Data Mining Project Proposal

Group Members: Polly McKim, Spencer Stucky, Malikat Coulibaly, Junhe Zhang

March 5, 2020

**Crime Arrest Data for Baltimore 2013-2017**

**Project Proposal:** We plan to analyze data of criminal arrests in Baltimore from 01.01.2013- 11.04.2017. We plan to use arrests as a proxy for crimes committed in the city and use this project to analyze where, when, and what types of crimes are being committed in Baltimore. By exploring this data set we hope to identify patterns in the types of crimes committed and areas where they occur. By identifying patterns, we hope to create predictive models which could alert us to areas that may need more policing or other forms of anticrime intervention. In addition to that, the statistics we will extract will help us determine whether the previous major reforms undertaken by the law enforcement were effective or not. We also plan to possibly explore if there is a disproportional amount of people being arrested from various groups, when taking into account the demographic makeup of Baltimore city during this time frame.

**Dataset:** <https://data.world/baltimore/3i3v-ibrt/workspace/file?filename=bpd-arrests-1.csv>

**Source of Dataset:** This dataset was provided by Baltimore City’s “Open Baltimore” project from the mayor’s office, which publishes various datasets and information statistics online, “[t]o support collaborative and transparent government.  Our group plans to use a version of the dataset taken from the data aggregate website “Dataworld.com” because this version includes a

filled in “neighborhood” variable which we believe will be valuable in our analysis. We will download the data as csv file and run the analyses in Python.

**Description of Dataset:** 18 variables (columns); 154,291 observations (rows) which each represent a different arrest; The variables in the dataset include: date and time of arrest, gender, race, age, geographical coordinates of arrest, neighborhood, offense committed. The data dictionary notes that the data set only includes the “top offense committed,” meaning the most serious charge related to the arrest rather than all charges resulting from the arrest.  We will need to keep this in mind when analyzing our data. The data dictionary also notes that this dataset does not include information for individuals arrest processed through the juvenile arrest system.

**Questions we hope to answer/explore:**

·      Do certain areas experience a higher rate of crime than others?

·      Are certain types of crimes more likely to happen in specific areas of the city?

·      Does time of year have an impact on the level of crime experienced?

·      Does time of day have an impact on the level of crime experience?

.    Is there a correlation between the (gender/race/age) and the type of crimes committed?

·      Are there any trends present in the dataset with regards to the overall amount of crimes being committed?  (Have crime rates for individual categories of crime improved/worsened?)

·      Do neighborhoods with lower socioeconomic factors have higher rates of crime?

·      Can we construct a model to predict which areas/times Baltimore should concentrate resources to prevent crimes from occurring?

**Sources:** <https://technology.baltimorecity.gov/About_Us>

<https://data.world/baltimore/3i3v-ibrt/workspace/file?filename=bpd-arrests-1.csv>