W200 Fall 18 | Project 2 | Group Project

Team Members: Jonathan Hilton, Thanh Le, Eddie Zhu

Project 2 Proposal

Github Repo: <https://github.com/UCB-INFO-PYTHON/W200Fall18_GroupProject_Jon-Thanh-Eddie.git>

For this project, our team decided to work on analyzing potential relationship between City of Chicago crimes dataset and City of Chicago traffic ticket dataset. We are also planning to use the Chicago demographic info for further analysis.

We have 3 datasets to explore.

1. Chicago Traffic Violation Dataset

This dataset contains the following columns, from which we will use violation\_location, zipcode and ticket\_queue\_date.

Columns: [ticket\_number, issue\_date, violation\_location, license\_plate\_number, license\_plate\_state, license\_plate\_type, zipcode, violation\_code, violation\_description, unit, unit\_description, vehicle\_make, fine\_level1\_amount, fine\_level2\_amount, current\_amount\_due, total\_payments, ticket\_queue, ticket\_queue\_date, notice\_level, hearing\_disposition, notice\_number, officer, address]

1. Chicago Crime Dataset

This dataset contains the following columns, from which we will use Date, Location Description, and Location.

Columns: [ID, Case Number, Date, Block, IUCR, Primary Type, Description, Location Description, Arrest, Domestic, Beat, District, Ward, Community Area, FBI Code, X Coordinate, Y Coordinate, Year, Updated On, Latitude, Longitude, Location]

1. Chicago Census Dataset

This dataset contains the following columns, from which we will use Zip code, Calculated population growth, Population density, Estimated median household income and unemployment rate.

Columns:[Zip code, Estimated zip code population in 2016, Zip code population in 2010, Zip code population in 2000, Population density, Estimated median household income, Unemployment rate]

We will attempt to explore the correlation between traffic violation occurrences and non traffic crime occurrences in the same locations and time frame. We will also explore the correlation between both crime rates and demographics at given locations. The result will determine community ‘hot spots,’ where both criminal activity and traffic incidents occur and forecast future potential occurrences at given locations, which will help the city strategically deploy and allocate police forces across the city to reduce both traffic violations and non traffic crimes.

We will most likely use numpy and pandas to extract the information from the csv files and use matplotlib and seaborn to analyze and visualize the findings. The final reports will first present the basic information about crime occurrences by location and trend, and traffic violations by location and trend, along with demographic information by location. Then, It will present the correlations we found between the independent variables. Finally, we will make recommendations on actionables based on the findings from our analysis, which will hopefully help Chicago police force strategically fight the crimes and traffic violations more effectively without additional investments.