**Finding the Perfect Place to Live**

**Introduction/Business Problem**

A group of college kids would like to find new places to live in the city of San Francisco, Ca since it is currently a favorable market and they have been quite successful with certain investments. They would like us to populate a list of at least 3 potential locations. The new living location must meet the following criteria:

1. Be within 2 miles of public transportation
2. A location with a low crime rate.
3. Be in a lively location with heavily frequented coffee shop and bar.

The main objective is to identify 5 potential locations that meet the above criteria and provide the client with the Top 3. San Francisco is a very diverse location with a plethora of “good attributes” but it also has quite a few issues such as concentrations of high crime, hordes of homeless people, and significant economic disparity.

Ideally, the apartment/condo/townhouse would be located in a neighborhood where its residents can afford to live and work but there should still be to public transportation for convenience.. Areas that have a lot of successful business’ tend to be more populated, diverse, and ideal communities so this is a good marker for a location.

**Data**

The Foursquare location data will be used to identify areas of successful business, based on frequency of visits and reviews, as well as locations of any “Bar” and “coffee shop.”

Crime statistics data from <https://s3-api.us-geo.objectstorage.softlayer.net/cf-courses-data/CognitiveClass/DV0101EN/labs/Data_Files/Police_Department_Incidents_-_Previous_Year__2016_.csv> will be used to determine ideal locations with lower crime rates. A choropleth map will be used to highlight neighborhoods with lower crime data. A JSON file from <https://cocl.us/sanfran_geojson> will be used to determine major landmarks and will be used to create the chloropleth map.

