BALTIMORE CRIME ANALYSIS

Intriguing insights into crime in the City of Baltimore. Data analysis and modeling in Python.

KOLTEN SAFRON

SYNOPSIS

PROJECT BACKGROUND

For as long as I can remember the world of crime has always been something that was intriguing to me. I probably have my mom to thank for this, as she also shared this same interest with me. She was always there to watch crime documentaries and loved to tell stories from her constant crime mystery books she read. So I sat there and I figured what better way to put my interest in crime and passion for data together, than to do a crime analysis. Thankfully for me, the City of Baltimore posts crime data online so I was able to get a great data set for free. This consisted of each major crime that occurred in Baltimore from 2011 to present. Including such data as, crime date, crime time, neighborhood, district, crime code, crime description, and weapon.

My goal with this project was to dive into Baltimore's Crime and see what trends I could discover. I wanted to go beyond just looking at the crime data, so I tried to get creative and consider what else could affect crime and how could I try and draw conclusions from it. To take it a step further I wanted to challenge myself and see if I could create a machine learning model that would predict the number of crimes that would occur in each neighborhood on a given day or given month.

In order to save everyone time, I prepared a quick synopsis at the start summarizing the key takeaways from my data analysis as well as a summary of the model. Please sit back and enjoy the read as we dive into this together.

KEY TAKEAWAYS

The key takeaways from my exploratory data analysis done over the period 2011 through 2021 can be summarized as follows:

- One of the toughest parts about creating this model and the general data analysis, was dealing with the large amount of missing values and inconsistent data in the data sets used. Addressing this was probably the single largest part of this project as a lot of data cleanup was necessary to get it into a usable form.
- Crime does appear to have trends and correlations that allow us to forecast the crime activity as shown by my predictive models below.
- My data analysis and research led to the following findings having a correlation with crime frequency. These are covered further in the exploratory data analysis section later.
    - Neighborhood or district
    - Crime description
    - Year over year change in crime
    - Seasonality of crime as represented by its monthly trend
    - Crime throughout the month (day of month)
    - Crime per hour of the day
    - Crime per weekday
    - Crime occurring inside vs outside
    - Public events such as NFL games
    - City statistics or data

PREDICTIVE MODEL

Two different machine learning models were created to attempt to predict the number of crimes that would occur in each neighborhood over a given time period (day & month). Both models were created using the XGBoost machine learning algorithm (decision tree-based). The models incorporated much of the data and findings as discussed in the key takeaways above. The performance of these models was judged based off of their mean average error. This is the average error or difference that existed between the predicted crimes per day/month for each neighborhood vs the actual number of crimes that occurred on that day/month for each neighborhood.

1) Crimes per neighborhood on a given month (Ex. Crimes in Abcd Neighborhood in January 2018)
    - Mean average error of the model – 0.62
    - Average crimes per neighborhood that actually occurred – 13.3
    - Model accuracy – 95%

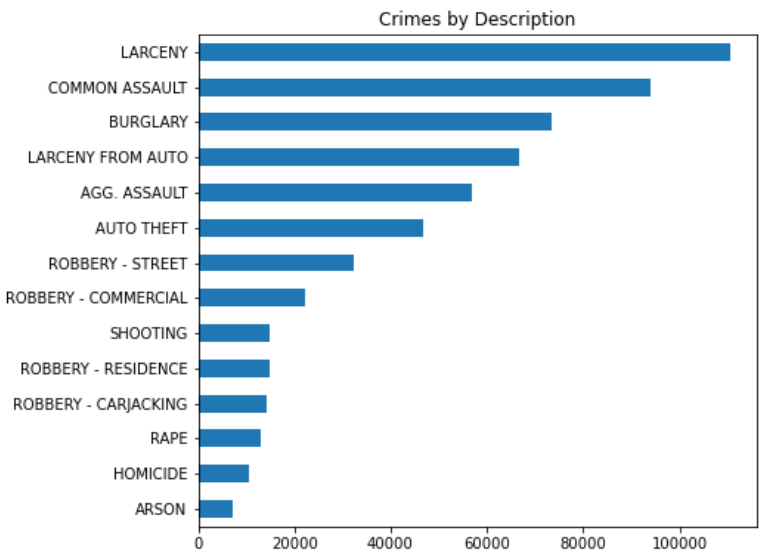
2) Crimes per neighborhood on a given day (Ex. Crimes in Abcd Neighborhood on January 8, 2018)
    - Mean average error of the model – 0.43
    - Average crimes per neighborhood that actually occurred – 1.57
    - Accuracy – 73%

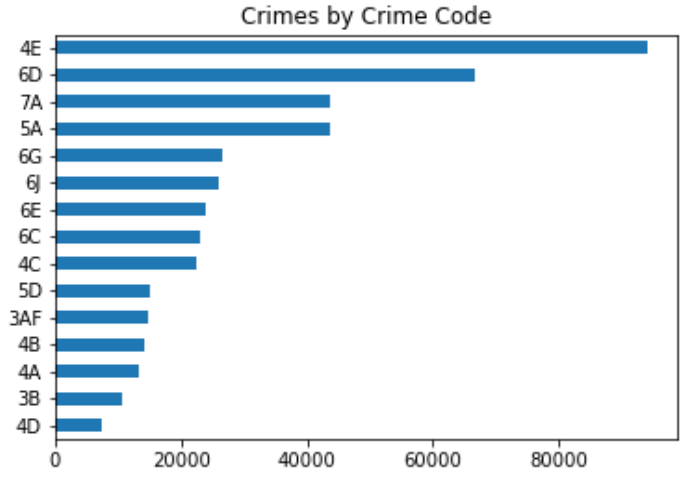
As you can see from the summary above, the model on a monthly level was much more accurate. This is to be expected as there are not enough crimes occurring in each neighborhood to allow for easy predictability. As seen above, only an average of 1.57 crimes occurring in a neighborhood each day in Baltimore.

