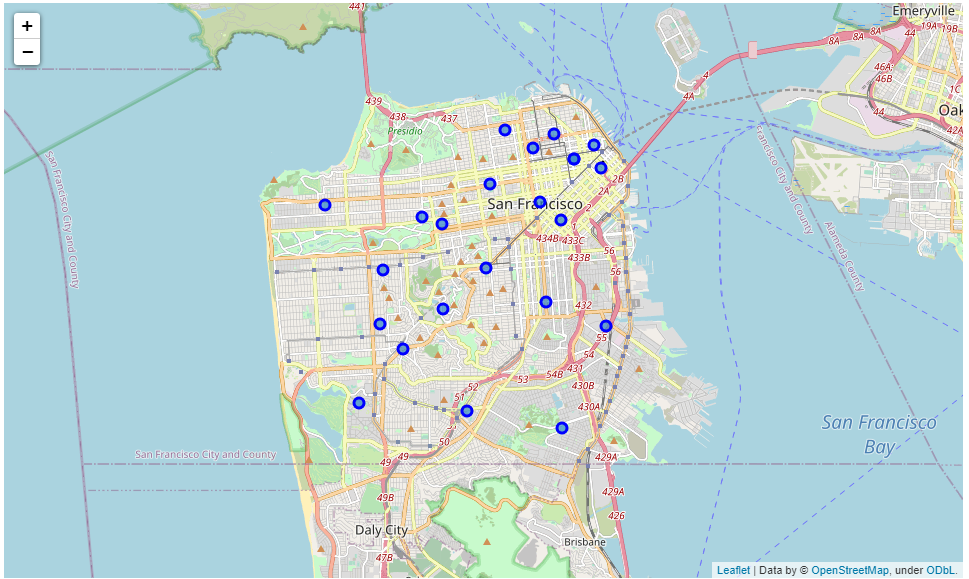
Selecting the Best Neighborhood in San Francisco for a New Yoga Studio

Looking for the ideal location to open a business can be a challenge. This method provides some data science techniques that can help.

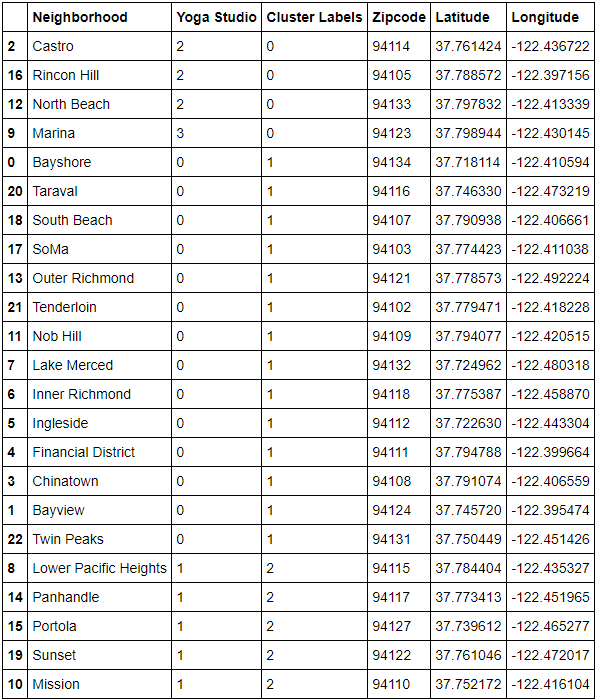
This is the approach that allowed for the selection of the best neighborhood in San Francisco to set up a new yoga studio during the pandemic.



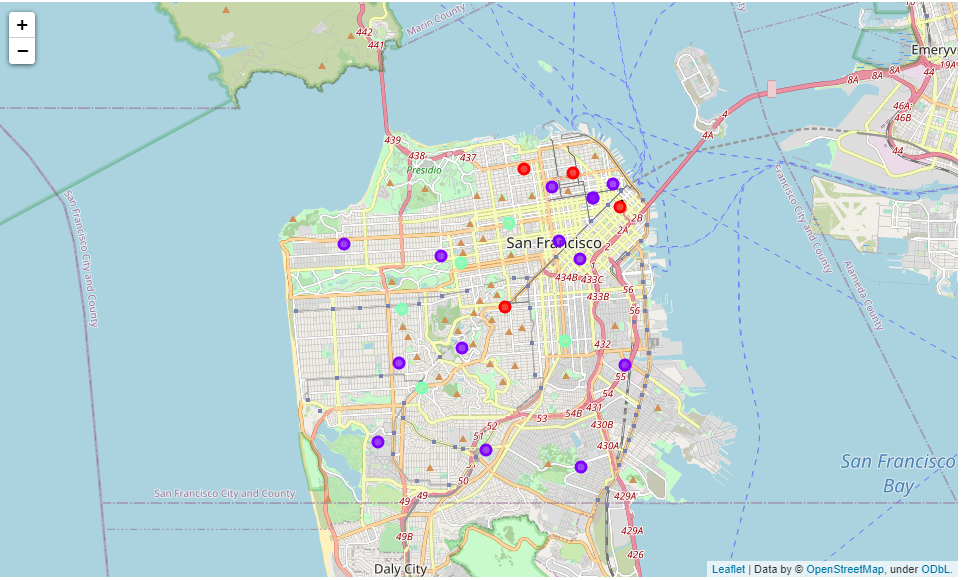
1. Introduction - Data science and machine learning techniques are used to determine the best location to open a new business. In this case, the business is a new yoga studio in the city of San Francisco. Additional conditions for the business includes nearby access to outdoor public areas (parks) to allow classes and instruction to take place during the current pandemic conditions and beyond.
2. Data and Methodology – Several steps of data processing were performed to get the results and to make the decision of the preferred neighborhood.
3. A user created JSON file was found on github.com that contained the zip codes and neighborhood names for each neighborhood.
4. The zip code information was used in conjunction with a geolocator function in a python script to determine the specific latitude and longitude for each neighborhood.
5. This data allowed for the creation of a map of each of the 22 neighborhoods using folium.



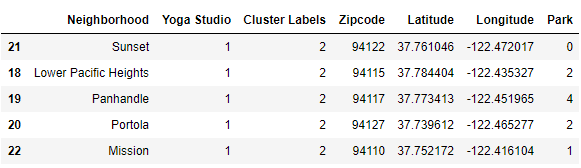
1. The Foursquare API was then used to get available venue data for each neighborhood.
2. This data was cleaned and processed with machine learning techniques (k means clustering) to find the number of yoga studios in each region and cluster them. This data is presented in tabular form.

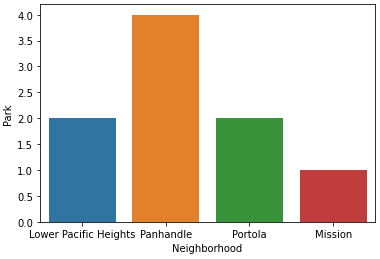


1. A second folium map was made showing the cluster 1,2, and 3.



1. Results – The business should be created in a region that has a market for a yoga studio. Regions with zero existing studios are ruled out as potential locations. These locations are classified as cluster one and a shown in purple on the map. Furthermore, cluster number 0 has the most regions and the most competition, and they have been ruled out as well. The ideal cluster is cluster number 2. Each region has only one existing yoga studio. Additional data for cluster one was then retrieved from the venues listed from the Foursquare data to find the number of parks in this region.





1. The Panhandle Region of San Francisco, with a zip code of 94117, has one existing yoga studio, but it is in the same neighborhood as 4 public parks. It would be the best region to set up a new studio in these current conditions based upon this data.