Top-Bottom Level Analysis of Baltimore City Crime

Demarcus Wirsing, University of Maryland Baltimore County

**ABSTRACT**

*In this paper the author summarizes the top level and bottom level analysis of crime within Baltimore City, and the interesting findings that 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 shed 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, give or take, 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. It is also home to many tourist attractions such as: 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. This fact has been reported by both “Forbes” and “USA Today”. This ranking is due to the amount of crime that is committed per residents in the city itself.

Due to Baltimore’s notorious reputation on 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 and that should be focused on. This top-bottom level analysis will focus on the crime reported in each district followed by a further drill down.

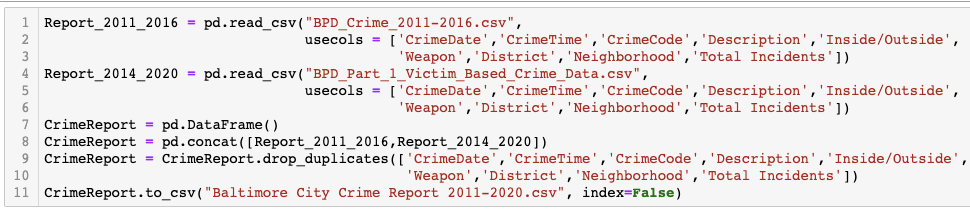
**METHODS, RESULTS, AND DISCUSSION**

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) if the pictures are hard to see.

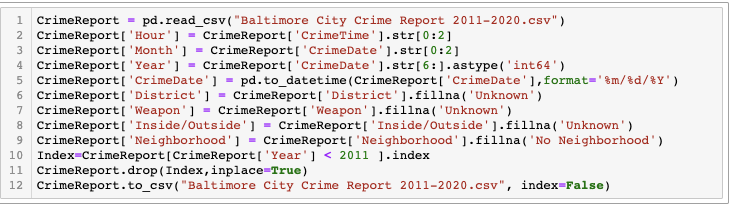
I first gathered two datasets from Baltimore City’s website “Open Baltimore”. This website houses data sources pertaining to different departments within the city that is accessible to the public and is frequently updated by the city’s workers. The datasets that I will be using are from the Department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

My analysis will be focusing on the period 2011 to the current year. In the first developments of my report, I am using python packages to create such analysis. These packages are Pandas, Altair, and csv. I will then continue on to read certain columns of data from both csv files. I am utilizing this method because, between the two files, only what is listed is shared.

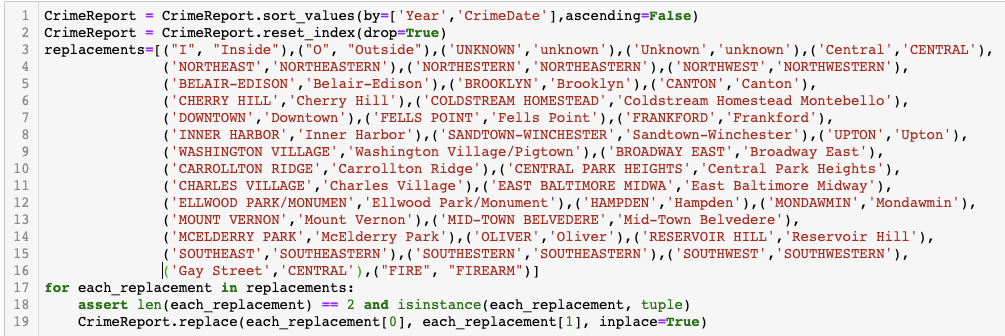
* CrimeDate
* CrimeTime
* CrimeCode
* Description
* Inside/Outside
* Weapon
* District
* Neighborhood
* Total Incidents



