## **The Wired: Reducing Baltimore’s Homicide Rate through Data Science**

Chris Murray, Rudy Ferrara, Tarun Chopra, Winston Lin

### **1.0 Background**

In 2015, homicides have increased in more than 30 US cities, and no city has been more impacted by the recent rise in inner city violence than Baltimore, where homicide rates have reached their highest levels[[1]](#footnote-1) since the 1990’s. Currently, Baltimore has a homicide rate of 48.97 per 100,000 residents, breaking the 1993 record of 48.77 homicides per 100,000 residents. Our project will explore how Baltimore city officials, law enforcement agencies can stabilize and reduce homicide rates by using data science techniques to help design and implement intervention programs.

Traditionally, most of the public policing around prevention of violent crimes were centered around the basic premise of police force working effectively with communities. During 1990s, three cities mounted widely publicized policing intervention efforts aimed at curbing higher homicide rates. Boston’s operation Ceasefire[[2]](#footnote-2) focused on firearm violence involving youth gangs, New York’s Compstat sought to restore order on the street and accountability for crime in police department, and finally Richmond’s Virginia project Exile entailed sentence enhancements through federal prosecution for violent and drug crimes involving firearms. Although, the above efforts of policy decision making helped in crime reduction, these programs were all launched after the crime rate had dangerously risen..

Lawmakers and government officials are exploring approaches along with preventing policing and rather than being more focused on the means of “policing” than its “ends”. One such model, Problem-oriented policing (POP[[3]](#footnote-3)) is created with the goal is to change crime prevention focus from being primarily reactive to a model requiring proactive action by police forces. POP framework also allows the focus on the identification of underlying problems that could be targeted at the root of crime and disorder in the society. POP is a holistic, data-driven approach to crime prevention that focuses on the root causes and indicators of crimes, rather than just the penalties for committing them. The body of research on POP is moderate, but growing: 44 papers have been published on the subject, many of which examine the effects of POP intervention programs on crime rates in Boston, New York, and Richmond. In addition, large cities, including Baltimore, have made city data publicly available in order to foster transparency, openness, and innovation that can help to improve the lives of citizens overall.

With vast improvements in technology centered around streamlining of data collection and data analysis, parts of the country are starting to leverage predictive analytics to fight crime and improve overall wellbeing of the society.For example, in 2013 Santa Monica started the Wellbeing Project[[4]](#footnote-4) with a simple concept of identifying key factors affecting the city wellbeing, collect the data on how the city measures up towards key wellbeing metrics. Based on the data analysis, The Wellbeing Project worked alongside with city and community officials totake actions on what is needed for improvement. In 2012, California introduced PredPol (predictive policing)[[5]](#footnote-5) leveraging data analytics platform to help fight crime. Since, its inception Santa Cruz police department had reported a 27% reduction in burglary and 19% reduction in property theft by 2012. Our goal, similar to Santa Monica’s Wellbeing Project and California PredPol project is to provide decision makers with a predictive tool that analyzes factors tied to higher homicide rates (e.g., poverty, unemployment, education, infrastructure/amenities, gun sales, citizens’ wellbeing) and whether the homicide rate will improve or worsen for each neighborhood in the city.

The central question in this domain is “How do we reduce crime and disorder in the city?” Interesting sub-questions include:

* What regions should we focus on?
  + Where, when, and how are these crimes committed?
  + Who is committing these crimes? Who is affected by them?
  + What are the root causes and indicators for these types of crimes?
* How do we reduce violent crime and homicide? Punishment or intervention?
  + What penalties will discourage these types of crimes?
  + What kind of intervention should be implemented?

This project will address the question of which regions should be prioritized for violent crime and homicide prevention programs. We chose this question because of its relevance to current events related to certain cases of excessive use of police force and the resulting increase in civil unrest and violent crime in major cities across the U.S. Using data science to answer this question more efficiently and accurately will help policy makers and police forces protect the health of their cities. Also, from a practical standpoint, there is a significant amount of published research examining actual POP interventions in major cities, thus providing us a good foundation for exploring our own ideas in this domain.

### **2.0 Research Design**

The metrics used as predictor variables will be divided into wellbeing metrics and crime metrics. Monthly (when available) values for the metrics will be stored for each neighborhood in Baltimore. The historical metric values will be extracted from various sources and backfilled to the year 2005. The metrics will be normalized per-capita so that correlations that occur across neighborhoods, and correlations between data from nearby neighborhoods can become part of the statistical model. Values will be repeated for each month in the year for metrics that only have data available at yearly level, for example high school graduation rates and average educational level per adult.

