Matthew Wilchek, DATS 6401, Summer 2020

Final Technical Report

## Washington, D.C. Crime Incident Analysis & Forecasting

**Abstract**

Data Analysis and visualization are essential tools for law enforcement today. Crime is also a dataset that will, unfortunately, always exist and grow. This project uses visualizations to help improve how Washington, D.C. crime data from 2010 to 2020 can be used for criminal analysis and information sharing within the law enforcement community. Of the various aspects analyzed in this analysis, it can be deduced that the Metropolitan Police Department (MPDC) have optimized resources to locations where crime is more prevalent and deterred the use of deadly weapons since 2016. This report will discuss the objectives, approach, requirements, and other conclusions drawn from this analysis.

**Objectives**

The primary objective of this project is to visualize crime statistics of the Washington, D.C. area from 2010 to 2020. There were 362,762 records available for analysis and visualization, which includes features such as coordinate locations, type of crime, whether a weapon was used, and if so, what, time of day, and local precinct area. Insight can be gain that will identify the rate of dangerous weapons used in crime, months, and seasons when a crime occurs more often, and what geographic areas of the city are more dangerous than others. Also, this analysis shows how the global pandemic has factored in the decrease of recent crime and what the rate of crime incidents occurring might look like ahead using time-series forecasting.

**Functional Requirements**

This project must be able to consolidate, preprocess, and summarize detailed criminal incident data and geo-process coordinates into a geodatabase. The summarized information must be organized so that powerful representations can be made, and end clients will have the option to make determinations. The end-user will be able to interface with multiple web pages and interact with various visualizations, map applications, and intelligence dashboards. The visualizations will include the ability to hover over or to zoom in some geographic regions for additional information.

**System Architecture and Description**

The data consolidation, preprocessing, and summarization was completed using a script using R, which manipulated several CSV files. The 2010 to 2020 Washington, D.C. crime incident data was collected from [Open](https://opendata.dc.gov/) Data DC and was available to download as ten separate CSV files. Different subset tables by season, offense, and time were made using the script, which was then copied into a Google Worksheet to be utilized for Google API visualizations. The master dataset was converted to a CSV file that contained all incident records that were geocoded into a geodatabase for the ArcGIS piece and extracted for the Tableau piece. All data, including the raw files, consolidated structured data, and summary tables, were ultimately stored online using a Google OneDrive, ArcGIS Public Portal, and Tableau Public.

**Development Platforms**

The user-facing front end of this analysis is an HTML based webpage that includes various tabs to organize the data by analysis methodology (e.g., statistics summary, cluster analysis, forecasting, etc.). Users can interact with the webpage using links, hovering over visualizations with additional information, or clicking into to drill down. The webpage will be stylized using Cascading Style Sheets (CSS) from a Bootstrap template. The visualizations will be created using Google API visualizations, ArcGIS, and Tableau, which will be discussed in greater detail in the following section.

**Proposed Visualizations**

The Google API visualizations JavaScript library will be used to create the summary visualizations for this analysis. Bar charts, column charts, and a treemap will be first applied to explore initial distributions to answer basic questions of the crime data. Followed by line graphs to answer how criminal activity has changed over time and by what offense type. The geospatial coordinate data of the criminal incidents is critical information to analyze to drive optimization of law enforcement resources. Therefore, cluster analysis will be developed using ArcGIS. By using ArcGIS, an interactive map application can be made to allow users to drill down to specific areas of Washington, D.C. to see what types of crime occur over others, and how areas have changed over time. Tableau will also be used to portray regions of the city broken out by time. Lastly, a forecasting tool with confidence intervals will be developed using Tableau’s analytic functions. The tool will allow users to see what crime in Washington, D.C. may look like and by the offense for the next year, especially if COVID-19 remains to be a problem in the United States.

**Experimental Analyses and Conclusions**

The data first revealed that theft is the most significant offense type committed in the Washington, D.C. area. Robbery, motor vehicle theft, and burglary were also quite high. Compared to the more severe crimes such as homicide, sex abuse, and arson were very low for the past ten years. The overall trend for most offense categories has been decreasing since 2016. However, there was a slight increase in homicide by five from 2018 to 2019. There was a significant decrease in the use of guns in 2016 of 2,124 incidents to 1,586 incidents in 2017. There was also a slight decrease in the use of knives. But the trend remains to be consistent so far. Therefore, we can infer that starting in 2016, the MPDC started making significant impacts on public safety and continue to do so. Both in decreasing the number of incidents relating to stealing and the use of deadly weapons.

In 2019 the number of crime incidents that were recorded during the winter was 8,004 and in the spring was 7,723. Since the pandemic, the numbers have dropped to 5,490 and 5,656, respectively. These numbers are the lowest number of criminal incidents recorded in the past ten years! Finally, based on the current trends and the pandemic still having a massive impact on the country, it is likely that we can conclude that decreases for the summer and autumn may be the lowest in the last ten years as well.