Top-Bottom level Analysis of Baltimore City Crime

Demarcus Wirsing, University of Maryland Baltimore County

**ABSTRACT**

*In this paper the author summaries the top level and bottom level analysis of crime within Baltimore City, as well as what interesting findings were made when reviewing the crime reports. The paper is based on personal observation along with data intake and modeling to display the different levels of crime within Baltimore City. In light of the findings from the report I hope to shade light on the data that is presented.*

**INTRODUCTION**

Founded on July 30, 1729 by Cecil Calvert, Baltimore was established as a port for shipping tobacco and grain. Fast forward to today Baltimore City is home to approximately (+ or -) 619,493 residents as reported by the United State Census Bureau in 2011. It is the number one most popular city in the state of Maryland. As well as home to many tourist attractions being the famous Inner Harbor, National Aquarium, Oriole Park, Baltimore Museum of Art, and many more. With the amount of populist along with notable tourist sites, Baltimore City is also amongst the top ten most dangerous cities in America reported by both “Forbes” and “USA Today” due to the amount of crime that is committed per residents in the city itself.

With being in the list of the most dangerous cities in America, I created a deep dive in the form of a top-bottom level analysis to highlight certain areas in Baltimore City that is most concerning that could be focused on. This top-bottom level analysis will focus on the crime reported in each district followed by a further drill down.

**METHODS, RESULTS, AND DISCUSSION**

Please click the GitHub link below to have access to my Baltimore City analysis [(Baltimore City CrimeReport)](https://github.com/marcusw0602/DATA-692/blob/master/Baltimore%20Crime%20Forecast%20Project/Baltimore_City_Crime_Report_2011-2020.ipynb)

I first gathered two datasets from Baltimore City’s website “Open Baltimore”. This website houses data sources pertaining to different departments within the city that is accessible to the public and is frequently updated by the city’s workers. The datasets that I will be using is from department of Public Safety, those sets are as followed “BPD\_Crime\_2011-2016.csv” and “BPD\_Part\_1\_Victim\_Based\_Crime\_Data.csv”.

My analysis will be focused from the year 2011 to current. Now in the first developments of my report I am reading in certain columns of data from both csv files. I am doing this because between the two files only what is listed is shared.

* CrimeDate
* CrimeTime
* CrimeCode
* Description
* Inside/Outside
* Weapon
* District
* Neighborhood
* Total Incidents

Since importing and combining the two files I then started to change and manipulate the columns. First line you will see me reading in the new csv file using the variable “CrimeReport”. The next six lines you will see me cleaning and adding columns in the data set, I changed all "NaN" values to unknown or not reported to see from a reporting standpoint how much data is missing. Last two lines you will see me filtering out any row of data that does not contain the year 2011 or greater. This is done due to the years before 2011 only containing partial entries. I did not want this to mess up the metric portion of my charts down the line in the project.