Chicago Crime Analysis

Group 5: A Data-Driven Approach to Understanding Crime patterns

Steven Delaney
Student
University of Colorado Boulder
steven.delaney@colorado.edu

William Cole Akers
Student
University of Colorado Boulder
wiak0179@colorado.edu

Problem Statement and Motivation

With this project we intend to analyze crime trends in Chicago using historical data ranging from the years 2001 to 2023. We aim to conduct in depth analysis of the crime trends present in this data set. Using statistical techniques we aim to understand how the patterns in crime have changed and further to infer future changes so that law enforcement could use the findings to help mitigate future crimes and plan for changing crime patterns in the years to come. Using the knowledge gained from the analysis we hope to uncover and present trends that show patterns across different locations and time periods.

The first question we aim to answer is to see what areas in Chicago are “hot spots” for violent crimes. We will analyze which areas experience the highest rates of crime as well as how these hotspots have changed over time. We can apply this knowledge to help law enforcement better understand what areas they need to concentrate on more heavily. This could be used to help create policing routes and patrolling areas putting an emphasis on areas that have been identified as hot spots.

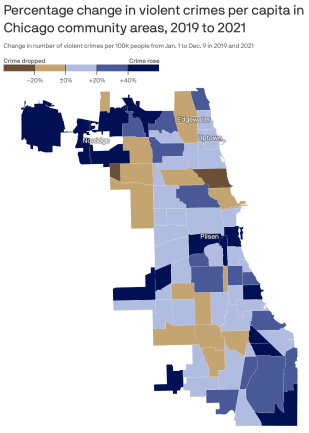

Figure 1: change in violent crime per capital 2019 to 2021

https://www.axios.com/local/chicago/2021/12/20/chicago-violent-

We also intend to answer the question is there seasonality to crime. That is to say we want to be able to tell if crime rates are increasing or decreasing during certain months and do these changes persist at the same season throughout multiple years. When patterns are identified law enforcement will know when to expect more or less crime activity. This can help with preparation of officers on the individual level allowing them to know to expect more or less crime. It can also help at an administrative level where the number of patrols could be increased or decreased to match crime rates of the current season.

Further, we intend to investigate which districts have the highest and lowest success rates for arrests. We can find which districts and wards have the most success in finding criminals and detaining them. This examination will allow us to see if there are noticeable discrepancies between districts. By identifying districts that do not have a high success rate for arrest administration can allocate more resources to these districts to help bring their arrest rates up to the rates exhibited in other districts.

We also aim to answer wether or not there is a correlation between location and type of crime. Using the data we can find out which crimes are most prevalent in certain neighborhoods or under specific environmental conditions. Armed with this knowledge officers patrolling certain locations can be better outfitted to recognize and mitigate specific criminal activities that are common in areas they patrol.

Lastly, we aim to identify trends in types of crimes that occur over the years. By sifting through the type of crime and dates of crimes we can look for trends in specific types of crime and identify which crimes were committed most in certain years as well as how those trends changed through the years. We can use this knowledge to potentially predict the changing trends in types of crime In future years.

Literature Survey

We have identified three previous studies that performed research on crime trends, law enforcement strategies, and data-driven crime prevention.

The University of Chicago Crime Lab studies conduct research on Chicago crime trends with a focus on policy interventions and the impact of law enforcement strategies on crime rates. These studies highlight evidence based intervention and the resulting reductions in violent crime. Additionally they evaluate programs like predictive policing and youth crime intervention. The University of Chicago Crime Lab studies give us an inside to historical crime trends and which intervention strategies worked well to mitigate crime rates.
URL**:** https://crimelab.uchicago.edu/resources/2024-end-of-year-analysis-chicago-crime-trends/

The Chicago Police Department (CPD) crime Statistics study focuses on annual crime statistics with an official record of reported crimes, arrests corresponding to these crimes, and success rates of law enforcement. These reports examine this data differentiating by district and category of crime. The key findings include statistics which show patterns in crime occurrence, arrest rates, and law enforcement responses.
URL**:** https://www.chicagopolice.org/statistics-data/crime-statistics/

The third study from the Journal of Data Analysis and information Processing focuses on long-term crime trends in Chicago. These trends are examined using machine learning techniques and statistical monitoring. This study revealed that crime rates fluctuate depending on socio economic conditions, seasonal trends, and neighboring infrastructure. We can use the information found by this study to help us check our findings for seasonal trends and as a model to check which statistics from our data set are needed to conduct this type of evaluation.

URL: https://www.scirp.org/journal/paperinformation?paperid=134329

Figure 2: Chicago skyline

Google maps image

Proposed Work

Data preprocessing and cleaning is the first task for this project. First we intend to detect outliers and remove any that may skew results of our analysis. Using statistical techniques like z-score analysis, interquartile range, and clustering to identify data points that could cause skewing.

Using cluster analysis we can group crime prone areas based on frequency and type of crime. After graphing and applying a clustering algorithm we will be able to evaluate clusters and see high density regions that may not be specific to locations defined as one of the attributes or if there are any clusters when plotted by a certain trait to see if there are any interesting correlations that we are looking for or that we did not expect.

Additionally we will check for any missing values and where possible fill in with either mean median or mode or handle in a case by case manner if very few missing values are found for either data or categories filling in missing values when possible. Filling in categories as unknown when no other reasonable solution is achievable.

We will then use tools to display the data and look for correlations that will help us answer the questions we presented in the problem statement section. Some questions we are attempting to answer like looking at the crime trends over the years, success rates of law enforcement, and seasonal trends replicate the studies we observed and summarized in the literature survey section. However, our project adds several unique observations that could be of use to law enforcement. These include looking for a correlation between location and type of crime as well as identifying hot spots for crime.

Dataset

The primary dataset for this project is sourced from Kaggle and was originally collected by the Chicago Police Department as part of the CLEAR (Citizen Law Enforcement Analysis and Reporting) System. The dataset spans from 2001 to the present, containing over 7 million reported crimes with the following attributes; ID, Case Number, Date, Area (location), IUCR, Type of crime, Description, Location (street, residence, or other), Arrest status, Domestic.

Dataset URL: Chicago Crimes Dataset –

https://www.kaggle.com/datasets/utkarshx27/crimes-2001-to-present

Evaluation Methods

We intend to evaluate our findings by cross referencing prior research done and comparing our findings to them. By using the existing crime research like the methods mentioned in the literature survey section we can validate our findings against existing solutions. We will look for correlations that answer our research questions with the hope that we find expected correlations or definitive analysis that disproves correlations we were searching for.

We further intend to create a predictive model that we can compare to other proven and existing prediction models. This model will help predict location based on the time the crime is committed and what type of crime was committed.

By validating against existing solutions we hope to glean how effective our model and data handling was. We believe that this method will be effective and accurate as a way to judge our performance with the proposed work.

Tools

Visualization: Tableau.

Analysis: Pandas, NumPy, Matplotlib.

Programming Language: Python.

Development Environment (IDE): VSCode, Jupyter Notebook.

Collaboration: GitHub, Outlook, Discord.

Milestones & Timeline

Milestone	Completion Date
Data Cleaning & Preprocessing	Mar 3
Initial Exploratory Analysis	Mar 10
Crime Hotspot & Time-Series Analysis	Mar 17
Arrest Success & Correlation Studies	Mar 24
Model Development	Apr 7
Final Report & Visualization	Apr 28