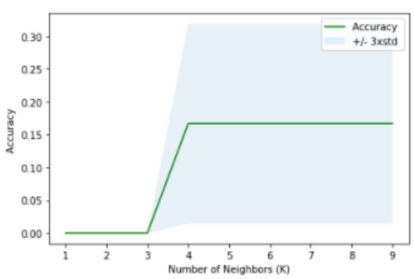
For KNN modeling, we wanted to generate 3 different scenarios with different numbers of independent variables. Specifically, we wanted to model with top 4 independent variable, top 8 independent variables, and top 12 independent variables.

For each of the 3 scenarios, we generated data set of independent variables, then used scikit-learn’s StandardScaler function to normalize the data. A preview of these data for top 4 independent variables are show on the following page.

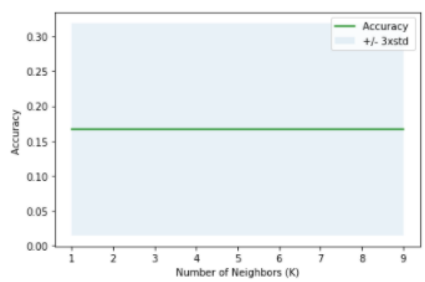
	Coffee Shop	Park	Café	Pizza Place
0	0.000000	0.095238	0.095238	0.000000
1	0.050000	0.050000	0.010000	0.000000
2	0.050000	0.020000	0.050000	0.010000
3	0.060000	0.060000	0.060000	0.020000
4	0.130000	0.020000	0.030000	0.010000
5	0.060000	0.050000	0.040000	0.050000
6	0.070000	0.020000	0.010000	0.020000
7	0.037500	0.012500	0.012500	0.062500
8	0.093750	0.125000	0.062500	0.031250

```
array([[-1.34479292,  0.8052657 ,  2.5712085 , -1.21568739],
       [ 0.07439828, -0.04325534, -0.8641893 , -1.21568739],
       [ 0.07439828, -0.60595877,  0.74795268, -0.69345349],
       [ 0.35823652,  0.14431246,  1.15098818, -0.17121959],
       [ 2.3451042 , -0.60595877, -0.05811831, -0.69345349]])
```

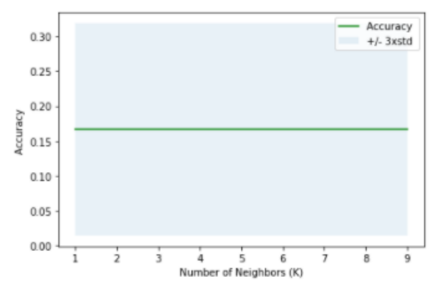
With independent variable data set created, we then split the data into train/test. With scikit-learn's KNeightborsClassifier function, we generated a KNN machine learning model. Iterating K (number of neighbors) value from 1 to 10, we trained the model with training data, and calculated the model's accuracy mean and standard deviation from the model's prediction with test data. The results of the 3 scenarios were then plotted to see which scenario and K value yielded best results. The plots are shown here:



Top 4 ind. variables



Top 8 ind. variables



Top 12 ind. variables

The highest accuracy mean was 0.1667 across all 3 scenarios. Therefore, this indicated that we could not find a KNN model providing reasonable accuracy for predicting COVID-19 case counts from venue categories.

DISCUSSION

Prior to discussing results of the analysis, we would like to remind readers that the goal of this analysis was to attempt to *see if any correlation can be found between COVID-19 case counts and Foursquare location data in San Francisco*, not to determine what Foursquare's location data correlates with the city's COVID-19 case count.

With the in mind, our results did not find any correlation between Foursquare's location data and San Francisco's COVID-19 case count by zip code. Neither set of data modeling showed any meaningful correlation between the data's independent variables (venue categories by neighborhood/zip code) and dependent variable (COVID-19 case counts by zip code).

This should not be a surprise to any reasonable reader. As we are aware, geo-location data aggregation companies such as Foursquare are in the business of selling data to customers who are mainly interested in using such data for sales and marketing purposes; therefore, geo-location data collected are typically geared toward commercial establishments, such as retail stores, tourist attractions, and restaurants/bars. Rankings of these places seldom correlate with public health statistics.

Demographic census data such as income, age, and/or ethnicity distribution may correlate with public health data in more significant and meaningful ways. Finding such data sources, cleansing and exploring that data, modeling and analyzing the subsequent modeling results can be the topic of a future project.

CONCLUSION

For this project, we performed the following steps:

- Downloaded relevant data from various websites/sources
- Wrangled the data in formats needed for analysis and modeling
- Performed exploratory data analysis and visualization
- Generated Linear Regression models
- Generated K-Nearest Neighbors models
- Analyzed the results

The goal of this analysis was to attempt to attempt to see if any correlation can be found between COVID-19 case counts and Foursquare location data in San Francisco. We analyzed the results of our data models and could not find reasonable correlation between the two data sets. As mentioned in discussion section above, this should not come as a surprise as geo-location data geared toward commercial purposes typically do not correlate with public health statistics.

REFERENECES

[1] Johns Hopkins University Corona Resource Center (https://coronavirus.jhu.edu/us-ma)

[2] San Francisco Department of Public Health (https://www.sfdph.org/dph/alerts/coronavirus.asp)