Chicago Crime Analysis

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

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

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

Abstract

In this project, we analyze crime trends in Chicago from 2001 to 2023 using the Chicago Crimes Dataset. Our goals were many fold in that we wanted to explore the data to see if crime is predicatble. To this end we dug into the data to uncover patterns that could explain how crime changes across time, location, and socio-economic factors. We used various techniques to clean and explore the data including clustering, time-series analysis, and predictive modeling to explore these crime trends. Additionally, we geocoded the data and explored how factors like household income interact with both the amount of crime and arrest rate.

Our findings revealed several key insights. First, we observed a steady decrease in reported crimes year-over-year, with an average decline of about 6% annually. Second, we identified strong seasonal patterns, with crime rates consistently spiking during the summer months. Using K-Means clustering, we were able to detect consistent geographic crime hotspots across the city. Our geocoding efforts further showed a correlation between lower household incomes and higher crime rates, suggesting that economic conditions play a significant role in crime distribution.

Initially, we aimed to predict the specific locations of future crimes. However, as the project progressed, we pivoted toward forecasting the number of crimes expected each month instead. This approach proved simpler to model accurately given the available data and, importantly, offers more practical value to law enforcement for resource planning and patrol scheduling as well as providing a data driven resource to lobby for increased funds. Our predictive model shows that monthly crime levels can be forecasted with reasonable accuracy based on historical patterns. Overall, our work demonstrates how data-driven methods can uncover meaningful trends and provide tools that support crime prevention and operational planning efforts.

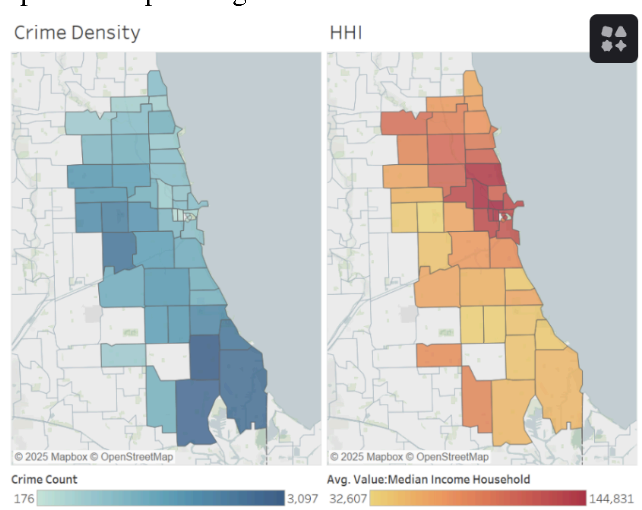

Figure 1:Crime density vs Household income produced in Tableau

Deliverables:

- A machine learning model that forecasts the number of crimes expected per month based on historical trends.
- Visualizations to better understand crime patterns and seasonal effects.

Applications:

- Helping law enforcement allocate resources more efficiently based on predicted crime surges.
- Supporting predictive policing efforts by incorporating seasonal and economic patterns.
- Providing insights into how socio-economic conditions correlate with crime, helping inform community support and prevention strategies.

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 will conduct an in-depth analysis of the crime trends present in this data set. Using statistical techniques, we seek 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.

Firstly, we will identify what areas in Chicago are "hot spots" for violent crimes. We will analyze which areas experience the highest rates of crime and what types of crimes are most common there. Additionally, we will explore how the crime landscape has changed ward to ward. We can apply this knowledge to help law enforcement better understand what areas they need to concentrate on more heavily in the present day. This could be used to help create policing routes and patrolling areas, emphasizing areas that have been identified as hot spots.

We also intend to answer the question, "Is crime impacted by seasonality?" That is to say, we want to be able to tell if crime rates increase or decrease during certain months and do these changes persist in the same season year to year. When patterns are identified, law enforcement will know when to expect more or less crime activity. This can help with the 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 the crime rates of the current season.

