Nathan Grant

12/11/2022

Impact of COVID on Violent Crime in Baltimore City

1. Introduction

Baltimore City had experienced a large spike in crime following the death of Freddie Gray in 2015. In 2014 the homicide rate in the city was 33.8 per 100,000 , compared to 2015-2020 where the rate jumped between 50.5-58.6 [1]. With the onset of the pandemic, these crimes increased even more and this analysis attempts to understand the dynamics of these two systems. Outside of various news outlets citing homicide statistics and other human accounts of the escalating crime scene in Baltimore, I could not find any other analysis that tried to find causal relationships between case data and crime.

By testing whether these two variables are causally related, you can get a deeper understanding of the systems of violence within the city. It will give you greater insight as to how reluctant people are to perpetrate violence given the severity of the virus. It can also give you insight into the periodicity of peaks in violence allowing for better preparedness for law enforcement and other municipal organizations.

I also tested to see if any of the policy implementations had an impact on the spread of COVID. This was so that in the future, policies that might have been effective could be applied to reduce the spread of a future virus.

2. Background/Related Work

In this analysis I hypothesize that there is a causal relationship between the percentage change in covid cases and daily call volume for things related to shootings. Since there is no daily data available for homicides in the city, I needed a proxy for something that was closely linked to this statistic. The Baltimore PD releases call logs every year with data about each call, including the reason, urgency, date, and location.

Other analysis has been done in this area including the work at Baltimore Neighborhood Indicators Alliance [2]. This group has done thorough exploratory data visualization on quarterly crime data for Baltimore. This analysis also takes into account geographical information as well so that users can see the crime centers of the city and the spatial dynamics of crime through time as well. They have not however gone to the lengths of applying descriptive or prescriptive analysis on this data. Their analysis serves as a solid barometer for what I should expect from my model.

3. Methodology

The goal of the analysis was to determine if there was a Granger causal relationship between case data and high priority calls related to shootings. There were a multitude of possible models I could use for this specific case, however I chose to use Cross Convergent Mapping [3]. This technique uses commonly used properties of dynamical systems to model the relationship between two variables. This technique is also more robust than other granger causality tests which is the reason I decided to go with it.

Before doing predictive analytics over this data, I first wanted to use this technique to see if they were causally related. It seemed that if I was to do predictive analytics first, then their conclusions would be meaningless if there wasn't an underlying relationship between the two. Also, the hyperparameters of this model have a story to tell about the interaction between the two variables. The lag parameter of the model sets the sampling period for when data is selected from the time series, indicating that some recurring pattern is happening that gives more information to predict the next time point. Less important but also relevant is the embedding dimension. This tells the model how many previous lags to take into account.

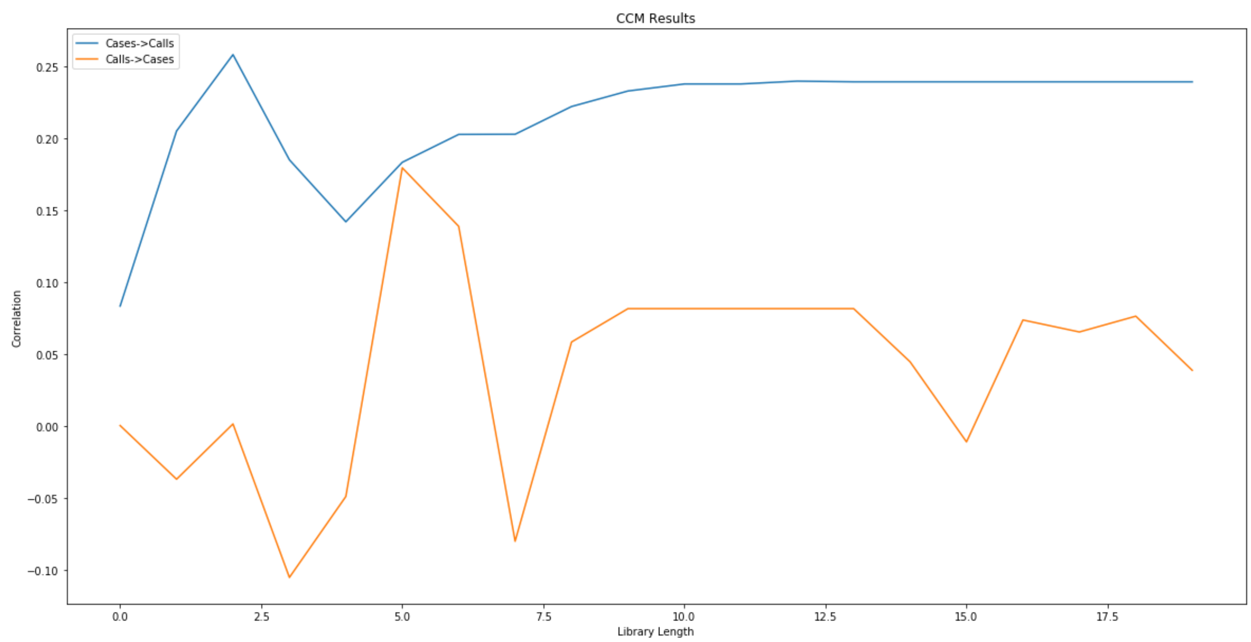
Both of these parameters are of interest because if there is a causal relationship, then they can give you insight into the cyclical nature of the dynamics. In terms of explainability and adoptability, it seems that this would be very useful for law enforcement agencies to be able to have a simple and powerful heuristic for predicting spikes in crime. It would be immediately understandable to tell people not well versed in machine learning that whatever the situation was last week on some day will give you a solid understanding of how it is going to be today.

