Yoga Studio in San Francisco

Anthony Ritz

Data Science Capstone Project

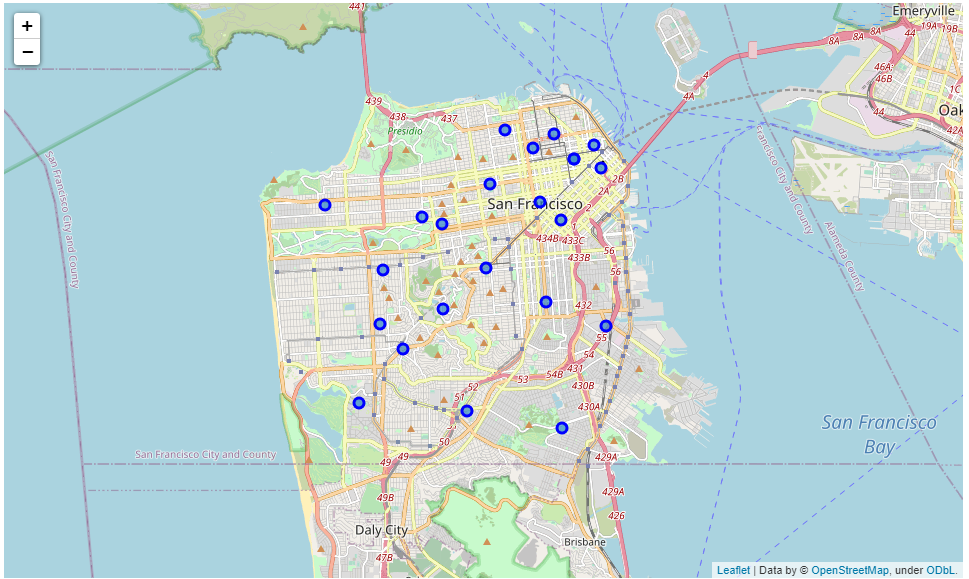
Coursera

December 14, 2020

**Introduction**

The purpose of this laboratory exercise is to determine an ideal neighborhood in San Francisco to open a new yoga studio. The goal is to determine which neighborhood in the city and county of San Francisco would be best suited to a new yoga studio during the current pandemic conditions. With most indoor activities being suspended until further notice, the most ideal location would be a neighborhood with a market and some competition for a yoga studio, as well as it being a neighborhood with access to close public parts so that that new studio could offer outdoor instruction until conditions improve. Data science tools and techniques will be used to determine this location.

**Data and Methodology**

A Python notebook is used to execute data science techniques to filter through data sets to determine the ideal location for the new yoga studio (Ritz, 2020). First, a user created JSON file is used to determine the geographic information (cdolek, 2020). The JSON file contains the neighborhood information of San Francisco. That is, it contains the names and zip codes for each of the neighborhoods. Once the data for the neighborhoods is input, a Python Geolocator function can be used to retrieve geographical data on each zip code and neighborhood. A map below (figure 1) shows the 22 neighborhoods of San Francisco that are a part of this report. This information is then used in conjunction with the Foursquare API to find all venues in each neighborhood. The Foursquare data is processed using data science and machine learning methods to find the number of yoga studios in each neighborhood. Likewise with the number of yoga studios, those neighborhoods that contain the desired number of potential competitors are also screened for parks within the neighborhoods.   
  
