Colin Michael

DSC 680: Project 1 Milestone 2 – Paper Draft

March 29th, 2024

Boston, Massachusetts is one of the most famous and historic cities in the United States of America. Boston has a population of approximately 675,000 people, including 35 colleges, two professional sports teams, and a Gross Domestic Product of $532 billion, eight-largest in the country (VisualCapitalist.com, BostonPlans.org). The City of Boston is a key metropolitan area of the U.S. but has struggled with containing crime. The objective of my paper is to analyze crime data for Boston, Massachusetts, with a focus on drawing actionable conclusions to help the Boston Police Department better confront criminal activity.

My primary data set is from Kaggle.com, titled “Crimes in Boston.” It is in csv format, including over 300k rows of data, one for each crime reported. Key columns of the data include offense type, district, occurred on date, hour, and street. I will load my csv data into Python and do my data manipulation and analysis there.

I will begin by looking at high-level trends in the data to an introductory understanding of the data. The data set includes 300k reported crimes from Boston, Massachusetts from June 2015 to September 2018. Here is a visual representation of the count of crimes reported by per month:

A graph with blue lines

Description automatically generated

The high-level trends in the data show crime counts increasing in the summer months and January and decreasing in the winter months.

Here is a bar chart of the count of the top 10 incident group codes:

A graph of a number of blue rectangular bars with white text

Description automatically generated

Motor vehicle accident response codes were the most frequently reported crime, followed by larceny and medical assistance.

Next, I applied the data science technique called time series forecasting to predict future patterns in crime data. Specifically, I used a Seasonal Autoregressive Integrated Moving Average, or SARIMA, time series forecast. SARIMA uses past timestamped data points, along with trends in seasonality, to predict future outcomes.

My SARIMA model specifically aims at predicting the number of incidents per month. The model returns a Root Mean Squared Error of .1673. Root Mean Squared Error, or RMSE, evaluates the values predicted against the actual values of a data set. A RMSE of .1673 suggests that the predicted values are close to the observed values in this model. The SARIMA model predicts this trend in seasonal crime data:

A graph with a red line and a pink line

Description automatically generated

The forecasted values of crime use empirical data to predict future trends in crime data for 2019.

Next, I will apply a clustering technique towards the Boston crime data. K means clustering is a technique that groups data points into a specified number of clusters. It helps visualize large data sets into actionable groups. I will use latitude and longitude as my spatial coordinates in my k means clustering exercise. I chose latitude and longitude because they are easy to visualize and provide the Boston police department with clear, actionable data points. I chose to apply 5 clusters to my model by using a silhouette score analysis. The model produces the following plot:

A diagram of a crime hotspots

Description automatically generated

The darker colored points in the above chart represent higher activity crime clusters. This k means clustering plot allows the Boston police department to focus their efforts on the latitude and longitude points.

In conclusion, I applied the data science techniques of SARIMA time series and k means clustering to help the Boston Police Department better deal with crime in the city. My SARIMA model found that criminal activity peaks in the summer months, as well as the month of January. My k means clustering helped group criminal activity in Boston into five clusters based on latitude and longitude. The analysis shows the latitudes and longitudes of the most active criminal clusters.

My predictive models can be used by the Boston Police Department to better allocate resources towards handling crime. My findings suggest that the BPD should increase its staffing in the summer months and January and focus on the high-volume latitude and longitude points.

My analysis is limited due to the nature of the dataset. It only records criminal activity from 2015-2018, which limits the effectiveness of the time series model. It would benefit my analysis to have a dataset that captures more years of data. Also, the dataset does not include data on if the incidents led to guilty verdicts. Bringing in additional data points on if the reported incidents led to guilty verdicts would help identify true criminal activity. This would make the analysis more accurate.