As one measure of performance, we will investigate which districts and wards have the highest and lowest arrest rates for crimes that warrant arrest. 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, we can compare differences in policies and practices between lower-performing districts and wards and higher-performing ones. This analysis may also bring to light if performance is a factor of policing quality or if a difference in resources plays a bigger role.

A third area of analysis is whether 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 the areas they patrol. Lastly, we will conduct a temporal analysis based on time of day. For each case the data provides a time/date stamp that is space-delimited and can be easily parsed. Time will be used to illustrate what crimes are most prevalent in a 24-hour period

Introduction

For this project, we wanted to dig into Chicago crime data from 2001 to 2023 and try to answer a few important questions. First, we wanted to see how crime rates have changed over time — is there a steady pattern, or are there big swings in certain years? We also set out to figure out if some areas of Chicago consistently have more crime than others, and if certain neighborhoods could be labeled as hotspots.

Another big question we had was about seasonality. We were curious if crime rates go up or down depending on the time of year, like if there's a spike during the summer months. On top of that, we wanted to see if income levels or property values had any connection to the amount of crime or the success rate of arrests in different areas. Finally, we wanted to see if it would be possible to build a model that could predict crime trends, especially the number of crimes expected each month, based on patterns from the past.

These questions are important because they can help law enforcement and city planners better understand how crime moves and changes over time. If police departments know when and where to expect higher crime, they can make smarter decisions about where to put officers and how to plan patrols. Understanding how income levels tie into crime rates could also help point to deeper issues that cities might need to address outside of just policing. And if crime can be predicted even a little bit, it gives agencies more tools to stay ahead of the curve instead of just reacting after the fact.

Related work

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 look at 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 that 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 socioeconomic 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 the results of our analysis. Using statistical techniques like z-score analysis, interquartile range, and clustering to identify data points that could cause skewing

Additionally,we will check for any missing values and, where possible, fill in with either mean, median, or mode. If no replacement can be found for the value, we will handle it case-by-case 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, and 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 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 plan to develop a predictive crime model and compare it against existing crime prediction models to evaluate its accuracy and effectiveness. Our model will use crime type and time of occurrence to predict where crimes are most likely to happen, leveraging historical data to identify patterns and trends. This prediction model will focus on tying locations to crime types and descriptions. It is our aim to create a model that might aid in narrowing in on the location of a crime when a vague description of the area of the crime is given.

To validate our approach, we'll test our model against proven methods like logistic regression, decision trees, and neural networks, assessing its performance based on key metrics like precision, recall, and accuracy. This comparison will help us gauge how well our data handling and methodology hold up against established standards.

By refining our model through this process, we aim to create a reliable tool for crime forecasting, one that can support law enforcement, resource allocation, and crime prevention strategies with data-driven insights. Specifically, we aim to help narrow down the location where a crime may have occurred based on the type of crime committed.

Tools

Visualization: Tableau.

Analysis: Pandas, NumPy, Matplotlib, Alteryx

Programming Language: Python.

Development Environment (IDE):
VSCode, Jupyter Notebook.

Collaboration: GitHub, Outlook, Discord.

Milestones & Timeline

Data Cleaning & Preprocessing	Mar (15-16)
Initial Exploratory Analysis	Mar (17-21)
Crime Hotspot & Time-Series Analysis	Mar (24-28)
Arrest Success & Correlation Studies	Mar (24-28)
Model Development	Mar (31-Apr 11)
Final Report & Visualization	Apr (21-28)

Milestones Completed

Targets:

- Data cleaned
    - Address dropped as it was partially censored data, making it unusable
    - Crimes that do not warrant arrest will be separated into their own category of analysis.
    - Rows with null latitude and longitude
- Initial exploratory analysis
    - We have derived a set of initial summary statistics that will lead to further exploration.

To clean the data, we dropped the address column as we found it to be extraneous information. This is because the data set offers longitude and latitude as well as a precinct that are much more useful for plotting on a geographical map. Additionally, we removed from our data sets crimes that do not warrant arrest, and instead, the repercussions are a citation or non-arrest-worthy penalty. Our original approach to find crimes where the punishments do not warrant arrest was to look for primary types that always resulted in a false arrest rate. However, we discovered that no such primary type existed.