The other avenue of research analysis was how I could tie in the policy implementations to see how they impacted the spread of the disease. Maybe, if I could find a causal relationship between COVID and crime, then I could also extend that relationship to look further into what it was about the policy changes that could have impacted crime.

For this task I took the moving average of the percentage daily change in covid cases. Then over this curve, I took the timepoints for when each policy was active and got the slope of the curve. A negative slope would indicate that the policy was effective in reducing the spread of the disease, and a positive slope the opposite. I then categorized each of the policies in terms of their stringencies with respect to the categories that were in each of the documents. The main things that I found had changed were the allowance of people into theaters, social distancing guidelines, and also restaurant occupancy limits.

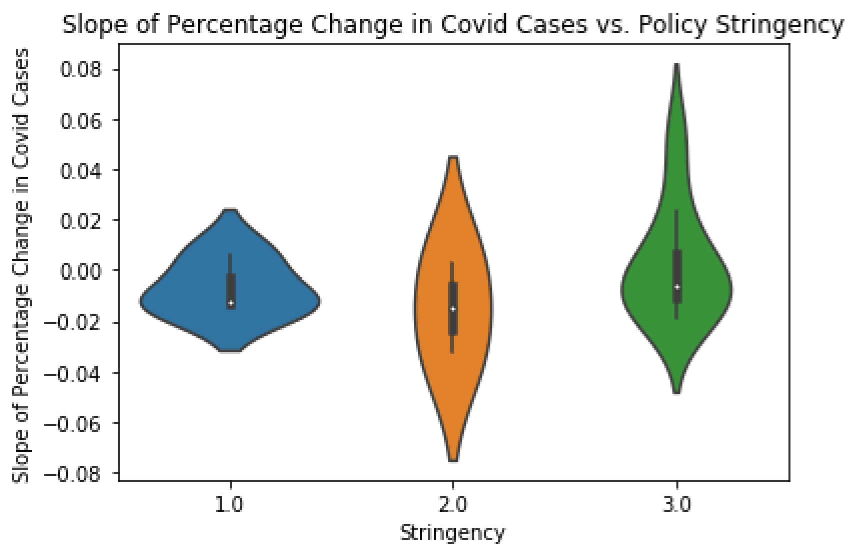
4. Findings

The results from the analysis were inconclusive. The final mode of validation for the CCM algorithm is to evaluate the plot for correlation of predictions as a function of library length. Library length is the number of timesteps incorporated in the training data, so we should expect that as we incorporate more and more data, the predictions get better and better. This was not the case however.



As you can see from the above graph, as library length increases the predictive power of cases to calls stagnates at quite a low correlation. If the correlation was higher, or we had more data, then we could say that these two have some sort of causal relationship. But under a model trained on up to 20 data points, there is no ground to assume that the two are causally related.

Under this model I did find that the optimal library length was 11 and embedding dimension 8. Since they are not causally related this has less significance, however this is still an interesting result. This shows that the most causal impact on the dynamics of today's number of calls is best predicted by data from 11 days ago (and also the previous 7 weeks before that).



As for the policy effectiveness analysis, there were no significant differences found between policy groups. There is a skew in the number of policies in each group. Group 1 has 3 policies, group 2 only having 2, and group 3 having 9 points (groups going from least to most stringent). Since the sample sizes were too small, there was no evidence to suggest that any of the groups were significantly different.

5. Discussion/Implications

Studying the dynamics of populations is an inherently human centered pursuit. While the findings were inconclusive, I believe that this line of research could have wider implications. For example we don't know the tradeoff between crime and restrictions on people staying indoors. We would expect that if we were in a martial law setting and everyone was forced to stay at home then there may be less of certain kinds of crimes. Or on the other extreme, if it's anarchy then maybe there would be more crime. The in-between ground is even more obfuscated.

A study like this would allow you to peek into the dynamics of the system of crime and human behavior because we have a population under two different conditions of interaction restrictions.

Also the scope of this study was limited in that only one county was used. Maybe when repeated over many different counties you could find different relationships between crime and COVID that are dependent on some other variable.