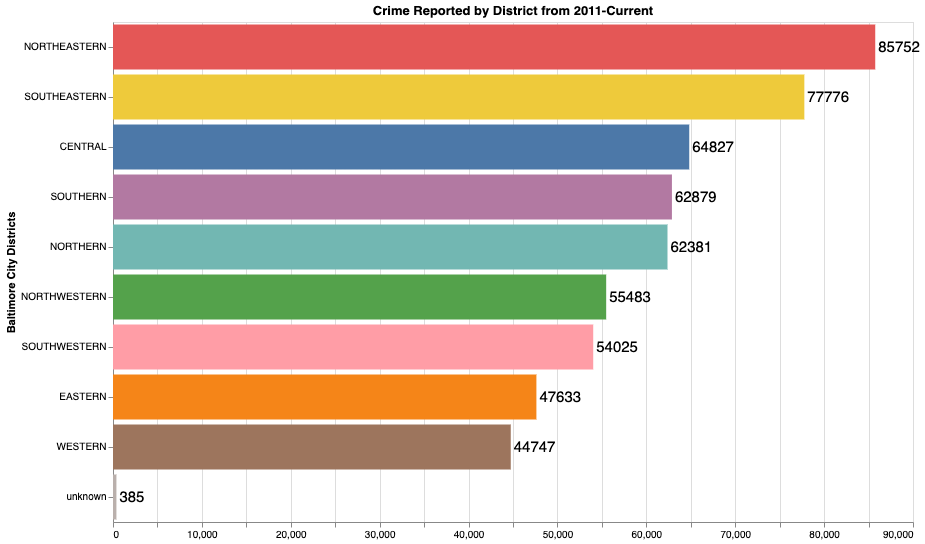
Since importing and combining the two files, I then started to change and manipulate the columns. The first line you will see me reading is in the new csv file using the variable “CrimeReport”. The next six lines you will see me cleaning and adding columns in the data set. I changed all "NaN" values to unknown or not reported to see from a reporting standpoint how much data is missing. In the 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 alter the metric portion of my charts further on in the project.



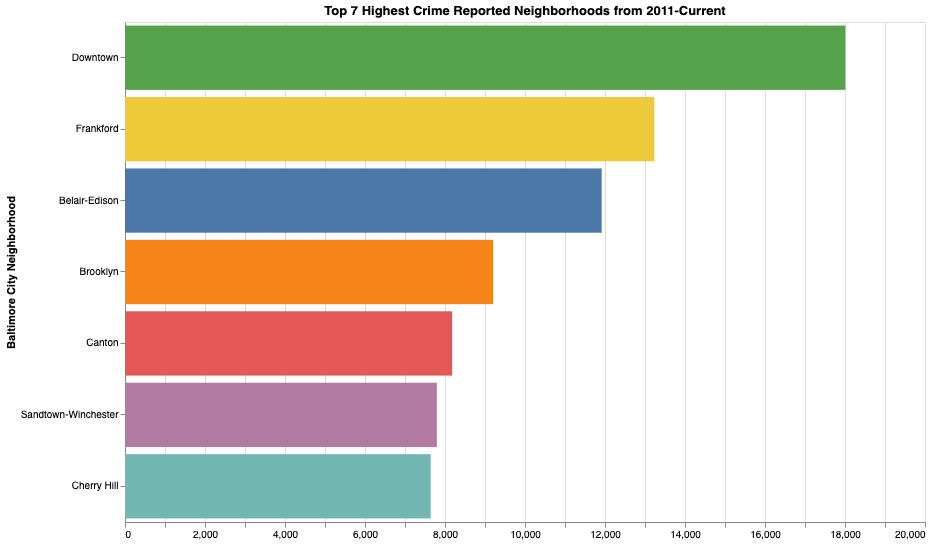
After making certain alterations to the dataset columns, I noticed that there were thousands of duplicates regarding the districts and neighborhood columns. When I discovered this, I went in and replaced all data strings to match a uniformed string so that the metrics later on are not misleading or incorrect.



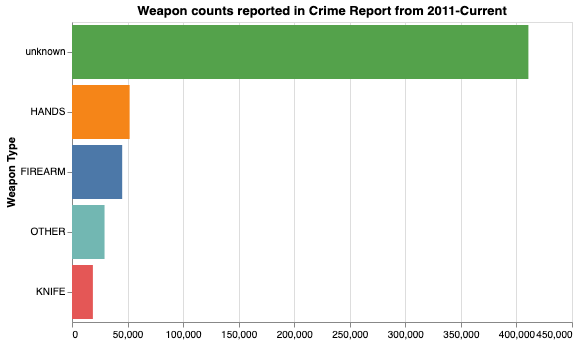
Once I cleaned the data, I began my analysis on the crime levels within the nine districts of Baltimore City. You will see from the bar graph below, that the Northeastern district, out of all 9 Baltimore City districts, is the most dangerous with a report of over 80,000+ reports from 2011-2020. You will also find that the Southeastern district is a close second will almost 75,000+ reports. At the very bottom you will see an unknown mark. This is the marker with the amount of criminal reports that were not reported with a district. There could be some questions as to why this is so. Furthermore, we could also raise the question as to how we can make sure each report is placed at the correct location.



