IS&T 5520: Data Science and Machine Learning with Python  
Dr. Langtao Chen  
Missouri S&T

**Group 4:  
Boston Crime Incidents   
Reflective Report**

Aaron Tucker

Alaina Martens  
Kwamena Ato Quainoo

Oluwatosin Martins Waleola

December 5, 2022

Crime significantly influences the life pattern of people in a given community. Community life, people’s choice of work, living and shopping are actively impacted by this canker. Many research confirms that the rate of crime is highly centered in special or “hotspot” areas in cities. These active places have been reported to be the primary sources from where crime spread throughout an entire city. In this study, we investigated the effects of these “hotspot” centers on the spread of crime in the city of Boston. We evaluated several crime reports obtained from the Boston Police Department within the year 2015 through to 2022. Our aims were; (a) monitor the trends in crime rates in the city of Boston, (b) determine the hotspots centers where crime rates occur, (c) determine the types of crime committed in these centers and the season in which these crimes are conducted. A successful evaluation will present the Police Department with the requisite knowledge needed to combat crime by presenting the department with information on crime rates based on geographical location, season and time and also equip the public with information necessary for their safety.

The visualization of our dataset showed us many interesting findings. The year 2015 presented the lowest count in crime rates. However, a surge in crime rates was observed from the year 2016 through to 2018 followed by a downward trend from 2019 through to 2022. The lowest crime rate observed was still 2015 for the entire studied years. Wikipedia reports that since 1992, crime rates saw a general decline in the United States until 2015. Our dataset confirmed this trend, however, we noticed that crime rates began to rise slightly from 2016 to 2018. The counts however began to reduce again from 2019 through to 2022. Many reasons can be attributed for the latter decline. It could be due to the era of COVID-19 seeing the increase rates at which people stayed indoors and the increase in police patrol during the season. Contrarily, CNN also reported in 2021 that the FBI recorded the lowest level of crime reportage since the year 1979. The drop in the reporting of crime in the entire country might also be the possible reason for the observed decreasing crime rates in the city of Boston.

Our data visualization also confirmed the impact of weather changes on crime counts. A surge in crime was observed during the summer months. The increase in crime rates is possibly due to the influx of people visiting the city of Boston to enjoy its fine weather and historical sites. Professor Daniel Semenza, a professor of criminology at Rutgers University reported that high crime rates in the summer months are due to the summer heat that causes people to spend longer time outside. These long periods increase human conflicts and the possible increase in crime. Furthermore, we observed that time is a major influence of crime in Boston. Crime rates were observed to be extremely high in the rush hours of 4-6 PM and lower in the early morning 4am through to 6am. Also, analyzing the data shows Larceny to be the most common crime in the city.

From our survey, the hotspots centers for crime in the city of Boston follow the order; B2 (Roxbury) > C11 (Dorchester) > D4 (South End) > A1 (Downtown) > B3 (Mattapan) > C6 (South Boston) > D14 (Brighton) > E18 (Hyde Park) > E13 (Jamaica Plain) > E5 (West Roxbury) > A7 (East Boston) > A15 (Charlestown). This information is important as it presents visitors and people looking to relocate with ample information that should determine where they should visit and how they should equip themselves if they do visit these dangerous centers. The safest cities in Boston are; Charlestown, East Boston and West Roxbury. The moderate safe cities are; Jamaican Plain, Hyde Park, Brighton and South Boston. The rest of the districts are classified as dangerous and thus residents and visitors are thus cautioned to be careful equipping themselves with the needed techniques to combating crime.

In conclusion, the following deductions were made from the BPD dataset;

* Larceny is the most common crime occuring in the city of Boston.
* Crimes are lower in the early mornings 4-6 am and higher in the evenings from 4-6pm.
* Crimes occur more on Fridays and less on Sundays.
* Crimes tend to increase in Summer, and reduce in winter.
* Districts such as Roxbury, Dorchester, South End, Downtown and Mattapan are most affected with crimes.

With all this being said we as a group had a pretty hard time working with this dataset. Due to inconsistencies between the 2015-2018 data and the 2019-2022 data. There were missing columns and they even changed how they formatted data between years. This caused us to have to drop a lot of our data and have to do a lot of data preprocessing. We spent most of our time trying to aggregate our data with little to no luck. It wasn’t until towards the end of our milestone 3 that we realized that we needed to do something with RFM analysis or clustering analysis. At this point we weren’t able to shift gears to redo all of our visualization or research work. We decided to work on predicting shooting instead. This ended up working better and we were able to do regression analysis and create a predictive model with Gaussian Naive Bayers. We ended up getting an AUC score of 68.57% on the model. If given more time we would do further tuning on this model. As well as dealing with the imbalanced dataset. If we continued this project in the future our goal would be to aggregate our data so we could answer our original research questions about predicting crime in locations in Boston. We were still able to see via visualization some interesting things but not so much as predicting the location or crime. Overall, this project taught us a lot about how important data preprocessing is. We didn’t process the data the best early on which affected how we developed models in the end. We were still able to learn an interesting story about shootings but it was not our initial goal.