EXPLORATORY DATA ANALYSIS

CRIME CODE AND CRIME DESCRIPTION

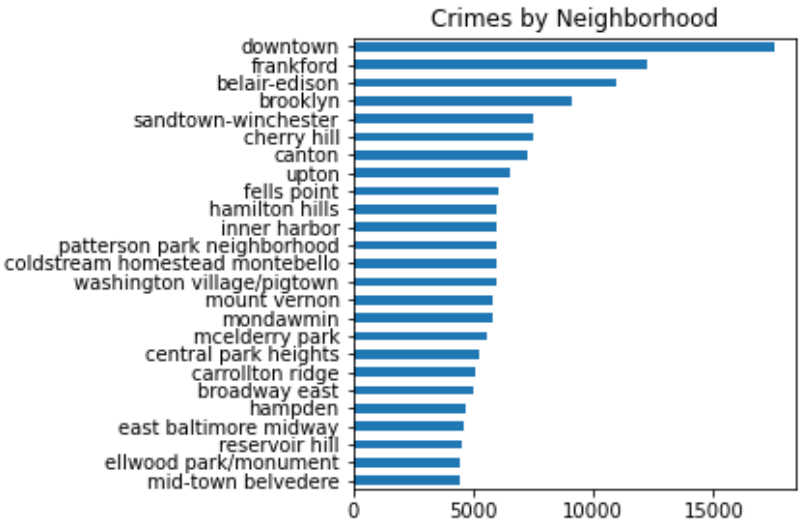
Data set included major crimes with 14 different descriptions (Ex. larceny, agg assault, robbery – residence, homicide) as well as 80 unique crime codes (descriptions broken down a level further). The top 3 crimes (larceny, common assault, burglary) made up 54% of the major crimes from the period. Of the 80 crime codes the top 15 were charted below. These 15 crime codes made up 87% of the total crimes, with the top 4 crime codes making up 48% of total crimes. We can see from this that there is significant concentration of crimes in a few categories. These categories as expected would be the less serious of the crimes included in this data set.

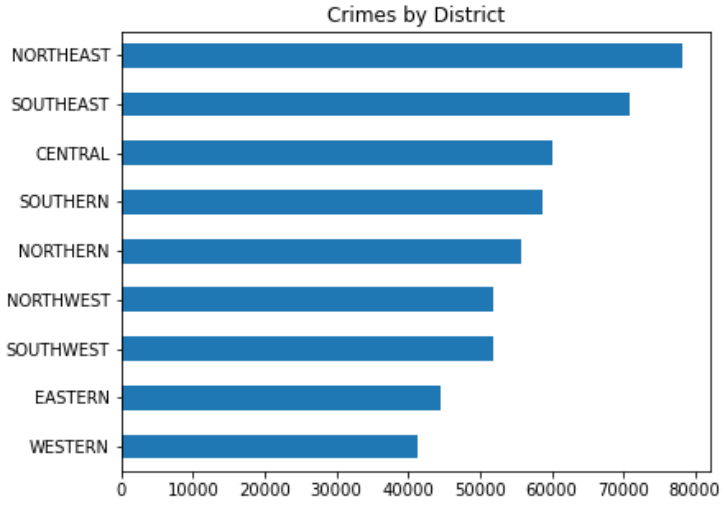




DISTRICT AND NEIGHBORHOOD

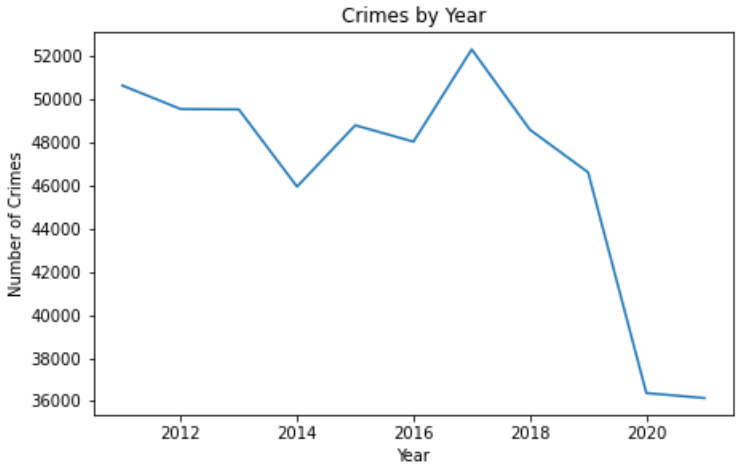
The city was broken down into 9 sections. The east side of town (northeast, east, southeast) consisted of the highest occurrences making up 38% of the total crime. With the northeast containing the highest occurrences of crime. When looking at neighborhoods specifically there were 280 neighborhoods left after some cleaning was done, the top 25 of which are charted below. These top 25 neighborhoods made up 33% of the total crime. From this we can again see there is a strong correlation with crime and the neighborhood.