We discovered this is caused by the different scales of crime types. An example of scaling that we encountered is theft. A theft where the stolen item is valued at $100 would not be a crime where arrest is warranted, but when the $ value is increased, the crime warrants an arrest at a certain point of the upscaling. We were able to find crimes that are not arrest-worthy by instead filtering through the descriptions of crimes and then dropping crimes with descriptions that never lead to arrest. We did this to get an accurate look at arrest rates by precinct where crimes that do not warrant arrest may be more common and lead to lower arrest rates.

Milestones To-Do

Still to do:

- Crime hotspot time-series analysis
- Arrest success and correlation studies
- Acquire additional dataset on Chicago property values
- Develop a predictive model
- Final Report & Visualization

We plan to use crime hotspot time series analysis to analyze the shift of crime hotspots over the years our data set takes place. Using arrest success and correlation studies, we will compare the success rates of different wards and precincts by using the total number of crimes that warrant arrest and the number of successful arrests in these predefined divisions. Our final steps will be running a machine learning model through our data, focusing on arrest location and the success of arrest rate. It is our goal to create a model that can project future changes in success rate by ward/precinct.

Results So Far

In our initial exploration, we have derived several summary statistics that will guide further exploration and experimentation.

One major finding is the consistent drop in the number of cases year over year, illustrated in Figure 1 below.

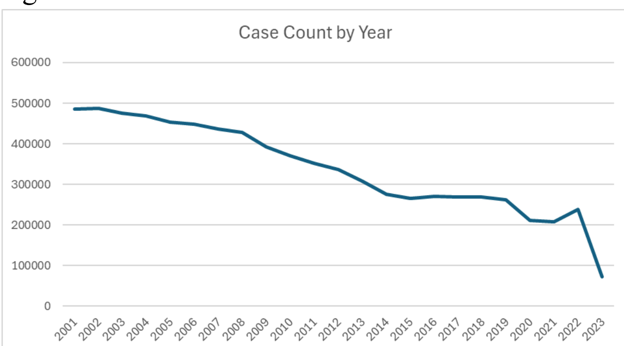

Figure 3 case count trend by year

The number of cases has dropped at an average rate of 6% each year compared to the previous year. This begs the question, why? Is this a result of better policing or changes in how crimes are reported?

Additionally, we have begun to conduct a geospatial analysis utilizing Tableau. Specifically, crime density by zip code. This allows us to easily identify areas with the highest and lowest volume of reported crimes. We aimed to take a more granular approach to our refinement of the data by separating the geographical space into more areas building on what previous research projects have already done. See Figure 2 below

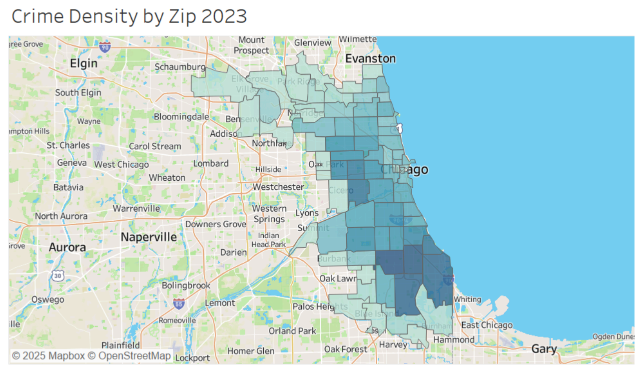

Figure 4 crime desity map

Figure 2 clearly illustrates that in 2023, South Central Chicago was a hot spot for crime, while the Upper West Side had a much lower volume. This opens several additional avenues we can go down. One next step would be to run the same analysis for multiple years to see if there is a geographical change in density over time or if hot-spots have remained consistent for several years. We will also explore potential reasons why the identified areas might have higher or lower crime rates. The factors we discover will play a major role in our predictive modeling.

Responses to Feedback

What method of cluster analysis will be used?