Figure 1: Map of San Francisco

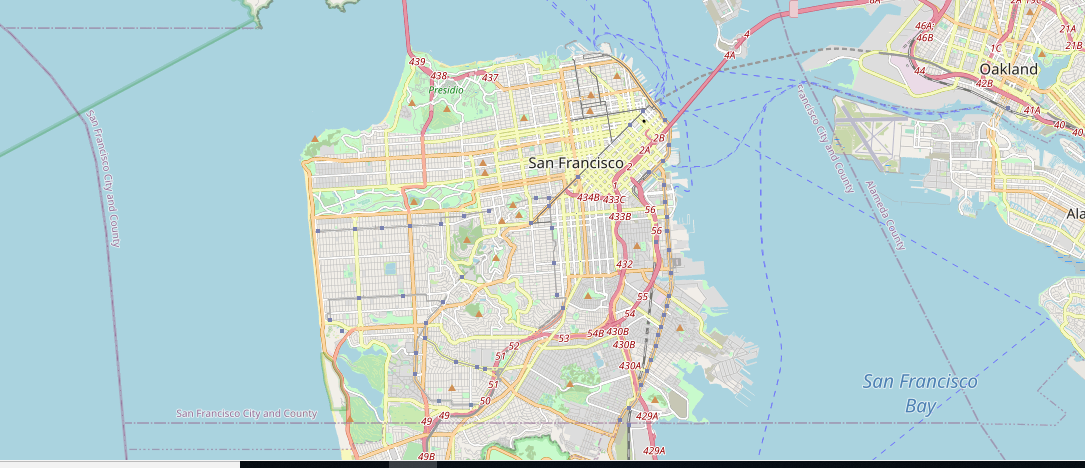


Figure 2: Cloropleth Map of San Francisco Crime Data

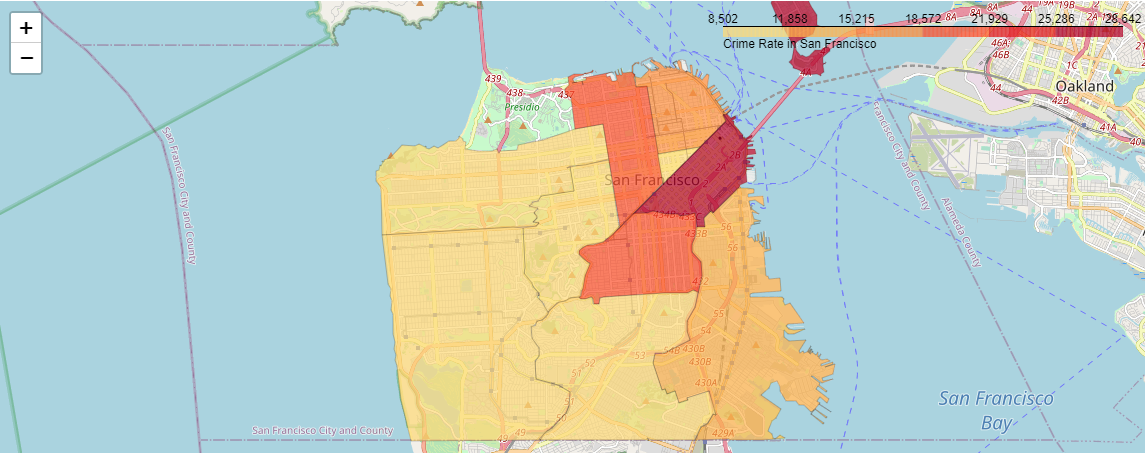


Table 1: Pandas Dataframe of Baltimore City Crime Data by Neighborhood

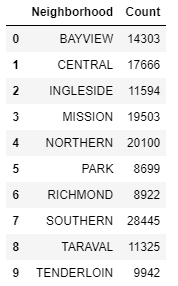


Table 2: Venue Dataframe Generated using Foursquare API



Jupyter Notebook for Data Compiling and analysis

<https://dataplatform.cloud.ibm.com/analytics/notebooks/v2/05d505bb-03da-40fa-a7d4-9c0c66d9da7d/view?access_token=a5a4aa7af1672ca1c8e5895f00d42cf2ca1b53e6e799db2ad7f8fc0f90ab99f0>

**Methodology**

**Determining the Safety of a Community**

Using the Crime Statistics data I was able to group the crimes by Neighborhoods in order to overlay it with a GeoJSON file outlined with Neighborhood Boundaries. This will visually make it easier to narrow down a neighborhood. The Number of crimes per neighborhood were represented b colors on the map. The darker the shading the more crime was present. The crime statistics dataframe used for this can also be examined to further assess the crime activity of each neighborhood.

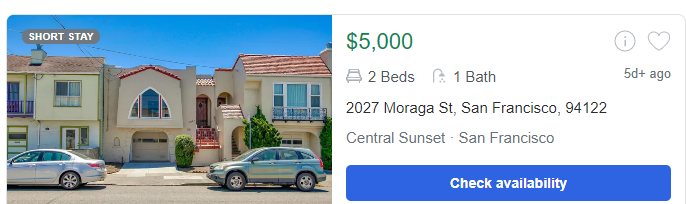
**Determining the Location and Success of Certain Business**

The easiest method for determining the success of neighborhood business was to use the Foursquare API developer app to call for the venue data for San Francisco Neighborhoods and assess which neighborhoods had a Bar and a Coffee shop as well as at what frequency the top 10 establishments were visited. If within the top 5, a bar and a coffee shop were listed it was determined the neighborhood met the desired criteria. If the top 3 venues had a frequency of at least 0.02, the neighborhood was considered “busy” and thus desirable.

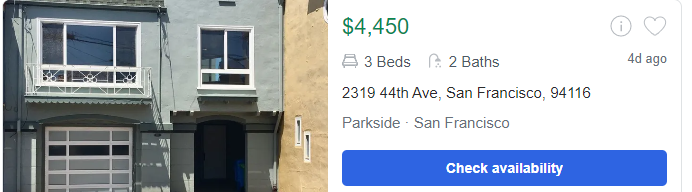
**Results**

Using the data and methodology outlined in this report and accompanying notebook it was determined that the Park neighborhood of San Francisco was the best fit for the client. Once the appropriate neighborhood was determined, a search of current real estate listings revealed 5 locations that met the client’s requirement and fell well within their budget of less than or equal to $5,000/month. They expressed a keen desire for renting but did not want to rule out an outright purchase. So rental properties took precedent.

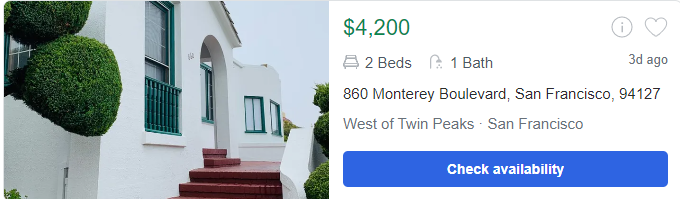
**Property 1**



**Property 2**



**Property 3**



**Discussion**

This method of determination can be used to streamline a variety of client desires and requirements. Investors that are looking to open restaurants, build new condos, or open a sports complex could use a similar method. Any area can be examined as well as long as the appropriate open source data sets exist or the client is willing to pay for the data set if it is available. Different types of maps could also be used to highlight the crime data and the venue data could also be superimposed onto the map of the crime data if needed. In this case, the requirements and the amount of data was limited and easily discernible.

**Conclusion**

The Park neighborhood of San Francisco is the best fit for the client’s desires and needs. Other neighborhoods would also be an option if the properties identified were not, once evaluated by the client, to their liking. There are many options and many tweaks that can be made to the data and/or method in order to comply with the desires of the client.