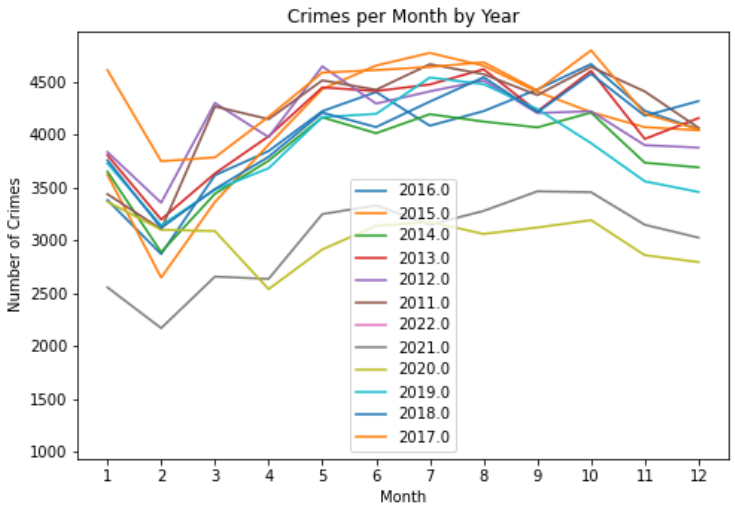




CRIME BY YEAR

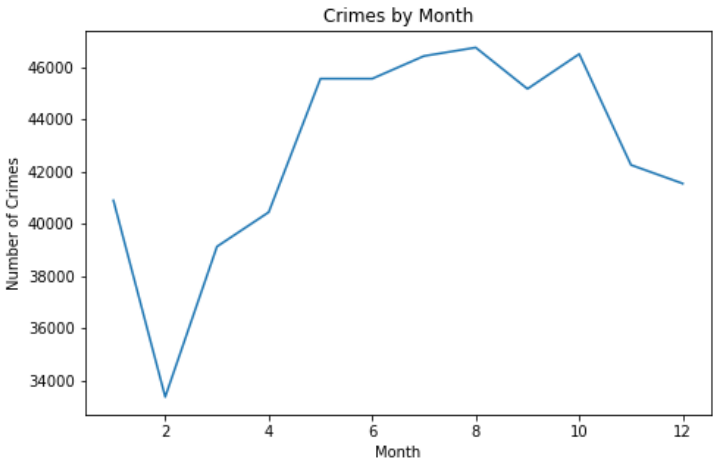
On a yearly level we can see crime has stayed relatively stable with the exception of a drop in 2016, jump up in 2017, and drop in 2020 & 2021. The significant drop in crime from 2020 & 2021 is likely a result of the COVID-19 pandemic. This is interesting given the negative affect on the economy and unemployment, people may become more desperate and crime would then increase. When looking at the significant increase in 2017 a likely contributor to the increase in crime may be the US Presidential Election. When reviewing the data, I saw that January 2017 was by far the highest amount of crime that had occurred in January through all the years, being 17% higher than the next closest. After January the crime started to level out over with the rest of the months. One thought that came to mind is that January 2017 would have been when President Donald Trump officially took office and as Baltimore (Maryland) has been historically a strong Democratic state people may have been upset leading to an increase in crime.

