How Did the COVID-19 Pandemic Affect Crime In Chicago? A Data Analysis Approach to Crime Data Correlations

Azal Alhadidi, Rekha Srini Bhupatiraju, Holland Hysmith

Introduction

The COVID-19 pandemic called for global shutdown in the year 2020, heightening a sense of isolation, fear, and anger among billions of people. With the ongoing friction of shutting down businesses and mask mandates, less people were able to carry out their normal lives. Interestingly enough, crime decreased at a rate of 37% worldwide after stay-at-home issues were ordered [**1**]. According to the American Law Institute in 1962, crime has been defined as an "offense by the Model Penal Code or any other statute of this State, for which a sentence of imprisonment is authorized" [**2**]. However, cities like Chicago that accounted for half of the U.S. homicides for 2016, create speculation if the pandemic had an effect on their crime rate as well [**3**]. Understanding the connections of densely populated cities with crime behavior is imperative to help identify potential motives and prevent further crime from occuring in the future.

As machine learning models and data analysis methods mature in the computational landscape, applications in criminal studies have utilized their techniques. A recent study was conducted to analyze crime in South Africa following a linear regression model. This model was designed to help predict crime across all nine provinces dependent on the population, number of police stations, and types of crimes committed [**Figure 1**]. A separate study observed 2013 crime statistics in the state of Mississippi by utilizing linear regression, additive regression, and decision stump algorithms. The goal of this work was to categorize which cities of Mississippi were more likely to have certain crimes occur and how many crimes happened based upon population [**Figure 2**]. Therefore, for this work, an array of data analysis techniques will be constructed to correlate crime trends in Chicago during the pandemic. Understanding connections such as locations where crimes were committed, types of crimes, and times in which they occurred can be a helpful resource for understanding how global phenomenon can impact large cities. This information can be used for future events to help mitigate prevention of crime occuring in the city of Chicago.

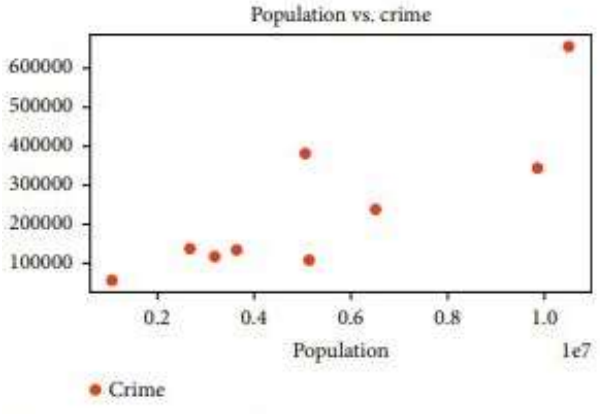

Figure 1: Linear Regression approach to crime based on population [**4**].

Algorithm	*Correlation Coefficient*	*Mean Absolute Error*	*Root Mean Squared Error*	*Relative Absolute Error*	*Root Relative Squared Error*
Linear Regression Model	1	0.004	0.006	0.0009%	0.001%
Additive Regression Model	0.97	116	168	26%	27%
Decision Stump Model	0.78	276	379	62%	62%

Figure 2: Results for Violent Crimes per 100k of Population [**5**].

Data Description

For this project, the analytical study leverages the Chicago crime dataset. Data is obtained from the Chicago Police Department's CLEAR (Citizen Law Enforcement Analysis and Reporting) system. This dataset is about the reported incidents of different crime types (except murdering) across Chicago state from 2001 to the present, with weekly updates. The dataset records may be updated by the authority based on further crimes' investigation further

on. This dataset can reflect enormous trends due to impacting factors, including the Covid19 pandemic. Hence, this report will focus on exploring the Chicago crime data during the pandemic and the effect of some applied precautions by the authority (e.g., lockdowns and isolation) on the crime rate and type.

The dataset consists of several necessary fields related to the reported crimes, including reporting code, type of the crime, demographic location, crime description, and other crime and police-related fields. The dataset contains many crime features (about 30 features) and a huge number of records (more than 6 million) as it has accumulated records since 2001. This would make the analytical process difficult and time and computational resources-consuming. Since we intend to explore the effect of Covid19 on the Chicago crime incidents, and since the pandemic spread in the US occurs around December 2019, we will utilize part of the dataset related to the crime dataset of the last two-three years. Therefore, the primary dataset will be diminished by writing a script for shortening it and putting the trimmed data in a new shape to serve our analytical needs

The dataset will be analyzed and preprocessed to figure out the exciting patterns using Python Jupyter Notebook with Pandas, Numpy libraries. The Matplotlib library is also utilized for presenting the analytical results. The Chicago crime dataset is available in CSV format on the Kaggle website[**6**].It is freely available for everybody to use under the following terms provided by the Dataset Source[**7**].

Methods

Our work would primarily be a data analysis to generate some descriptive statistics, rather than any prediction problem, because we are interested in quantifying the effect of recent events. Below are some questions that we will consider, though we may discover more as we explore the data.

1. Which categories of crimes had their rate affected by the onset of the pandemic, and how?
2. How did the location of crimes change during the pandemic, as more people spent more time at home?
3. How did the rate and proportion of domestic incidents change during the pandemic?
4. Was the proportion of arrests to crimes affected? Did it change only for certain categories or locations?

If time permits, we could consider geographical location as a factor as well. We will have to do some preprocessing. We will extract the data for our chosen date range (January 2019 - present) and extract categories of data. We may need to combine some categories if the data is too fragmented to show meaningful patterns.

To generate quantitative measures, we can run linear regression on rates of crimes overall and by category. We can also use similar methods to analyze the arrest rates and compare the change in locations, to see if there's a proportional increase in crime within residences. We'd also generate a number of visualizations to illustrate the temporal differences - we can do this with line graphs that include our regression results, as well as comparing pie charts at certain points in time to see how proportional data (such as crime location) changes over time. Success would be the observation of a statistically significant change in the crime and arrest rates we choose to analyze. We hypothesize that we will see a decrease in all types of crime and arrests, but seeing some types increase would also be a meaningful result and could be considered a successful result.

Conclusion

This study will provide a quantitative view of the effect of the pandemic's impact on crime rates and response. If we find no statistically significant patterns, this could be indicative of a lack of an effect that could still tell us something about the population's behavior. Any result could be informative for governments and other organizations as it could inform their response to future events.

References

1. Boman, J.H., Mowen, T.J. (2021) Global crime trends during COVID-19. Nat Hum Behav 5, 821–822. https://doi.org/10.1038/s41562-021-01151-3
2. Lauritsen, J. L., & Cork, D. L. (2017). Expanding Our Understanding of Crime: The National Academies Report on the Future of Crime Statistics and Measurement. Criminology and Public Policy, 16(4), 1075–1098.
3. Sanburn, Josh (2016) Chicago Is Responsible for Almost Half of the Increase in U.S. Homicides. TIME Magazine.
4. Obagbuwa, I.C., & Abidoye, A.P. (2021) South Africa Crime Visualization, Trends Analysis, and Prediction Using Machine Learning Linear Regression Technique. Hindawi Applied Computational Intelligence and Soft Computing. https://doi.org/10.1155/2021/5537902
5. McClendon, L., & Meghanathan, N. (2015). Using machine learning algorithms to analyze crime data. Machine Learning and Applications: An International Journal (MLAIJ), 2(1), 1-12.
6. https://www.kaggle.com/chicago/chicago-crime
7. https://data.cityofchicago.org/