We used K-Means clustering for our analysis, specifically applying it to latitude and longitude data to identify geographical outliers in reported crimes. This method helps us group crimes based on relative location, allowing us to see patterns that might not be obvious at a glance.

We've organized the crimes into five distinct clusters, giving us a view of crime distribution that is purely data-driven across the city. By analyzing these clusters, we can pinpoint high-crime areas, detect anomalies, and compare crime density across different regions. This approach helps us understand how crime is spatially distributed and whether certain areas exhibit unique trends or outlier behavior.

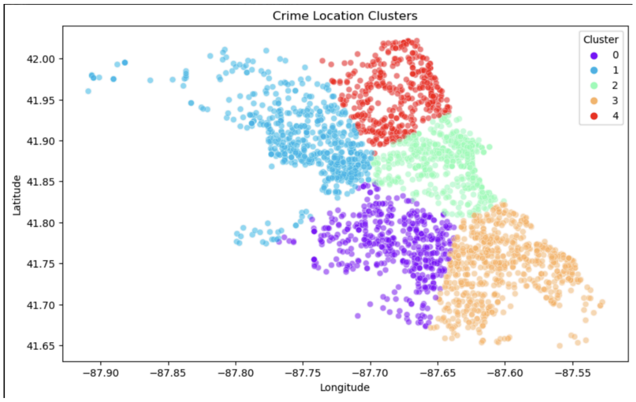

Figure 5: K-means clustering

Through cluster analysis, we grouped crime-prone areas into five distinct clusters based on the longitude and latitude of where the crime occurred. This allows us to analyze crime geographically from an overarching viewpoint. This allows us to double-check our findings for high-density crime areas that are larger in scale and do not conform to pre-defined borders that are present in the dataset.

Each of the five clusters represents a grouping of crimes that occur in geographically similar areas as determined by k-means. Using these larger groups to look at arrest percentages will help us discover if there are larger trends that exceed precinct/ward or zip code limits.

Moving forward, we plan to analyze each cluster to determine common characteristics, evaluate potential influencing factors, and explore how crime shifts within and between clusters over time.

How does this build on previous work, and what differentiates it from these studies?

Our work will most closely mirror that of *"The Windy City's Dark Side: A Statistical Exploration of Crime in the City of Chicago"* (Odooh et al., 2024.) Not only was the same data set used in this research, but many of the same questions were asked and explored. Particularly correlations in location and crime type, the impact of seasonality on crime, and crime density based on location. We will differentiate our work by taking economic factors into account, namely correlations between property values and how this impacts arrest rates and crime density. Economic impact is not a factor considered in the previously referenced work.

Additionally, we also consider how departments have analyzed crime throughout the years and what they consider to be their measure of success. Tableau allows us a unique opportunity to easily create geographic maps that these studies do not include in their analysis. If they do have the map, Tableau allows us to add more distinct borders to examine the data with a finer-toothed comb. This allows us more granularity than previous studies, which we hope will lead to better accuracy when pinpointing success rates. The more granular approach is inspired by the suggestion for a continuation in the future research directions section of the Journal of Data Analysis and Information Processing study mentioned in the literature review section of this proposal.

Our graphs help visualize the changes in time and arrest rates in a more readily apparent way for consumers of the report to get a more complete understanding of the studies in a much shorter time frame. We also include a more expansive scope of crimes committed, whereas the studies before focused on violent crimes. This differentiation leads to different hypotheses and conclusions for example, the previous studies focused on the lethality of shootings. Previous studies also picked the years they would include with an aspect of randomness. We plan to analyze in three-year chunks, which will give us a more complete look at the changes in the attributes we are observing.

Finally, our most clear-cut distinction from other studies is the development of a machine learning model that will help to predict the location of a crime based on the type of crime reported. While previous studies have graphically illustrated the data, none of our references have attempted to use the data to make a predictive model. This model could be applied by law enforcement to increase response times to crimes by helping to predict the area of the crime when an emergency line caller gives vague or nondescript locations. In terms of building off of previous studies and differentiating our work from what has been done previously, this model sets our exploration apart from the work covered in our literature review.