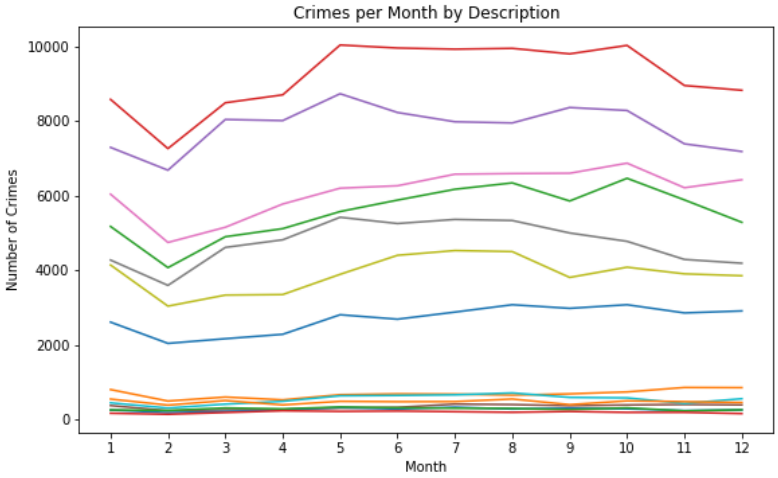




CRIME BY MONTH

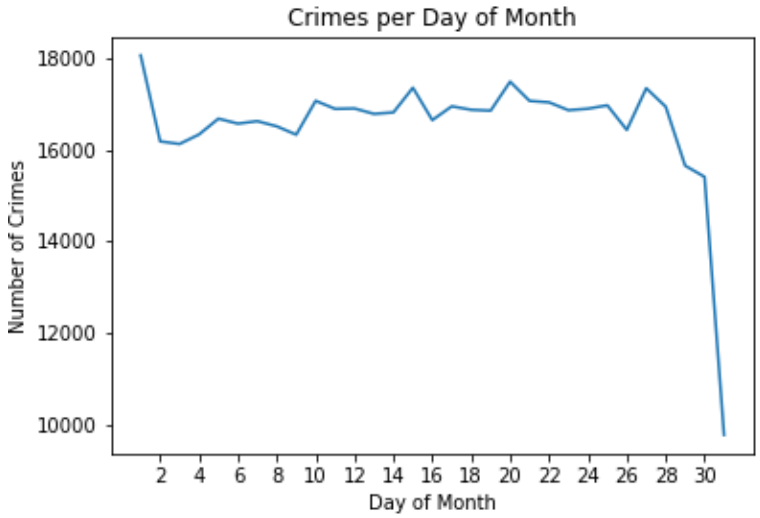
Crime was at its highest during May through October, afterwards it would slow down an average of 16% for the winter and early spring. February was the slowest month from crime, down from the average for the year by 21% or down from the high crime rate months of May through October by 27%. When looking at the crime descriptions the crimes all appear to follow the similar seasonal patterns. It does not appear that certain crimes are significantly more common during different months.





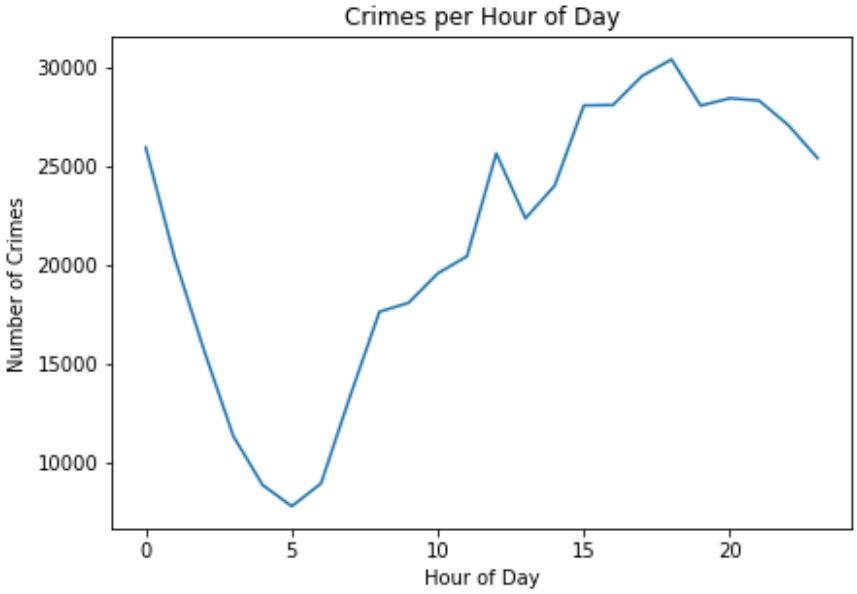
CRIME PER DAY OF MONTH

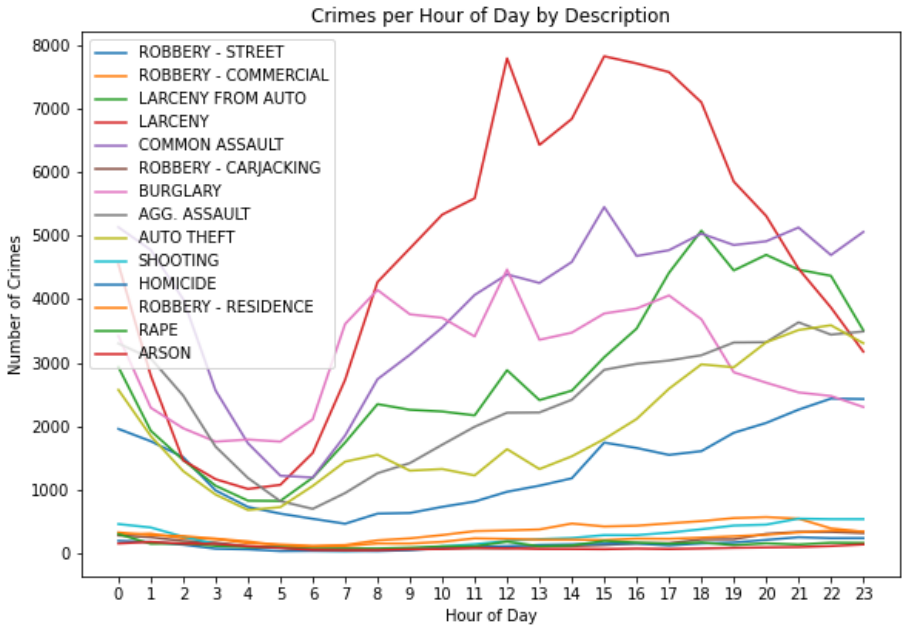
When looking at the crime as broken down by the day of the month we can see that the first day of the month being the highest crime rate by a large margin. It was 8% higher than the average and 5% higher than the next closest. We can see crime tails off towards the end of the month, noting a large drop in crime on the 28th forward. The drop after the 29th is likely due to fact that not all months have 31 days, and as such the 28th would be the last day where all months share that day. It is interesting that crime is the slowest towards the end of the month and the highest at the first day. To me this seems counter intuitive as I would expect crime to be higher towards the end of the month as bills become due, not slowest at the end of the month.