 Figure 1: Central Coordinates of Neighborhoods within San Francisco

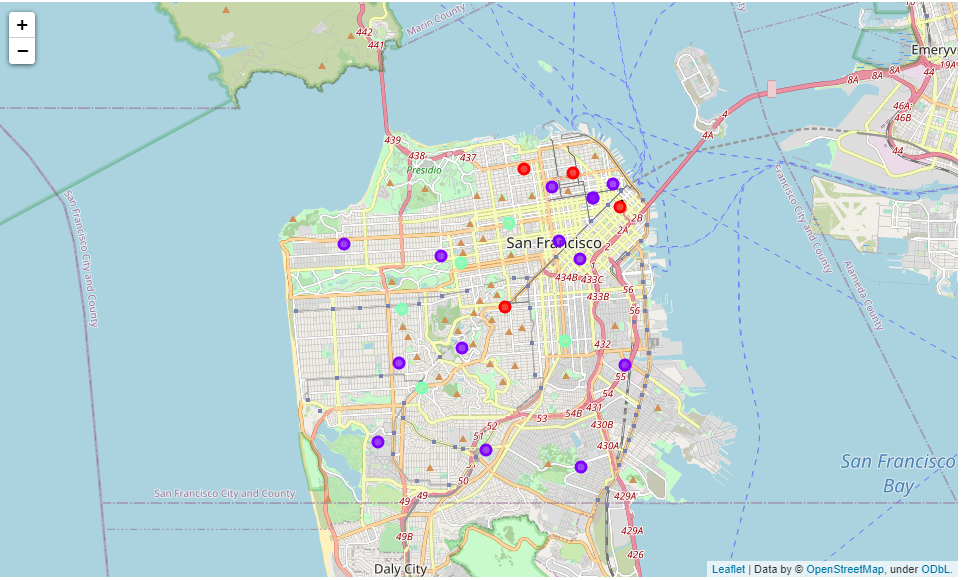
**Results**

The data for each of the neighborhoods, differentiated by zip code, is processed and information is retrieved on the venues for each region. The number of yoga studios in each region is used as the primary factor of a k means clustering function. The neighborhoods are divided into three clusters using this method. Results are shown in table 1, and the locations of each neighborhood are displayed on the map of figure 2. Additionally, table 1 contains the number of parks for each region.

**Table 1. Cluster data for Yoga Studio Data**



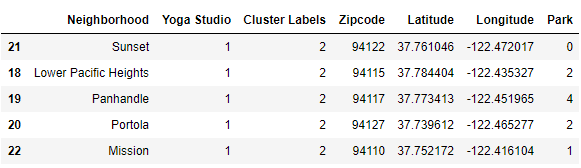
**Figure 2: Map of Each Neighborhood with Cluster Data.**



**Discussion**

The data and techniques from above give three clusters of regions to pick an ideal location for the new yoga studio. The largest group, in purple in figure 2, contains the majority of the results. These regions contain zero yoga studios. These would likely not be adequate for a new yoga studio as there is not competition and perhaps no market for such a facility. The red regions are cluster zero, they contain the most existing studios, and may likely be an overly competitive market. Regions from cluster 2, the green neighborhoods from figure 2, are the chosen neighborhoods. These regions have existing yoga studios indicating a potential market, but not so many as to be overly competitive. The names, zip codes and geographic information are singled out and presented in table 2.

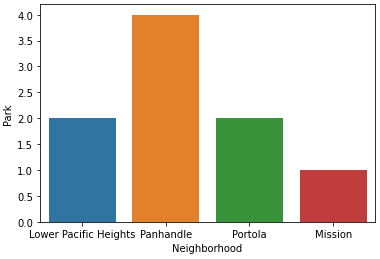
**Table 2: Neighborhoods from Cluster #2.**



One of the neighborhoods can be selected from cluster #2 by the number of parks in the region.

Figure 3 shows the number of parks for each region excluding Sunset as it has zero parks.

The neighborhood called Panhandle has 4 parks and one yoga studio. It would be the most ideal location for a new yoga studio in San Francisco.



**References**

Cdolek. (2020) Github user: cdolok. SanFrancisco.Neighborhoods.json. Retrieved from https://gist.github.com/cdolek/d08cac2fa3f6338d84ea

Ritz, A (2020) Github user: aanthonymritz. Coursera Capstone.ipynb. Retrieved from https://github.com/anthonymritz/Coursera\_Capstone/blob/master/Coursera%20Capstone.ipynb