To avoid legal and ethical issues, metric values will only be stored at the aggregate level - per neighborhood, with monthly historical values. The data will be extracted only from public data sources. No personally identifiable data will be stored.

**2.1 Wellbeing Predictor Variables**

Higher community wellbeing may lead to lower crime rates. According to the Urban Health Institute, “the new Police Commissioner has indicated that the health and wellbeing of the community is important because it impacts crime and opportunities to reduce crime.”[[6]](#footnote-6) Wellbeing metrics will be used as predictor variables in statistical model to predict future homicide rates.

* High school graduation rates (Baltimore City Public Schools)
* Unemployment rate (Bureau of Labor Statistics)
* Housing Affordability Index - Mortgage (National Association of Realtors)
* Rent as & of Income (REIS)
* Home Ownership Rate (U.S. Census Bureau)
* 311 calls by type (<https://data.baltimorecity.gov>)
  + Rat Rubout
  + Bulk Sanitation
  + Sanitation Property
  + Dirty Alley
  + Parking Complaints
  + Dirty Streets
  + Street Lights Out
  + Potholes
* Average education level per adult (U.S. Census Bureau - yearly data only)
* Number of 911 calls (https://data.baltimorecity.gov/)

**2.2 Crime Predictor Variables**

Increases in rates of certain categories of crime may lead to increases in the crime rates in other categories. For example increases in drug crime rates might lead to increases in violent crime rates.[[7]](#footnote-7) Monthly historical values per neighborhood will be stored. Historical values will be obtained from the Baltimore Police Department.

* Property Crimes per-capita
* Traffic crimes per-capita
* Drug Crimes per-capita
* Violent Crimes per-capita
* Illegal weapon sales/ownership per-capita
* Homicides per-capita

**2.3 Forecasted Metrics**

Violent crimes and homicides will be forecasted by the models that are developed.

* Violent Crimes per-capita
* Homicides per-capita

**2.4 Statistical Models**

Models will be developed to forecast the number of violent crimes and homicides in each neighborhood for the next twelve calendar months. Actual data for the current calendar month will be replaced in the data as it becomes available from the Baltimore Police Department, and the remainder of the current month will be forecasted. The current month will therefore have a combination of actual and forecasted data.

The standard deviation for the monthly number of violent crimes and homicides will be calculated. Crime rates that are forecasted to rise or fall more than a certain number of user chosen standard deviations will be flagged in the output. This is to reduce the noise in the output resulting for random fluctuation, and to increase the likelihood that the output data will be meaningful and actionable for customers.

Models will be generated using any available and appropriate statistical techniques, such as time series analysis and regression. The models will be developed based on historical data. The model developers will look at monthly historical data up until specific points in time without taking into consideration the historical data after that point in time. The models’ forecasted values will then be tested against the monthly historical data available for the following year. The goal is to create models that can be used to forecast the monthly number of crimes per neighborhood for one year.

The models will be continuously updated and improved through time. As the models are improved, predictor variables will be added and removed. For example new predictor variables will be evaluated that quantify information about family structure, demographics, community services (after-school programs, churches, counseling, etc.), and vacancy rates.

### **3.0 Conveying Research Findings**

We propose the development of a software model that we call **Anderton,** after the character in the movie *Minority Report* who is in charge of the police department’s PreCrime unit. The web-based graphical user interface will show a heat map of forecasted crime for each neighborhood of Baltimore, as well as an overall city-wide crime forecast. The user will be able to select a time period of a month, a quarter, or overall for the next 12 months. Each neighborhood is colored one of 5 shades from green to red, where green indicates a much better crime forecast and red a much worse crime forecast than the city-wide forecast. Clicking on a neighborhood brings up individual forecasted statistics for that neighborhood.

Left unchecked, increased crime rates can lead to a downward spiral that may significantly damage a city for a long period of time. An uptick in crime leads to lower economic investment, which eventually lowers the tax base and decreases the resources available to fight crime. It is important to break this trend quickly before it get out of control. Anderton will help spot these trends early on and enable officials to act before the problem becomes much more difficult to manage. Armed with this knowledge in advance, officials can restructure the budget to provide more resources where they are needed or redirect patrol units to spend more time in trouble areas.