CRIME PER HOUR OF DAY

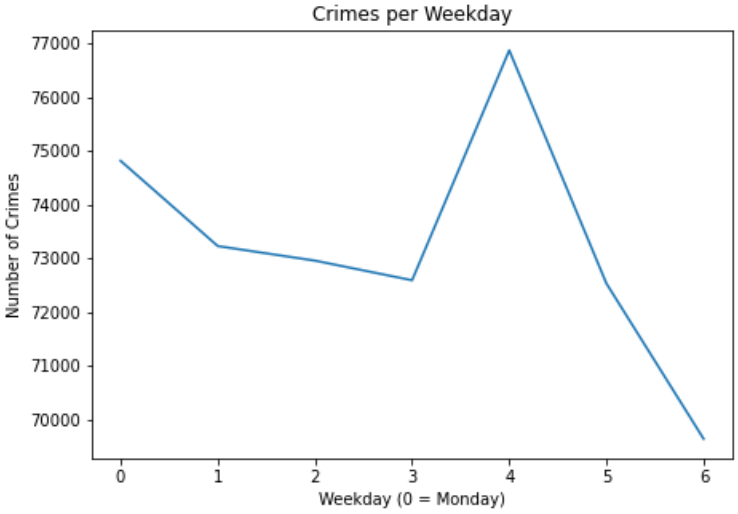
The time of the day, in hour increments, also has a large affect on the occurrence of crime. Crime is more common during the hours of noon to midnight. That half of the day makes up 63% of the crime that occurs. With the hours of 3AM to 6AM being the slowest making up only 7% of the total crime, yet it represents 17% of the total hours in the day. You can see that certain crimes tend to follow this overall trend of being slower after midnight through the morning, however some do not. For example, Larceny which this crime occurs almost twice as much from the hours of 11 to 17 compared to any other point of the day. Is the conclusion to draw here that Larceny is more likely to occur when people are gone to work?

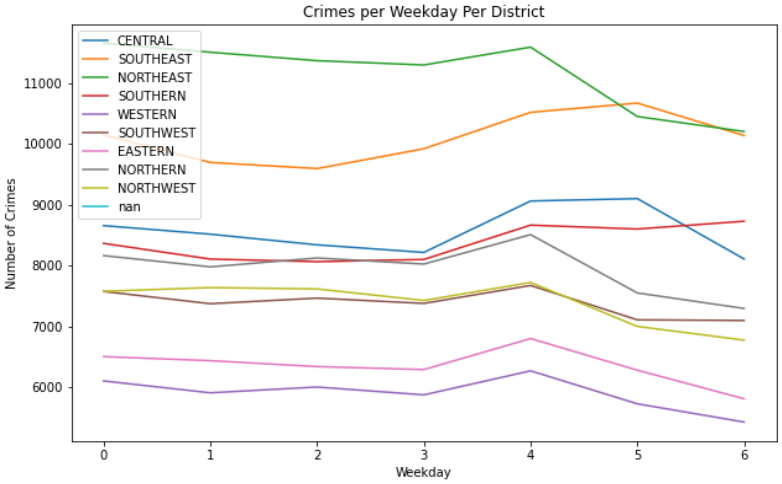


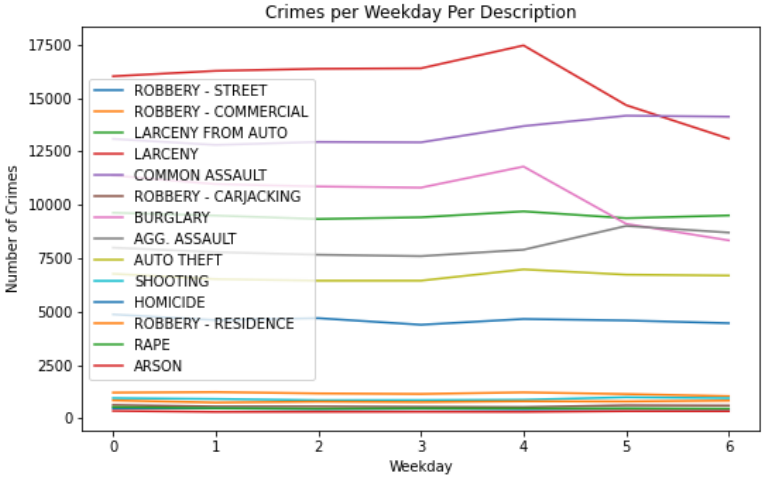


CRIME PER WEEKDAY

Crime is fairly consistent across the week with only small deviations. Friday has the highest crime with an increase of 5% over the average, whereas Sunday has the lowest with a decrease of 5% over the average. We can see that crime per district by weekday don't all follow the same pattern in each area. Most notably the Southern district, where crime actually is at its highest on the weekends, which is different from most all of the other districts. We can also see that certain crimes are more popular depending on the weekday. Take larceny and burglary, both of these crimes fall off significantly come the weekend where all the other major crimes tend to stay relatively consistent. A likely reason for this is that more people are at home for the weekend so those crimes aren't as likely to occur.

