Ariel Shurygin

Project Write up

Introduction

Understanding the sources of crime is a body of work spanning many fields of study, economics, psychology, and sociology to name a few. Looking at crime from any of these perspectives always has one similarity, crime is a cause and effect relationship. Certain actions taken by the federal reserve, or an individual, or society at large influences the crime rates in neighborhoods across the USA (figure 1). During the COVID-19 epidemic the United states saw a surge in crime rates not seen in many years. This study makes some assumptions about pivotal moments during the pandemic and analyzes their impacts on certain kinds of crime in Denver county Colorado.

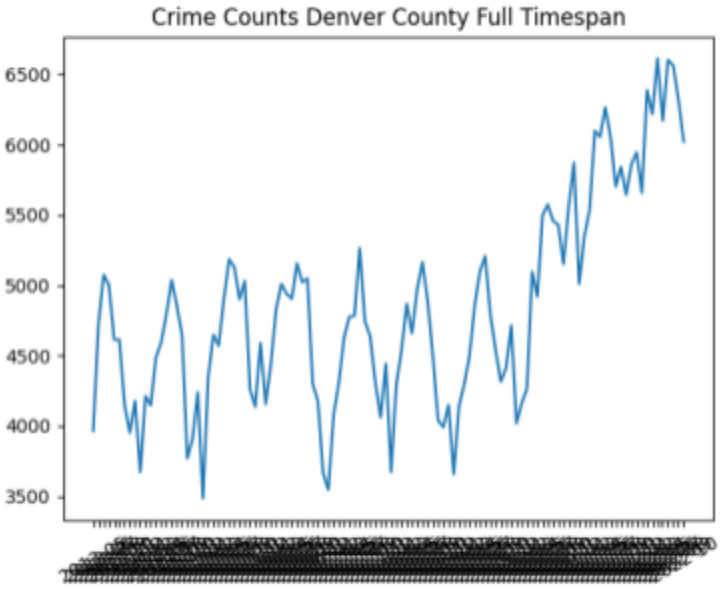

Figure 1: Crime Counts Denver County 2013-2022

Background/Related Work

Many conversations have been had in popular media about crime during the pandemic, mostly focusing on fear mongering around prisoner releases (1) or lawlessness in the streets (2). Many media outlets had ulterior motives to promote some sort of political action in one way or the other as is often the case in American politics. However there is a depth of research around the push and pull factors around crime (3). In particular many studies point out the positive correlation between unemployment rates and crime. The clash between popular reporting about Covid induced crime and existing studies was the primary motivation behind each of the three hypothesis which are as follows:

1) The introduction of mask mandates increased public disorder crime by 15%, as those who actively broke masking mandates put strain on police forces
2) Lock down increased domestic abuse by 20% as victims no longer could avoid their abusive partners and were more likely to be stuck in the home with them.
3) Stimulus checks reduced robbery and financial crimes by up to 10%, as people could finally put food on their tables, even for a week or two

Denver County Covid-19

Denver county is the county surrounding Denver, the largest city in Colorado, home to around 700,000 people in 2020 and with a land area of 154 miles squared. Over the course of the pandemic the county followed the same trend as most other major metropolitan areas in the United States, enacting masking policies in response to large rises in COVID cases and removing them once those cases reduced. Here is a chart of the weekly case count, the total case count, and whether or not there was a masking order in the county.

Figure 2: Weekly/Total Covid cases (red/blue) and masking mandates (orange)

Methodology

Before deciding how I would predict changes in crime rates I looked at crime before the pandemic, specifically 2013 - 2019. I noticed a fairly predictable trend with peaks in crime rates in August and troughs in January. Using the seasonality of crime I could create a distribution of crime for a particular month. For example the month of August has higher than average crime so it would be modeled by a slightly different distribution than January crime. Assuming

