COVID-19 and Crime Rates

Objective

Covid-19 forced people within their homes for almost 2 years which begs the question, with such forced measures how have crime rates and violence reacted to the new environment and situation? Using predictive modelling, can a correlation between such seemingly unrelated topics be uncovered? Has the news accurately represented the situation within New York along with police brutality?

Background

Finding a correlation between isolating catalysts such as an epidemic can lead to increased response time in violence and crime rates. Studying covid is especially important due to its abrupt emergence which could have heightened the possible response of the population. This study can be applied in many fields such as stress analysis, household relationships, etc which allows us to be more aware towards these situations and understand them at a greater magnitude. Few drawbacks that could exist are that the number of reports could have decreased in parallel to the increase of cases which would yield a null result in this study. Others include, lowered frequency of obtaining data due to covid, higher stress levels leading to more paranoia etc.

Resources:

https://data.cityofnewyork.us/Public-Safety/NYPD-Arrest-Data-Year-to-Date-/uip8-fykc

https://www1.nyc.gov/site/nypd/stats/reports-analysis/use-of-force-data.page

Solution

I will use pandas, numpy, most recent python, seaborn, and sql to visualize and predict how pandemics can affect police presence including arrests and police brutality. I will produce models to show crime rates, police brutality and arrest reports per borough as well as covid models for the same categories and see if there is a correlation with covid statistics and police presence in the city. I will also output a predictive model that will take in a borough, positive test cases and see if there is a relationship between that and crime rates. I will use error functions such as RMSE and MSE.

1. Create line graph mapping positive covid results through time for each borough
2. Create a choropleth to represent the same information better
3. Create a line graph mapping police brutality and covid
4. Create a line graph mapping police arrests and covid
5. Create a line graph mapping borough, crime type, crime rate, and covid
6. Create an algorithm that calculates “police presence” that takes into account police brutality, arrests, police reports etc. that will act as a weight for each data and outputs a singular metric

7. Create a predictive model that establishes a correlation between covid, a borough and the police presence factor (linear regression or neural network)
8. Evaluate the model using RMSE, MSE, Confusion Metric, and log loss

Data:
https://data.cityofnewyork.us/Public-Safety/NYPD-Arrest-Data-Year-to-Date-/uip8-fykc
https://www1.nyc.gov/site/nypd/stats/reports-analysis/use-of-force-data.page
https://data.cityofnewyork.us/Public-Safety/NYPD-Hate-Crimes/bqiq-cu78
https://data.cityofnewyork.us/Health/DOHMH-COVID-19-Antibody-by-Borough/x98t-3bbk/data
https://data.cityofnewyork.us/Public-Safety/NYPD-Shooting-Incident-Data-Year-To-Date-/5ucz-vwe8

Security and Privacy Considerations: I will not be working with PII data however mis-use of this data can have some serious implications. To mitigate incorrect interpretation of data I've included multiple visualizations to allow the reader to completely understand the data being predicted along with written explanations for each graph and result.