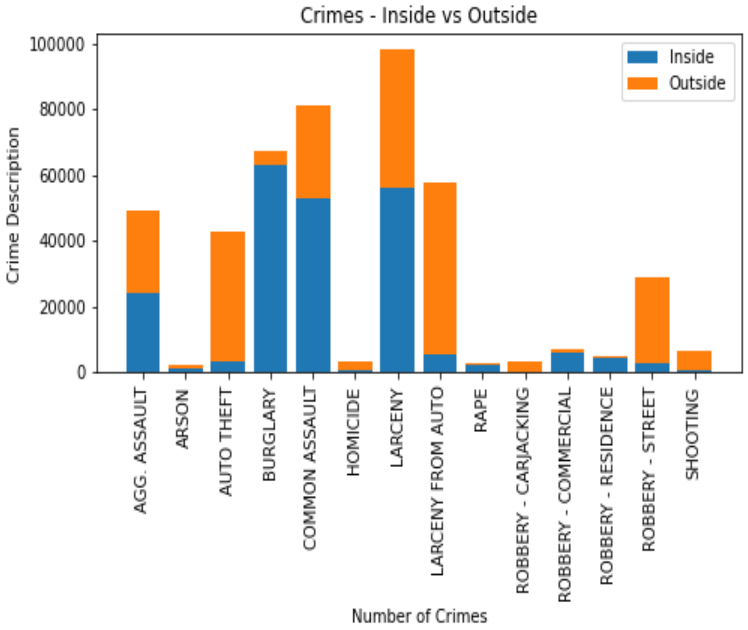


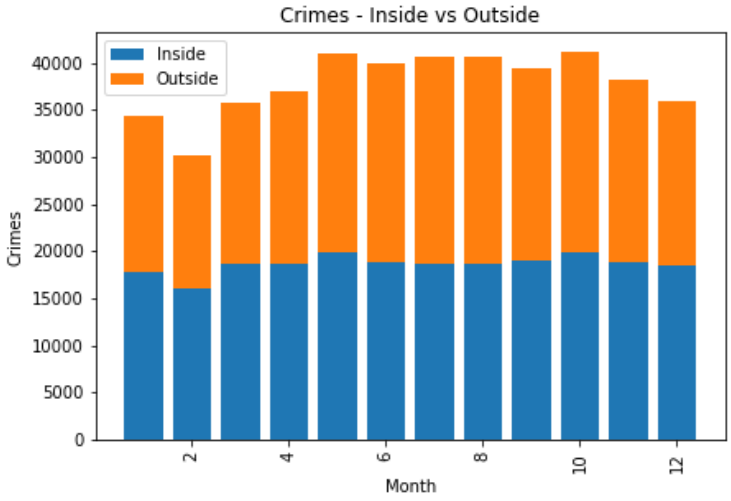




INSIDE VS OUTSIDE CRIMES

Of the total crimes 43% of them occurred inside where as 57% occurred outside. Excluding any auto related crimes, 54% of the crimes incurred inside whereas only 46% occurred outside. We can see that the amount of crime that occurred inside stayed relatively consistent throughout the year (with the exception of February as it is the slowest month for crime in the year as noted earlier). So the significant increase in crime that occurs from May through October happens primarily outside as expected due to the nicer weather in those months.