Main Techniques

Data Cleaning:

Before beginning our analysis, we first needed to clean and prepare the dataset to ensure more accurate results. One of the first steps we took was identifying and removing outliers that could skew our findings. We applied both z-score analysis and interquartile range (IQR) methods to detect extreme values and either remove them or handle them appropriately.

To address missing data, we used simple imputation techniques. Depending on the type of data, missing values were filled with either the mean, median, or mode to maintain consistency without introducing significant bias.

During our exploration of the dataset, we also made a few specific decisions based on data quality issues we observed. The address field, for example, was partially censored and therefore unreliable for location-based analysis. We chose to drop this column and instead rely on the latitude and longitude attributes, which provided cleaner and more precise location information for mapping and clustering tasks.

Finally, we removed any rows that were missing latitude or longitude values, as accurate geographic data was critical to several parts of our project.

By thoroughly cleaning the data, we ensured a more reliable foundation for the rest of our analysis, visualizations, and modeling efforts.

Clustering:

To better understand where crime was most concentrated, we conducted a hotspot analysis focusing on several of the most common crime types, including theft, battery, assault, narcotics, and robbery. We filtered the data by primary crime type and then rounded the latitude and longitude coordinates to two decimal places to group nearby incidents together into geographic bins.

Using these rounded coordinates, we generated heatmaps to visualize the density of crimes across different parts of the city. The heatmaps revealed distinct areas where specific types of crimes were especially common. By creating separate heatmaps for different crime types, we were able to see that different kinds of crimes tend to cluster in different parts of Chicago.

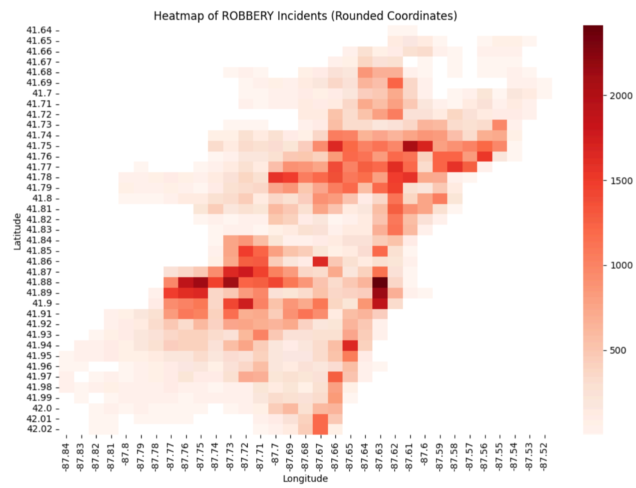

Figure 6: heatmap of robbery by long/lat

To refine our hotspot identification, we applied K-Means clustering to the location data, grouping crimes into five geographic clusters. This clustering allowed us to move beyond simple zip codes or precinct boundaries and instead identify natural groupings based purely on where crimes were occurring. We also analyzed how crime counts within each cluster changed over time by plotting yearly crime trends for each cluster from 2001 to 2022.

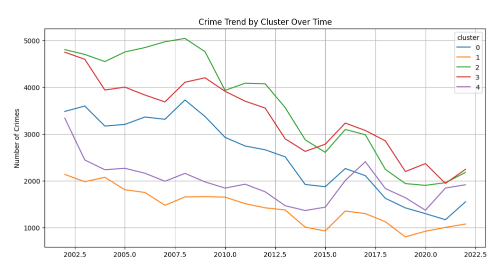

Figure 7: crime trend by 5 K-means clusters

For additional visualization, we created an interactive map using Folium, plotting a random sample of

100,000 crime incidents colored by cluster. This map made it easy to see where crimes were most heavily concentrated and how the clusters were distributed across the city. (A link to a video demo for this map is in our visualization section.)

