Matthew Wilchek, DATS 6401, Summer 2020

Progress Report

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

The 2010 to 2020 Washington, D.C. crime incident data was collected from <https://opendata.dc.gov/> and was available to download as ten separate CSV files. I then developed a pre-processing script using R programming, which merged all 10 CSV files into one master data frame. After, I subsetted the data to only columns that would be interesting for the analysis/visualizations. Then I corrected a couple of data types to ensure the formatting would be appropriate for visualization tools (such as dates). I originally planned to convert the data to a Google Sheet for the Google Visualization API piece. However, the total file size is about 40MB and too big to upload. Therefore, I will create an additional pre-processing script that will generate summary tables with numerous calculations of all the incidents for various visualizations and then create a separate Google Worksheet with the tables as different tabs, similarly to our first project.

As a preliminary exploratory data analysis (EDA), I uploaded the data into a Tableau workbook to first see various distributions of the data. The overall quality of the data after my first pre-processing script is excellent; no data gaps of the total 383,709 records. I also began working on developing a map analysis that breaks out all the incidents by the D.C. district and a forecasting piece using the Tableau Analytics functionality. The mapping functionality for Tableau isn’t the greatest, so I intend to use an ArcGIS trial account to develop and publish a cluster analysis layer that will provide greater geographical insight.

I have constructed the frame for my webpage using a free Bootstrap template and edited all the pre-made content with an outline of analytical tabs for the different visualization pieces I will develop. The first tab/home page will be an about section of the project/data, then a summary tab of high-level statistics using the Google Visualization API, followed with the ArcGIS Map Application, and ending with two additional tabs of Tableau Dashboard visualizations. All of my project’s progress can be found here: <https://github.com/mwilchek/dc-crime-analysis>