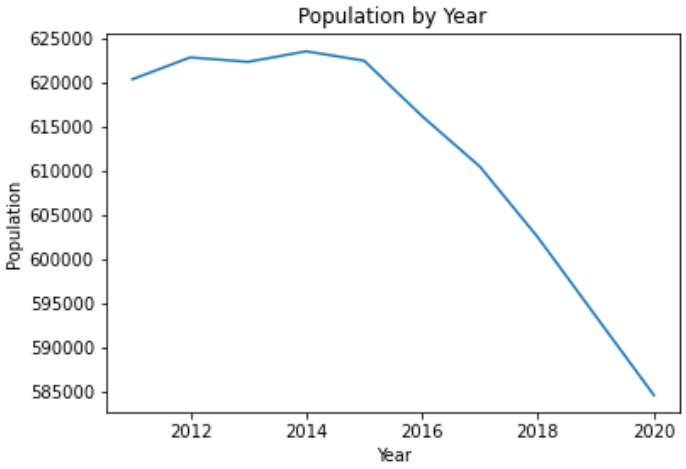


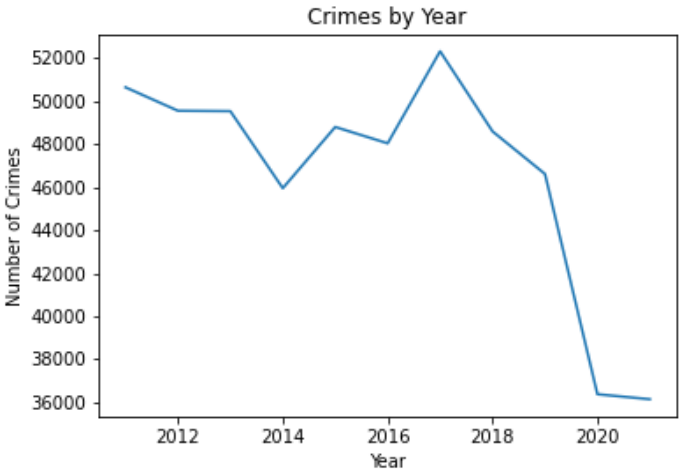
NFL GAMES

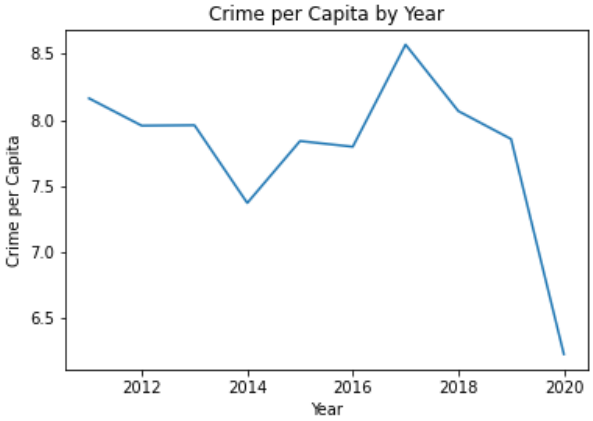
Looking at a large public event like a NFL football game, there appears to be little to no correlation with an increase in crime. In the downtown area, where the football stadium was, there was barely any change in crime. However interestingly enough when the final model was produced, the model did put a fair amount of value in the NFL games, showing that this likely does in fact increase crime it may just not be as apparent as other variables.

BALTIMORE CITY DATA

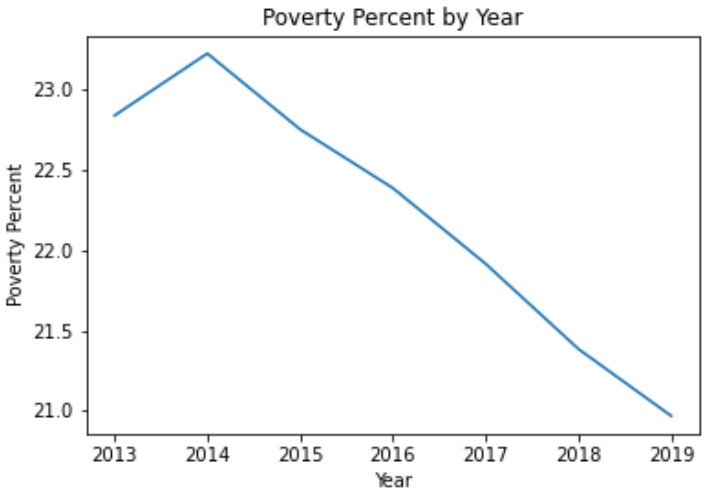
Population – The population has been decreasing consistently. This should lead to a decrease in crime, however this is not perfectly apparent when comparing crime by year with population. When looking at the crime rate per capita we can see that even though the population decreased 2.3% combined in 2017 & 2018 the crime rate increased an average of 0.5%.