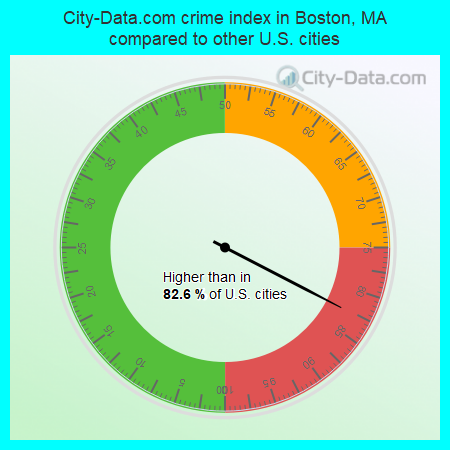
**Ten Questions from Audience:**

1. What is the biggest limitation on the data set?
   1. The limited scope of the data. In the future, I would want a much larger data set than 2015-2018. I would love to see how the results would look with 20 years of data.
2. What additional data points would you want?
   1. I would want to pull in the temperature at the time each incident was reported. I am very curious to see how weather impacts crime.
3. Biggest surprise from the data findings?
   1. I did not expect for crime to peak in the summer. My first reaction is that people are outside more in the summer, which leads to more criminal confrontations.
4. What inspired you to do this research?
   1. I live in Boston and take a lot of pride in the city. I want to see Boston prosper, and to do this Boston needs to handle crime better.
5. How do crime rates in Boston compare to similar cities?
   1. Portland, Oregon and Boston, Massachusetts have comparable populations, both approximately 650k. In 2022, Portland had 101 homicides, while Boston had 40 homicides (OregonLive.com, NBCBoston.com).
6. What are different measures that could impact spikes in crime in Boston?
   1. Big events like sports games or conferences lead to increases in temporary population, which I’d expect to impact crime.
7. Do you think the BPD would be receptive to these findings?
   1. Yes, I tried to make my discoveries very clear, concise, and consumable by non-data parties.
8. Why did you use a SARIMA model?
   1. I wanted to be sure that I took seasonality into account.
9. How could these findings by used to reinforce biases?
   1. I worry that the latitude and longitude clustering findings would reinforce biases that certain groups of people living in areas are more inclined to commit crime.
10. What initiatives has Boston taken to curb crime?
    1. In 2023, Boston joined a program designed by University of Maryland Violence Reduction Center to reduce gun violence. The program aims at bridging the gap between academic findings and actual police behavior (Dailyfreepress.com)

**Appendix**:

Criminal Information in Boston

According to City-Data.com, Boston has a crime rate higher than 82.6% of U.S. cities. The 2022 rate of crime increased by 9% compared to 2021 (City-Data.com).



Per 100,000 people, in 2022, Boston had 176 rapes, 770 robberies, 2,965 assaults, 1,213 burglaries, and 321 arsons. There are also 328 registered sex offenders living in Boston as of April 07, 2024.

Niche.com did an analysis on public sentiment on crime in 2024. 51% of people felt pretty safe in Boston, 23% felt somewhat safe, 22% felt very safe, and 4% felt not safe (Niche.com). Also, 40% of people reported that they felt Boston police are very visible and responsive, 25% felt police are visible but slow to respond, and 17% felt police are not around much, but are quick to respond when needed.

The neighboring city of Providence, Rhode Island reported that 74% of people felt pretty safe, 11% felt somewhat safety, and 11% felt not safe at all. In Providence, 26% of people feel the police are very visible and responsive, 26% feel police are visible but slow to respond, 26% feel police aren’t around much and aren’t reliable, and 16% feel unsure.

**Citations**:

Boston by the numbers colleges and Universities. (n.d.). http://www.bostonplans.org/getattachment/1770c181-7878-47ab-892f-84baca828bf3

Boston, A. (2018, September 4). *Crimes in Boston*. Kaggle. <https://www.kaggle.com/datasets/AnalyzeBoston/crimes-in-boston>

*Boston crime rates and statistics*. Niche. (n.d.). https://www.niche.com/places-to-live/boston-suffolk-ma/crime-safety/

*Crime rate in Boston, Massachusetts (MA): Murders, rapes, robberies, assaults, burglaries, thefts, auto thefts, arson, law enforcement employees, police officers, crime map*. Crime in Boston, Massachusetts (MA): murders, rapes, robberies, assaults, burglaries, thefts, auto thefts, arson, law enforcement employees, police officers, crime map. (n.d.). https://www.city-data.com/crime/crime-Boston-Massachusetts.html

Koop                                                                                                                                                                                                      Graphics/Design:                                    , A. (2023, May 18). *Mapped: The largest 15 U.S. cities by GDP*. Visual Capitalist. https://www.visualcapitalist.com/us-cities-by-gdp-map/