Yuting Guo Assignment 3

Analytic Goals: used clustering method to analyze and visualize crime report dataset for user knowing how safety in different areas-Los Angeles. It clusters the total number of crimes occur in one area and shows it by region on the map. The heatmap and folium map is interactive, it will highlight the areas with high regional crime rate.

Libraries Used: pandas, folium (https://python-visualization.github.io/folium/#) and HeatMap (https://python-visualization.github.io/folium/plugins.html).

Dataset: Crime data in LA from 2011-2017 (https://www.kaggle.com/cityofLA/crime-in-los-angeles)

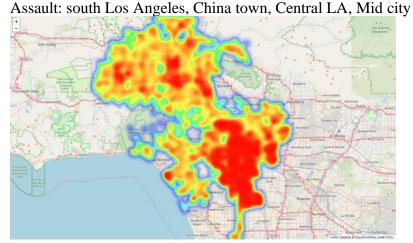
Parameters and Options:

During the data cleaning, I used Jupyter Notebook to drop the unknow location. In order to fit to folium HeatMap, I have to split the location from one column into latitude and longitude two columns and group the crime data category into 'ASSAULT', 'WEAPONS', 'THEFT', 'VEHICLE', 'CRIMES AGAINST CHILDREN'.

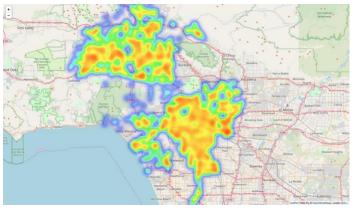
The difficulties were the dataset have street location rather than region attributes, that is the main reason I choose to use HeatMap instead of Choropleth maps. Fortunately, the data set include latitude and longitude. On the other hands, grouping crimes together took a lot of time because the crime description had various types to figure out.

The limitations were due to the large dataset have 1584316 rows, it will show ineffective visualization due to the large dataset, I have to narrow down the time range between the near year, 2016-2017. In order to showing better visualizations, I also separated into those five crime categories as mentioned above.

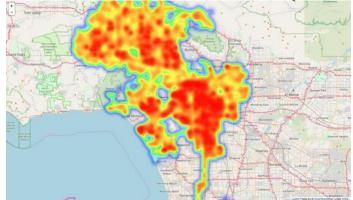
I will list the high incidence area in different crime data type based on the visualization.



Weapons: Downtown, Boyle Heights, Van Nuys, San Pedro,



Theft: Downtown, China town, Venice, Hollywood



Vehicle: Central LA, Downtown, South LA



Crimes Against Children: Panorama City, Mid City, South LA

