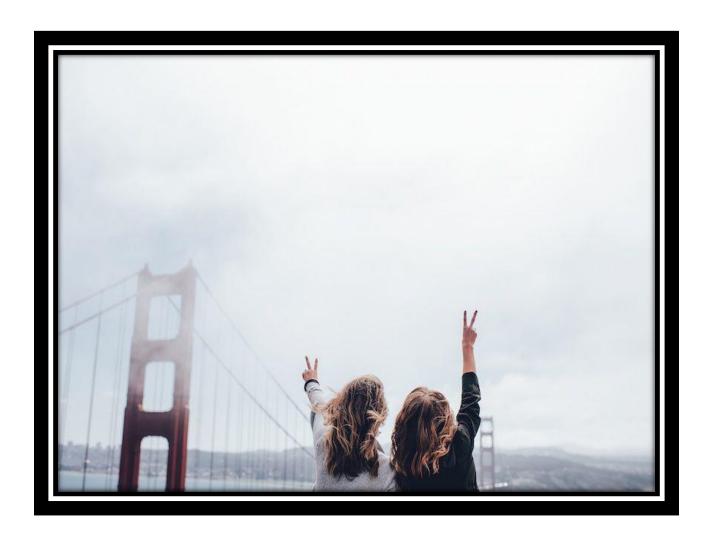# Capstone Project

The Battle of Neighborhoods (Week 1)

Neighborhood to move in SF Bay Area for Millennials

Data

There are four main types of data we will need for this project:

1. SF Neighborhood Data
2. SF Diversity Data
3. SF Rental Data
4. Things to Do Data

SF Neighborhood Data

For the SF Neighborhood, we will need three types of data:

1. Longitude/Latitude
2. Zip Codes
3. Neighborhood Names

We will get the Zip Code and Longitude/Latitude data from geonames.org. They have San Francisco data on the following webpage https://www.geonames.org/postal-codes/US/CA/075/san-francisco.html

Once we have that data, we will have to merge it with the neighborhood data which is made available by www.healthysf.org on their webpage http://www.healthysf.org/bdi/outcomes/zipmap.htm

SF Diversity Data

For this project, we will consider diversity as the average population of a region divide by the maximum population for a given race. The larger this number, more diverse the neighborhood.

Please Note: This theory fails if only two races live in the neighborhood with equal population. However, this type of demography doesn't exist in a diverse area like San Francisco

To calculate this information, we will take the census data from "Bay Area Census". This information is available at http://www.bayareacensus.ca.gov/small/PL94-171CensusTracts.xls

This data is available by TRACT which is a subsection of the zip code. By attaching "94" as prefix to this, I can get the census by zip code and merge it to SF Neighborhood data

SF Rental Data

Government maintains median rent for a given zip code for the whole country. This data is available at https://www.hud.gov/sites/dfiles/Housing/documents/2020MedianGrossRentbyZipCode.xlsx

Since the data is available at Zip Code level, it can be easily joined with the SF Neighborhood data

Things to Do Data

To get this information, I will be using the Foursquare API. Since I have the neighborhood, longitude, latitude data available in "SF Neighborhood Data" I will be able to get the "venue Category" from the explore function.

I can further count the number of "Venue Category" to calculate things to do in that Neighborhood