The policy effectiveness study requires more than just the policy categorizations in order for it to be meaningful. Since not much was changing in between policies, we need something like foot traffic data to see exactly how much interaction there was during those policies and see if we can attribute the changes in cases to policy intervention.

6. Limitations

One potential assumption that is flawed in the data is that the number of calls for homicide or shootings is proportional to the number of homicides. There was a large spike in call volume during the months of December 2020 to March 2021 that I couldn't find a reason for. Meanwhile the homicide statistics during that time did not reflect that large increase [4]. This led me to think about the shortcomings of using this as a proxy for high priority violent crimes. Another thing that impacts the proportionality of the two variables is that if people were more likely to be at home during this time, then maybe they were also more likely to call if something was happening. If multiple people were calling for the same reason then this would also throw off the numbers.

Another assumption was that of the categorization of policies. Some simplifying assumptions were made so that I could lower the total number of groups. The reasoning behind this was that policies were not that different amongst things that are commonly known as large spreaders of the virus. Many of the policy changes had to do with the ability of local government officials to have meetings together in person, which seemed like a low risk event in comparison to say increasing restaurant occupancy from 25 to 75 percent. Under different categorizations, even if there were more, then maybe there would be larger differences between groups.

One last assumption was that the policy term ended when the next one started. In reality the effects of one policy would have been felt for some time afterwards but the line between starting one and stopping the other is blurry, so I went for a simplifying assumption instead. For some policies this could have impacted whether or not the policy had a positive or negative slope, but for most the same trend would have been found.

7. Conclusion

Restate your research questions/hypotheses and summarize your findings. Explain to the reader how this study informs their understanding of human centered data science.

The results from the study suggest that there was not enough information to conclude whether or not COVID cases and violent crime calls were causally related. I was also unable to find whether or not the policy implementations were effective in stopping the spread of COVID.

This study served as an interesting first step towards modeling crime dynamics in Baltimore city. Because it's hard to get high frequency data for crimes, this project had to find other avenues to make inferences about the crime climate. Because of the large volumes of news articles talking about homicide deaths in Baltimore, the goal of the study was directed to try to find out if a reason could be pinned down for their increase. By understanding some part of the dynamics, policy makers and law enforcement agencies could then use the information to better help reduce crime rates.

8. References

A list of publications (blogs, articles, research papers) that you refer to in your text.

1. https://en.wikipedia.org/wiki/Crime_in_Baltimore
2. https://coronavirus-bniajfi.hub.arcgis.com/
3. https://www.nature.com/articles/srep14750
4. https://www.baltimorepolice.org/crime-stats

9. Data Sources

1.https://data.baltimorecity.gov/datasets/911-calls-for-service-2020/explore

2.https://data.baltimorecity.gov/datasets/911-calls-for-service-2021/explore

3.https://www.kaggle.com/datasets/antgoldbloom/covid19-data-from-john-hopkins-university

4.https://data.cdc.gov/Policy-Surveillance/U-S-State-and-Territorial-Public-Mask-Mandates-Fro/62d6-pm5i

5. https://github.com/nytimes/covid-19-data/tree/master/mask-use

6.https://governor.maryland.gov/wp-content/uploads/2020/04/Masks-and-Physical-Distancing-4.15.20.pdf

7.https://governor.maryland.gov/wp-content/uploads/2020/07/Gatherings-10th-AMENDED-7.29.20.pdf

8. https://governor.maryland.gov/wp-content/uploads/2020/08/2020-08-03-11-08.pdf

9.https://governor.maryland.gov/wp-content/uploads/2020/09/Gatherings-12th-AMENDED-9.1.20.pdf

10.https://governor.maryland.gov/wp-content/uploads/2020/09/Gatherings-13th-AMENDED-9.18.20.pdf

11.https://governor.maryland.gov/wp-content/uploads/2020/09/Gatherings-14th-AMENDED-9.28.20.pdf

12.https://governor.maryland.gov/wp-content/uploads/2020/10/Gatherings-15th-AMENDED-10.16.20.pdf

13. https://governor.maryland.gov/wp-content/uploads/2020/11/EO-11.10.20.pdf

14. https://governor.maryland.gov/wp-content/uploads/2020/11/Order-20-11-17-01.pdf

15. https://governor.maryland.gov/wp-content/uploads/2021/01/Gatherings-18th-AMENDED-01.28.21.pdf

16. https://governor.maryland.gov/wp-content/uploads/2021/03/Gatherings-19th-AMENDED-02.23.21.pdf

17. https://governor.maryland.gov/wp-content/uploads/2021/03/Gatherings-20th-AMENDED-3.9.21.pdf

18. https://governor.maryland.gov/wp-content/uploads/2021/04/Gatherings-22d-AMENDED-4.28.21.pdf

19. https://governor.maryland.gov/wp-content/uploads/2021/05/Gatherings-24th-AMENDED-5.14.21.pdf