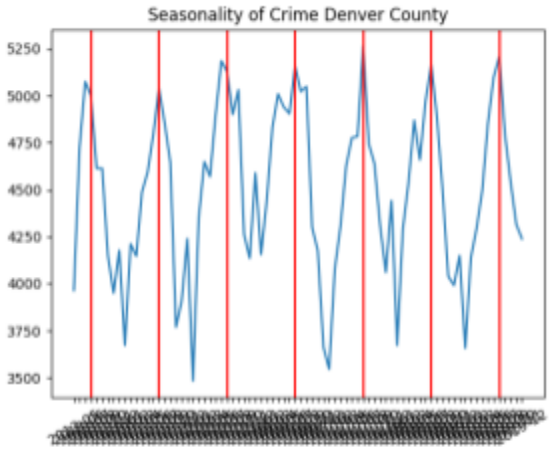

Figure 3: Crime rates Denver County with each August highlighted

negligible shifts in Denver County politics we can use a normal distribution from year to year with mean and standard deviation to model the likelihood of a certain amount of crime in that month. With the generated crime distribution cumulative density functions can test whether or not the pandemic had a statistically significant impact on crime.

Making accurate predictions about crime is a difficult human-centered issue as it attempts to turn every victim and perpetrator into a number. Without ethnographic research as to why an individual commits crimes the personal stories are lost in a fog of data. Unfortunately the research time frame did not permit flying to Denver county for interviews with victims and those in prison for a deep dive into the Denver crime scene. To compensate for these ethical issues I attempted to choose statistical analysis methods that made as few assumptions as possible. In short I made two major assumptions. Firstly, crime for a particular month follows a normal distribution, meaning the crime rates of august last year do not impact this august and they have an average value they tend to stay around. Secondly, I assumed the crimes themselves were independent, meaning if domestic abuse rose this was not necessarily correlated with a rise in robbery crime. I believe for the crimes studied in each of the three hypotheses the independence of crime rates does apply, while this assumption may not hold for all crime combinations.

Findings

After digging into the data I found that my findings did not line up with research in the field, nor did it follow my own expectations. I was unable to disprove the null hypothesis for my first two predictions, and actually showed the opposite trend in my third hypothesis. Let it be known before figures are shown that a lack of data in this analysis may have contributed to a false narrative being painted. Statistics in combination with incomplete information can give incorrect or misleading results (2).

Figure 4: Statistics without the full story may be misleading

Hypothesis 1: The introduction of mask mandates increased public disorder crime by 15%, as those who actively broke masking mandates put strain on police forces.

In short this hypothesis could not be shown in the data, in this scenario the null hypothesis was that the introduction of mask mandates had no statistically significant impact on public disturbance crime when

compared to pre pandemic years. The mask mandate was first introduced in August of 2020, meaning that if we compare the number of public disturbance calls to police in August 2020 vs all other pre pandemic years, there should be some statistically significant difference, here are the results:

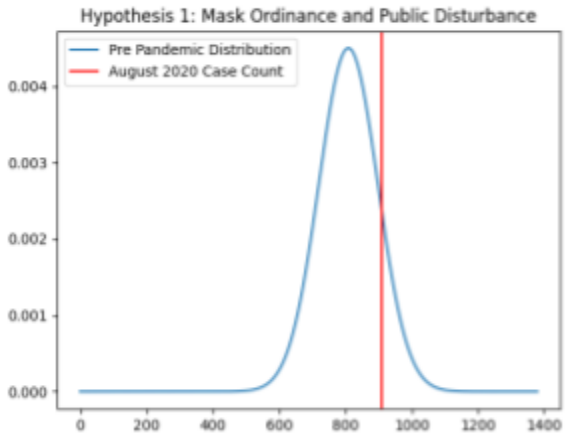

Figure 5: Case Distribution August pre Pandemic Compared to Observed Value

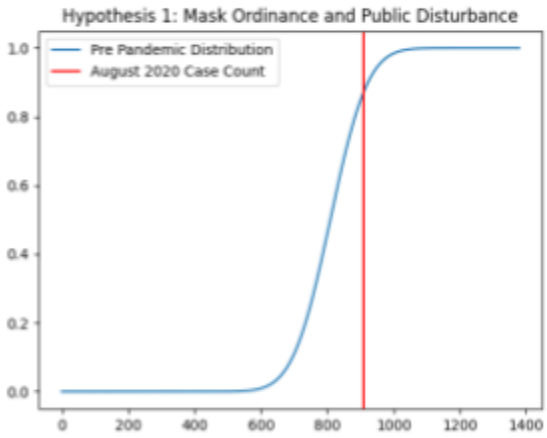

Figure 6: Public Disturbance CDF

In short we could not disprove the null hypothesis because there was a 13% chance that the month of August had a public disturbance case count at or above the observed value in 2020, which was 909 cases of public disturbance. This is the importance of taking into account standard deviation. While the observed crime rate in 2020 (909 cases) was 12% higher than the average rate for the years before it, the standard deviation of the pre pandemic distribution accounted for this shift.

Hypothesis 2: Lock down increased domestic abuse by 20% as victims no longer could avoid their abusive partners and were more likely to be stuck in the home with them.

Yet again this hypothesis could not be shown with statistical significance. The null hypothesis in this case being no significant change between domestic abuse rates during the lockdown as opposed to the same period of time before. In this case the lockdown occurred From March 26th to April 26th 2020 (4).

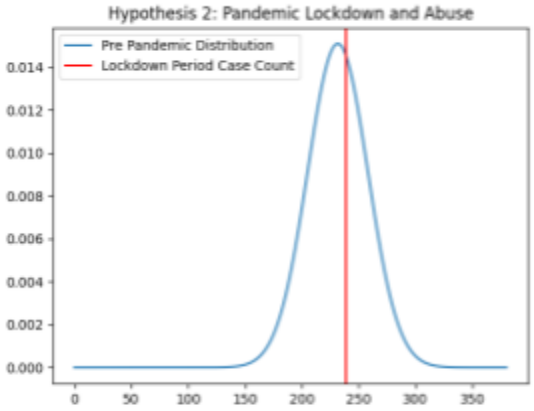

Figure 7: Case Distribution March-April pre Pandemic Compared to Observed Value

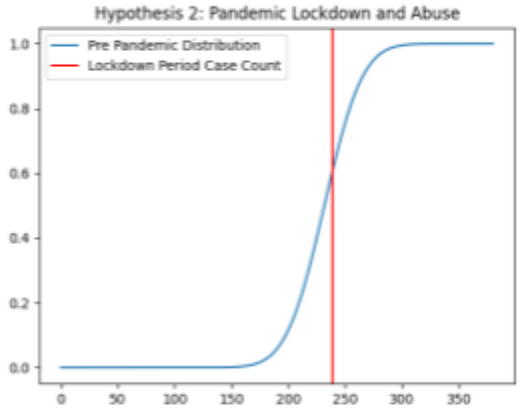

Figure 8: Domestic Abuse CDF

Again we could not disprove the null hypothesis because there was a 38% likelihood that the observed domestic abuse case count (239 cases) could appear in the pre pandemic distribution. Again while the observed value was 3% higher than the average for before the pandemic, this was covered by the standard deviation between the years of 2013-2019

Discussion/Implications

Why are your findings important or interesting; How could future research build on this study?

This section should include a thoughtful reflection that describes the specific ways that human centered data science principles informed your decision-making in this project.

Limitations

This is a required section for your report.There are often many, many limitations for any study. If you honestly tried to list them all, this might end up being the longest section. You should prioritize and list the ones that are most likely to have a significant impact on your results. Specific license issues could be a limitation, depending on what data sources you used. Flaws in the data, data cleaning techniques, potential assumptions and/or how you handled missing values could be a limitation. Statistical techniques often have specific assumptions and preconditions; if you're not certain all of the data meets those requirements - this is a good place to make that clear.

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.

References

1) https://www.foxnews.com/politics/kamala-harris-wants-low-risk-inmates-released-amid-fears-of-covid-19-spreading-in-prisons
2) https://www.foxnews.com/media/what-happens-land-becomes-lawless-dr-oz
3) https://www.ojp.gov/ncjrs/virtual-library/abstracts/economic-factors-crime-and-delinquency-critical-review-empirical
4) https://covid19.colorado.gov/public-health-orders-and-executive-orders

Data Sources

A list of links to the relevant data sources that you used.