Introduction and Business Problem

San Fransisco

The idea about relocating to a different city is always terrifying. People need to research a lot before actually deciding a place to stay in city where you have never been to. The research basically means a variety of information needed to be gathered. This could be area, region, business, position size, among several other considerations like community study. This can be termed as a search algorithm request that usually returns the required features such as population rate, median house price, school ratings, crime rates, weather conditions, recreational facilities etc.

Getting an application which could render things simple by providing a comparative study with given variables within the locality would be helpful and good.

The user can use this project when renting apartment or buying house in a locality based on the distribution of various facilities available around the neighborhood. As an example, this project would compare 2 randomly selected neighborhoods and analyze the top 10 most common locations in each of those two neighborhoods based on the number of people visiting each of those locations. This project also uses the K-mean clustering of unsupervised machine learning algorithms to cluster locations based on location categories such as restaurants, parks, coffee shops, gyms, clubs etc.

DATA

Sources of Information:

FourSquare API:

This API has a database of more than 105 million places. This project would use Four-square API as its prime data gathering source. Many organizations are using to geo-tag their photos with detailed info about a destination, while also serving up contextually relevant locations for those who are searching for a place to eat, drink or explore. This API provides the ability to perform location search, location sharing and details about a business. Foursquare users can also use photos, tips and reviews in many productive ways to add value to the results.

HTTP requests would be made to this Foursquare API server using zip codes of the San Fransisco city neighborhoods to pull the location information (Latitude and Longitude).Foursquare API search feature would be enabled to collect the nearby places of the neighborhoods. Due to http request limitations the number of places per neighborhood parameter would reasonably be set to 100 and the radius parameter would be set to 700.

Folium- Python visualization library would be used to visualize the neighborhoods cluster distribution of Chicago city over an interactive leaflet map.Extensive comparative analysis of two randomly picked neighborhoods world be carried out to derive the desirable insights from the outcomes using python's scientific libraries Pandas, NumPy and Scikit-learn.

• Unsupervised machine learning algorithm K-mean clustering would be applied to form the clusters of different categories of places residing in and around the neighborhoods. These clusters from each of those two chosen neighborhoods would be analyzed individually collectively and comparatively to derive the conclusions.

Packages:

Pandas - Library for Data Analysis

- NumPy – Library to handle data in a vectorized manner
- JSON – Library to handle JSON files
- Geopy – To retrieve Location Data
- Requests – Library to handle http requests
- Matplotlib – Python Plotting Module
- Sklearn – Python machine learning Library
- Folium – Map rendering Library