The Battle of Neighborhoods – Week 1

Problem Background:

For this project, the chosen location is Manhattan, NY in United States. Manhattan is the most densely populated of the five boroughs of New York City. Manhattan Island is divided into three informally bounded components: Lower, Midtown, and Upper Manhattan. Manhattan has been described as the cultural, financial, media, and entertainment capital of the world. It boasts of several world renowned tourist attractions which draws travelers from all across the world.

Airbnb is an online marketplace where members can use the service to arrange lodging or homestays for short or long durations. Airbnb properties exist in many countries and Manhattan has a lot of options for stay using Airbnb.

Problem Description:

Domestic and International travelers who travel for work or vacation generally tend to stay closer to the places of interest and usually have little or no knowledge regarding the location and place of stay. Also, as with any tourist city in the world, the crime rate in Manhattan is usually high.

One of the major problems faced by travelers during the course of stay in a new city is their personal safety. While being in the proximity of the points of interest, is it possible to come up with a solution wherein a safer neighborhood can be determined within the available Airbnb properties using the past crime data?

With the available data, this project is attempting to find correlation between the locations of Crime, with the geo locations of Airbnb properties along with Foursquare data of Manhattan listing out the points of interest. This assignment examines and analyzes the data using tools available in Python to explore the following:

1. Demographics of Airbnb listings
2. Exploratory analysis with trends in crime in Manhattan
3. Access to nearby places of interest

Combining all the above, it is attempted to find the correlation between the listed Airbnb properties, available Crime data and points of interest (using FoursquareAPIs) using several Machine Learning algorithms which are applicable (once the data analysis and visualization is complete)

Data:

For solving the above problem, following data is used:

1. http://tomslee.net/airbnb-data-collection-get-the-data - Details of Airbnb listings (in .csv format) with their location coordinates and ratings included for the City of New York.
2. https://data.cityofnewyork.us/Public-Safety/NYC-crime/qb7u-rbmr - New York City crime data acquired from NYC Open Data website (.csv file) which includes the type of crime, location of crime including the co-ordinates is used to overlay on the Map of Manhattan using Folium Libraries in Python as per the available data.
3. Data set for New York Neighborhood derived from https://geo.nyu.edu/catalog/nyu_2451_34572 - .json file is used to map the locations using geopy and folium libraries with neighborhoods superimposed
4. Exploring cuisines of New York city using https://en.wikipedia.org/wiki/Cuisine_of_New_York_City - web scraping to analyze demographics, cuisines using BeautifulSoup4 library (parsing data from html pages)
5. Foursquare APIs – New York city geographical coordinates are used as input, Foursquare APIs are made to acquire top venues in the prescribed radius
6. Combining and overlaying above data on a folium map, available Airbnb properties in a safer neighborhood which are closer to the places of interest is derived.