sYRACUSE

IST-652 FINAL PROJECT

justin pate

**SUID: 445414259**

**jdpate@syr.edu**

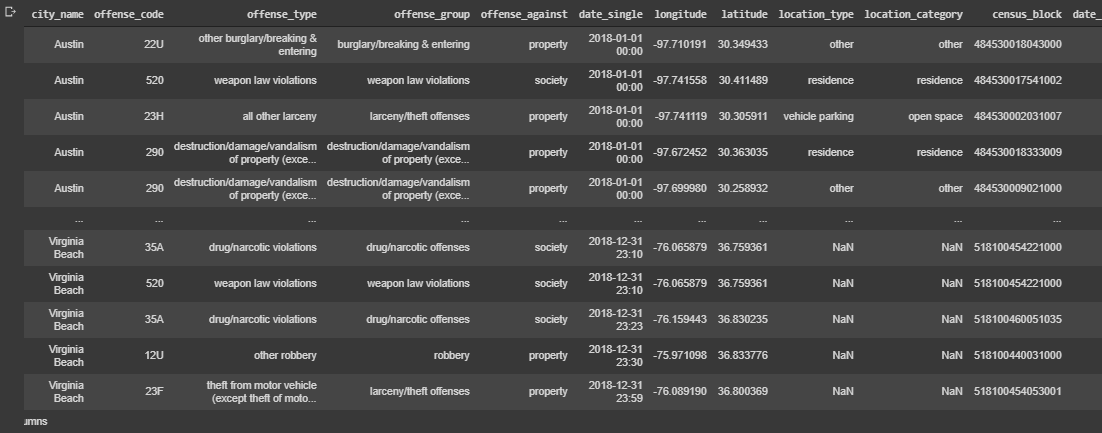
**PROJECT SUBJECT**: CRIMES IN MAJOR CITY ANALYSIS

DATASET:

# Introduction and Overview

The purpose of this analysis is to explore the open crime data source of a select group of major cities based on open crime data for 2018.

# Describe the data and its source(s), including any preprocessing

* 1. The open crime database came from <https://osf.io/zyaqn/>
  2. This is a publicly posted open database that includes crime statistics. Wiki does a good job at describing the dataset in the following Link <https://osf.io/zyaqn/wiki/home/>
  3. Pre-Processing?
     1. The datasource is relatively clean with a few fields that are someone duplicates that you would not want to consider in your analysis. Offense\_code is directly related to offense type. For charting, it is necessary to sample as the datasource is extremely large. The data source by default is in a \*.gz file so will need to be opened using compression = “gzip”.



# Describe your methods of analysis, including the questions that will be answered, in what fields the data will be used, and what the resulting output will be

* + 1. Using python functions in google Collaboratory the following questions will be answered.
       1. Can the type of crime and density be identified by city so that law enforcement can alter a plan for prevention and detection?
       2. Can the latitude and longitude values be used as a tool for home buying and law enforcement concentration as well?

# 3. Include an overall description of the program

# 4. If your project is a group project, describe the tasks and roles of each member of the group (2 points)

Using variations of Visual Data representation and aggregates, conclusions can be made regarding city and crime type and density. The first primary tool used is a mapping tool where a function is created in the program that controls the initial map zoom and the re-centering of the map based on the data to be displayed. The second primary tool used is plotly.express which is a phenomenal interactive visualization tool that can be imported into python.

VISUALIZATIONS FOR ANALYSIS:

FIGURE 1: SHOWS RELATIONSHIPS THROUGH CITY, AND CRIME TYPE VALUES OVER DATA

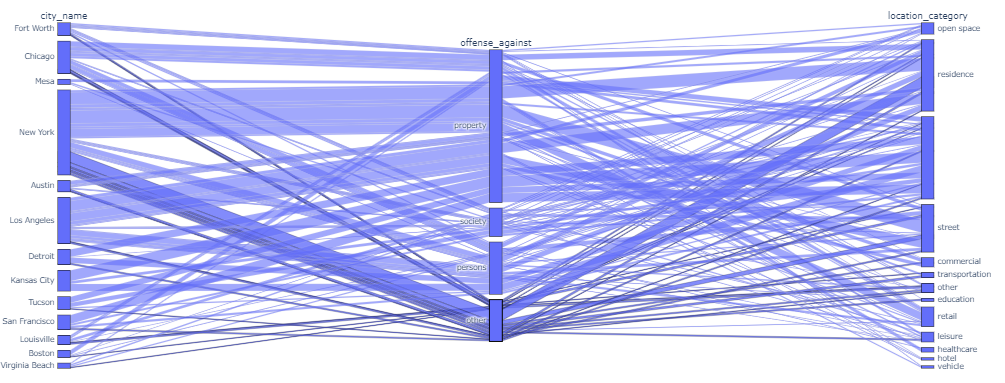


FIGURE 2:

Stacked bar by crime type by state

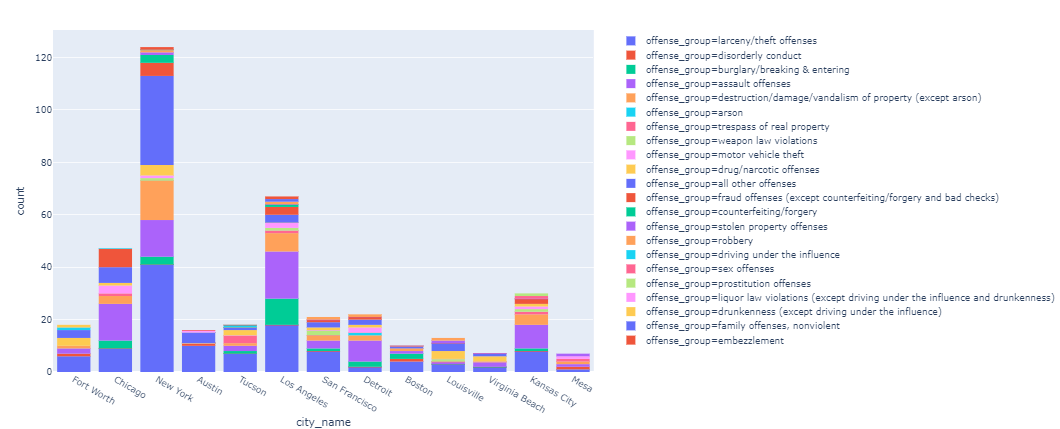


Figure 3: All up data comparison field by field matrix

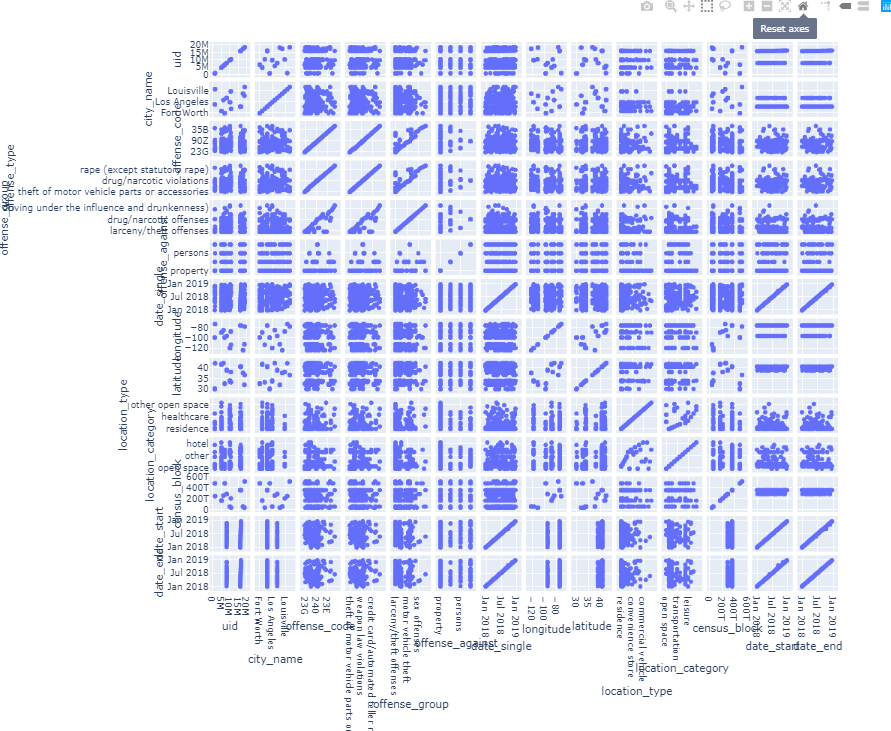


Figure 4:

Barchart by crime count by city

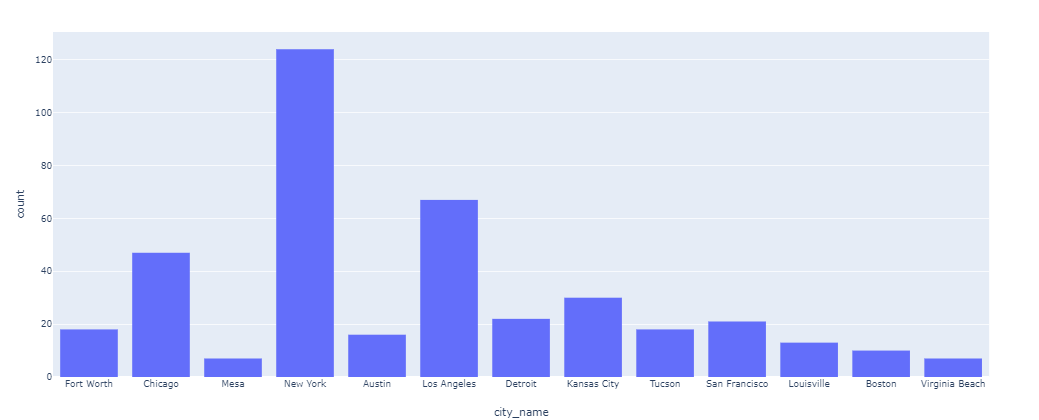
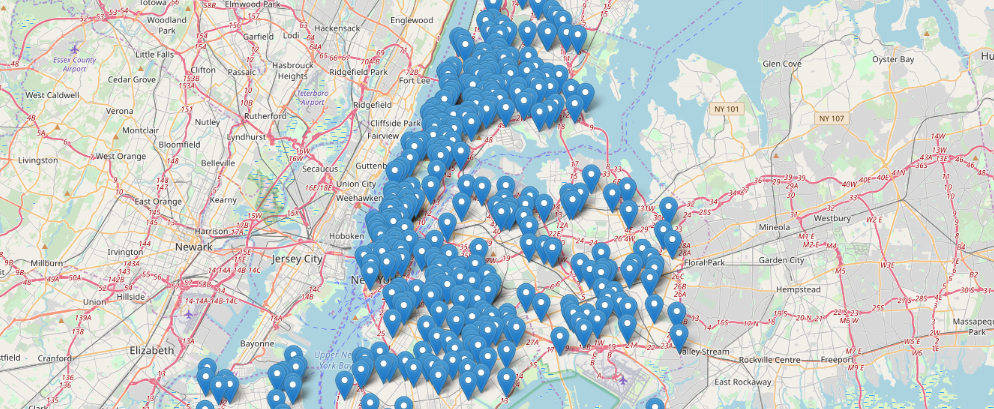
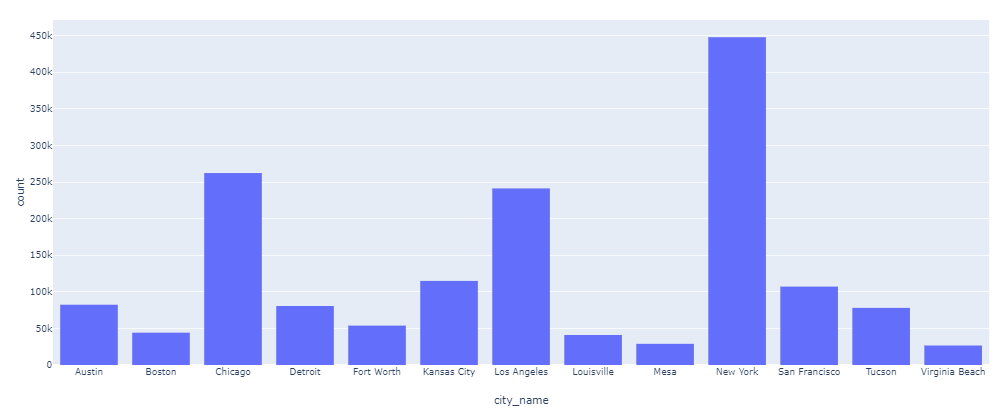


Figure 5: Sample crime map of city

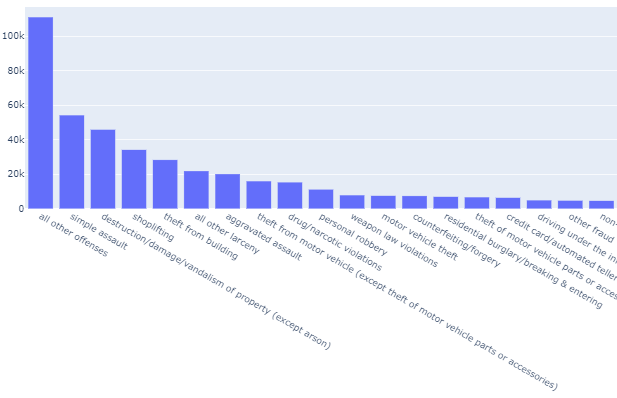


# 5. Draw conclusions from your results about your data

Some simple aggregation and visual analysis below allow some very confident and quick conclusions. New York has the highest number of reported crimes.



Break down the crime in only New York.



The largest number of identified crimes are simple assault at 54,219 reported.

If I still wanted to live in New York, I could use the mapping tool to find a location with less crime density.

Law Enforcement could use the same tool in planning officer daily routes.