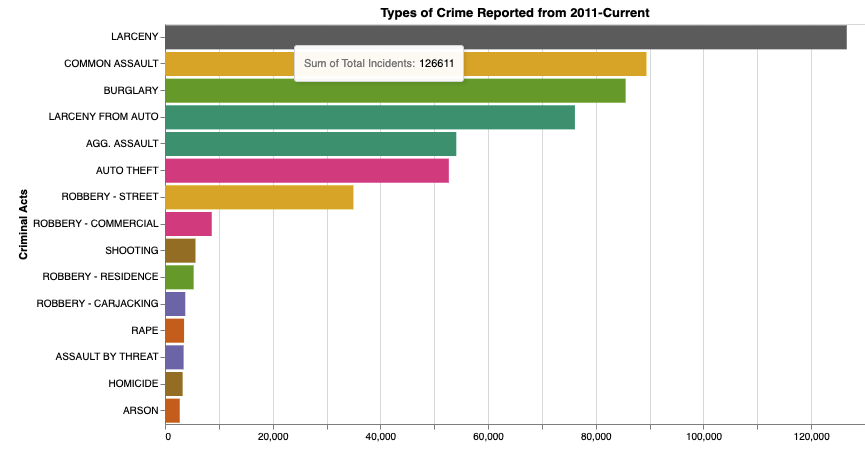
Next, I wanted to compare the amount of crime by neighborhood to see which areas are most dangerous at a smaller level. We see here that the Downtown area contributes the most crime with 18,000+ record incidents. It would be interesting to investigate what makes Downtown the most active neighborhood throughout the nine-year span.



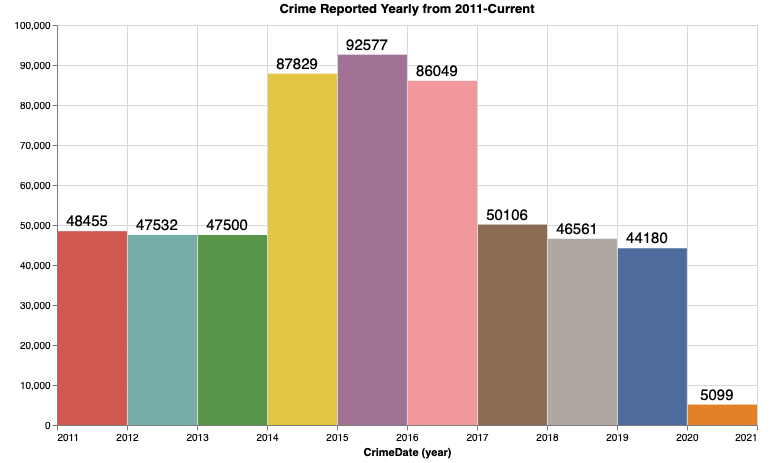
What I wanted to query next was what are the weapons, or non-weapons, being used to commit these crimes. As you can see, between the Knife, Other, Firearm, and hands, no type’s total reached over 50,000 from 2011-2020. Another interesting find is that more crimes were committed where the weapon used was by an unknown type that was not entered in. On the other hand, you will see that there were close to over 400,000 reports that did not list whether or not a weapon was used in the criminal act. A question that could be asked here is why and how are we reporting almost 400,000 unknown weapon types listed from 2011-2020. The strange but interesting part is that there is a substantial amount of weapon types, plus those unknown, that is not being properly reported whether it happened in or outside. Along with the other graphs, unknown weapon types are very significant.



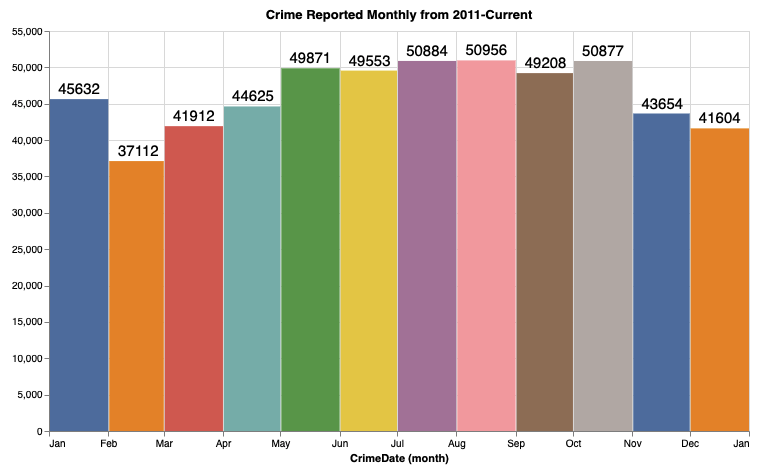
What I wanted to capture next was what were the actual criminal charges being reported and how frequently were they being committed. As you will see from the report, in almost a 9 year span the most common charge in Baltimore City is Larceny (theft of personal property) with 126,611 cases reported. The most shocking find is seeing that Homicide is the second to last common charge given in Baltimore City. I have structured the code in a way that when you click on a specific bar, the associated line to that bar will only show. Upon first observation of this chart it is shown that there are, more often than not, criminal acts occurring due to some category of theft happening.