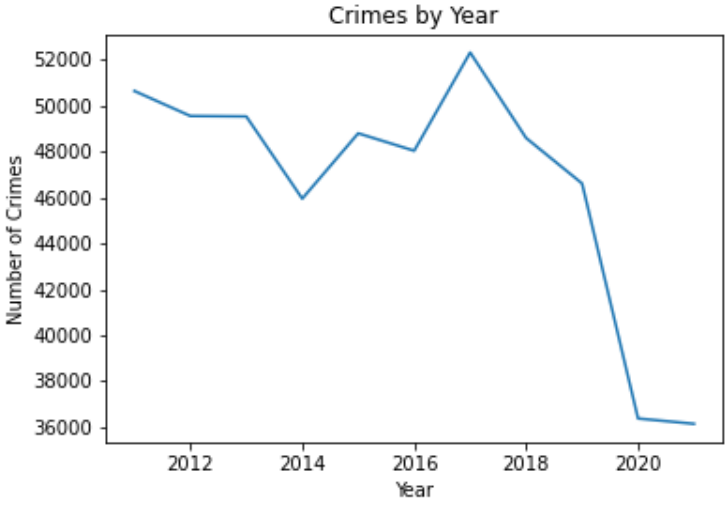




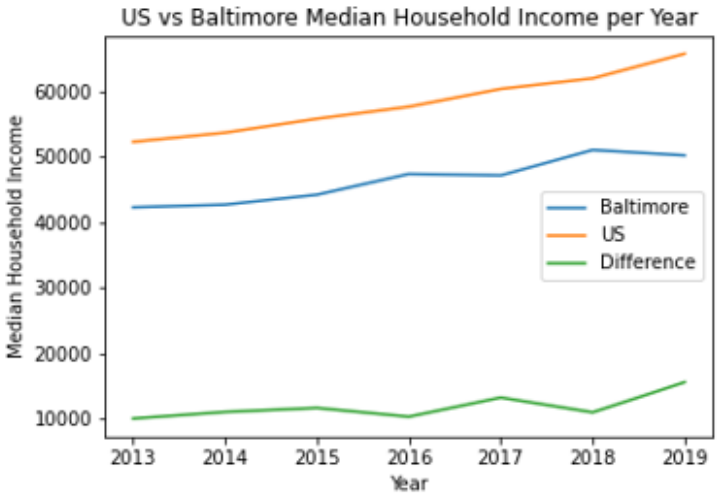


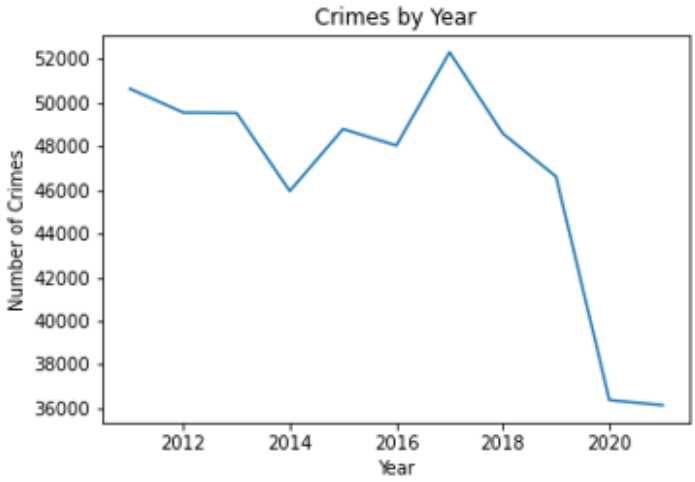
Poverty Population – One of the surprising findings was that the amount of the population below the poverty threshold has been consistently dropping since 2014 (23.2% of the population in 2014 vs 20.9% of the population in 2019), but there doesn't appear to be a correlation with the drop in crime.



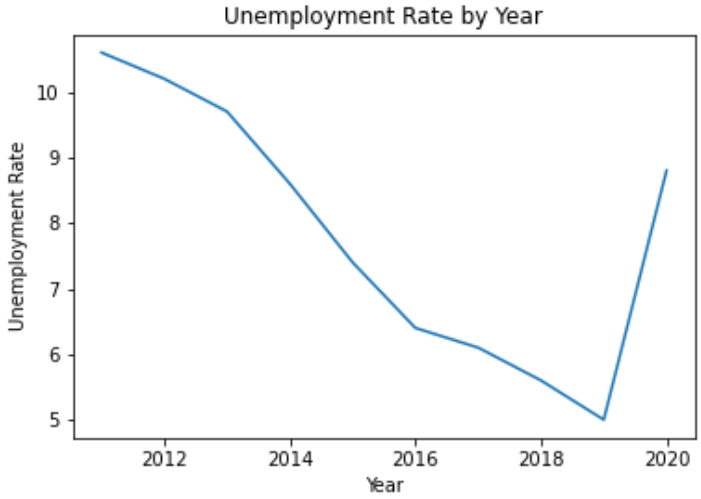


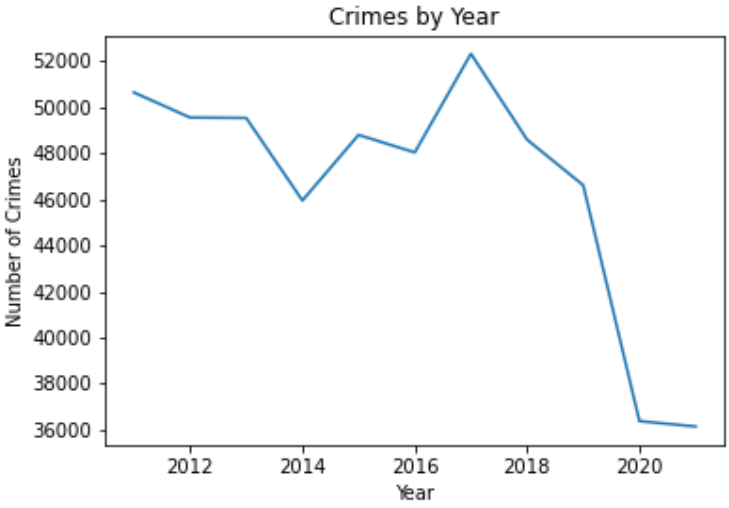
Median Household Income – Incorporating the median household income in Baltimore vs US seemed to provide little correlation with the overall crimes. However, in the predictive model both of these data points had a strong enough weight assigned to them, to indicate that is a correlation exists.





Unemployment rate – There does appear to be a correlation with unemployment rate, as the unemployment took a significant drop from 2011 to 2015, dropping from 10.6% to 7.4%. We can then see that through those years, there was a corresponding decrease in the crimes that occurred.





REFERENCES

LINKS MY TO WORK

Main repository for project - https://github.com/kds55/Baltimore-Crime-Analysis

Project code as completed in python - https://github.com/kds55/Baltimore-Crime-Analysis/blob/main/baltimore_crime_analysis_code.py

DATA SOURCE

City of Baltimore Crime Data:

- January 2013 to November 2016 - https://data.world/data-society/city-of-baltimore-crime-data
- January 2017 to present day - https://data.baltimorecity.gov/datasets/part1-crime-data/explore

Baltimore City Data:

- Poverty, household income - https://datausa.io/profile/geo/baltimore-city-md
- Population - https://worldpopulationreview.com/us-cities/baltimore-md-population
- Unemployment - https://msa.maryland.gov/msa/mdmanual/01glance/economy/html/unemployrates.html