While Anderton will initially be designed using data specifically for Baltimore, the model itself is generalized and could be ported to use data from other cities. City officials and police departments nationwide would be well-advised to adopt our model in order to gain valuable advance notice of future crime trends. Some benefits include:

* Knowing where to focus efforts - Determine which issues have a disproportionate effect on the community. Solving certain easy problems can have a big benefit to the community.
* Resource allocation - Help determine the best use of limited policing resources, such as where patrols should direct their efforts to have the most impact on reducing crime.
* Budget planning - As more and more cities grapple with balancing their budgets, police budgets are on the chopping block along with other city services. Our model can help officials set more realistic budget requirements and help determine when and how much overtime pay should be authorized.

### **3.1 Limitations and Risks of the Model**

Anyone familiar with weather forecasting know that forecasts are not always accurate, but people still make decisions based on what the weather models predict. Our crime model is similar in that a specific forecast is not guaranteed to be accurate, but on average using the model can result in better decision making. Our model’s accuracy improves with more data. Homicides are, statistically speaking, rare events so they are difficult to predict with any specificity but can be forecast in the aggregate.

One limiting factor for forecast accuracy is the quality of data, especially in geolocation information. Many of the datasets we plan to use include street-block-level addresses, which provide sufficient location accuracy for our purposes. Other datasets, however, are broken down by police district, school district, zip code, or simply do not have geographic information at all. We will normalize geolocation information into geohashes but we would like to see more standardized and granular geolocations where possible. Another limitation is the age of some data. While many datasets are available near-real-time, others are several months or up to a year old. More timely access to the data should result in improved forecast accuracy.

We recognize that some critics will dismiss the idea of crime forecasting altogether as something out of science fiction, or as no better than existing policing techniques. There is a reluctance to try something new and not well-understood. However, we believe our model will prove itself over time. As new datasets and more accurate data become available the model will continue to improve. Eventually we expect our model will replace Problem Oriented Policing as the new standard policing method.

### **3.2 Future work**

Although our proposal is focused on Baltimore, we believe our model can be applied to other cities with relative ease. Most of the work in porting Anderton to other cities involves normalizing the data into a format that it can use. More standardized data formats would greatly reduce this task. As new data sources become available they can be incorporated. One datasource we are very interested in pursuing is the ShotSpotter gunshot detection system. Although Baltimore decided against implementing the system[[8]](#footnote-8) due to cost, other cities such as Pittsburgh, Denver, and Washington have purchased it. We believe data from this system will dramatically improve our violent crime forecast. We also would like to incorporate ideas from Matt Gerber’s research in using Twitter to predict crime[[9]](#footnote-9), and to expand our model to provide property crime forecasts in addition to violent crime.

We look forward to developing our model and ultimately hope to play a small role in reducing crime in Baltimore and nationwide.

1. <http://www.baltimoresun.com/news/maryland/baltimore-city/bs-md-ci-homicide-per-capita-20151117-story.html> [↑](#footnote-ref-1)
2. [http://policeandcommunity.org/pdfs/Archives/2010/Archives-Effectiveness/2005\_Did-Ceasefire-Compstat-Excile-](http://policeandcommunity.org/pdfs/Archives/2010/Archives-Effectiveness/2005_Did-Ceasefire-Compstat-Excile-Reduce-Homicide_Rosenfeld.pdf)

   [Reduce-Homicide\_Rosenfeld.pdf](http://policeandcommunity.org/pdfs/Archives/2010/Archives-Effectiveness/2005_Did-Ceasefire-Compstat-Excile-Reduce-Homicide_Rosenfeld.pdf) [↑](#footnote-ref-2)
3. <http://www.campbellcollaboration.org/lib/download/228/> [↑](#footnote-ref-3)
4. <http://wellbeing.smgov.net/> [↑](#footnote-ref-4)
5. <https://www.predpol.com/governing-magazine-features-santa-cruz-police-and-predpol/> [↑](#footnote-ref-5)
6. <http://urbanhealth.jhu.edu/_PDFs/HealthBaltimore.pdf> [↑](#footnote-ref-6)
7. <https://ncadd.org/about-addiction/alcohol-drugs-and-crime> [↑](#footnote-ref-7)
8. <http://www.baltimoresun.com/news/maryland/sun-investigates/bs-md-sun-investigates-guns-20150207-story.html> [↑](#footnote-ref-8)
9. <http://uvamagazine.org/articles/predicting_crime_140_characters_at_a_time> [↑](#footnote-ref-9)