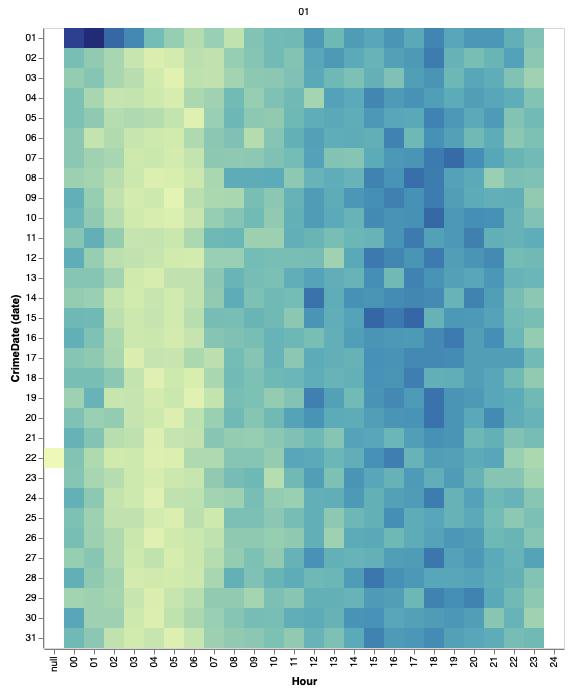
Soon after, I wanted to look at the levels of crime from the standpoint of years to months. As the graphs have shown, from 2011-2013, criminal activity became very consistent and stagnant during that three-year mark. When you continue onward from the ending of 2013, you see that there has been a spike in 2014, preceded by the highest recorded year in 2015 and then a slow descent in 2016. There is then a drastic collapse back down, followed by a slow decrease. I would compare this level of crime to the time of the Freddie Grey incident, along with the trial, and then the riots that soon took place within those three years (2014-2016).



After looking at the years, I queried the months and found that from the start of the year, criminal activity is somewhat high in January, and towards the middle of the year during the Summer months. During May-Aug you see the surge and consistency of crime occurring. I would attribute the rise and consistency in crime in the Spring and Summer months, to an increase in activity and more people gathering due to more desired temperatures.



Finally, after looking at the crime level from year to month I wanted to examine the crime levels per hour, ranging between the days in a month. I am only showing one figure, however, this figure stays true to the rest of the notebook on my GitHub link above. Looking at the chart below, I decided to create a heatmap for how much crime is committed over the course of the day in hourly terms. You can see below that the number of criminal acts is relatively low during the early morning, roughly between 2am-7am. As for the hot spots, these are roughly between 12pm-11pm. According to this, you will notice that there is a color change, or if you will, a gradual color change between 7am-12pm.



**CONCLUSION**

As expressed earlier, this was a top-bottom analysis of the crime levels in Baltimore City used to bring awareness to certain aspects, within the city, that may require additional investigation. We have seen that crime spiked tremendously during the span of 2014-2016, there are missing/unknown values in the datasets that should not be there, and there are criminal acts being committed in large numbers through a nine-year span. This data is still being researched and compared, as I search to compare and contrast different aspects of life such as: finances of residents, city projects/programs, as well as laws being implemented to try and grasp why Baltimore City is in the most dangerous cities in Americas list and how to change that.

**Reference:**

*Usatoday.Com*, 2020, https://www.usatoday.com/picture-gallery/travel/experience/america/2018/10/17/25-most-dangerous-cities-america/1669467002/.

2020, https://www.forbescom/pictures/mlj45jggj/7-baltimore/#7945c5915487. Accessed 19 May 2020.

*Data.Baltimorecity.Gov*, 2020, https://data.baltimorecity.gov/Public-Safety/BPD-Part-1-Victim-Based-Crime-Data/wsfq-mvij.