Finally, we grouped crimes by cluster and identified the top five crime types within each cluster. This analysis helped us understand not just where crime was happening, but also what kinds of crimes were most common in each hotspot area. This result was almost always robbery, as the most common crime primary type.

	cluster	Primary Type	count
0	0	ROBBERY	53310
1	1	ROBBERY	29983
2	2	ROBBERY	75912
3	3	ROBBERY	70534
4	4	ROBBERY	41378

Temporal Analysis:

To investigate how crime patterns change over time, we conducted a detailed temporal analysis across multiple dimensions, including seasons, time of day, and long-term monthly trends.

First, we categorized each crime into one of four seasons (Winter, Spring, Summer, or Fall) based on the month it occurred. We then grouped the data by year and season to examine how crime counts shifted over time. This analysis revealed strong seasonal trends, with crime rates consistently peaking during the summer months across most years.

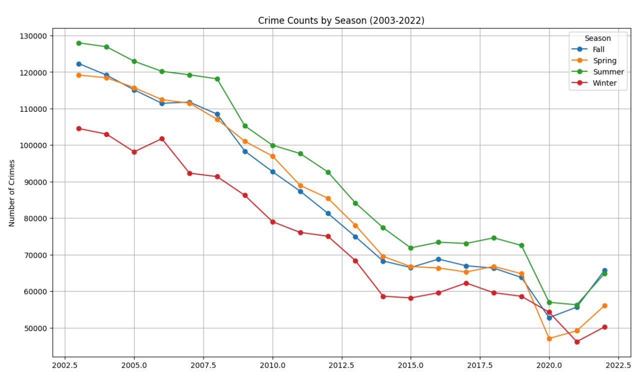

Figure 8: Crime counts by season

Next, we looked at the time of day that crimes were occurring. We extracted the hour from each crime's timestamp and plotted crime counts by hour for each season. This revealed that crime activity tends to rise sharply during late afternoon and evening hours, with some variation depending on the season. For example, summer months showed a slightly later peak in crime compared to winter months.

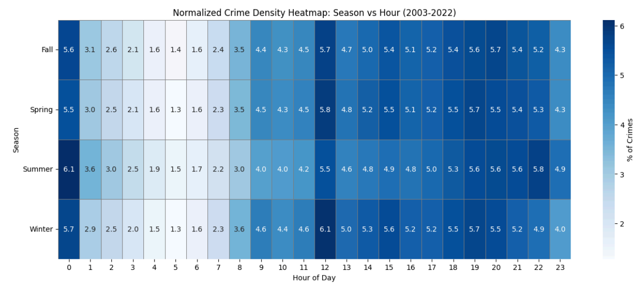

Figure 9: normalized Crime density Heatmap

We also extended this analysis by examining crime patterns at the ward level. For each ward and season, we identified the specific hour of the day when crime was most likely to occur. This produced a clear view of how peak crime times shift across different areas of the city depending on the time of year. We visualized these patterns using both heatmaps and bar plots, making it easy to see the differences across wards and seasons.

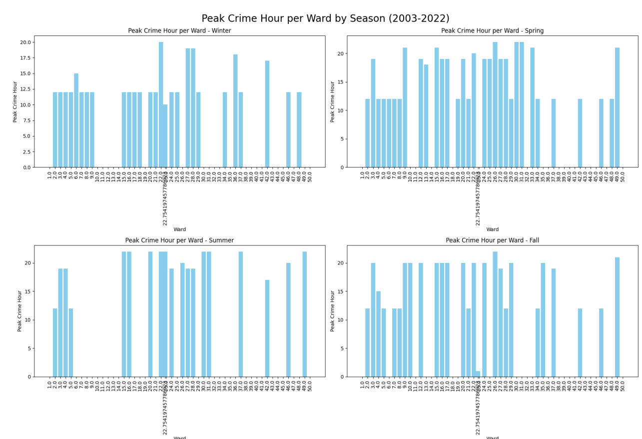

Figure 10: Peak crime hour per ward by season

Finally, we performed a seasonal decomposition of monthly crime counts using an additive model. This allowed us to break down the time series into its underlying trend, seasonal component, and residual noise. The decomposition confirmed a strong yearly seasonal pattern, with crime counts rising and falling in a predictable cycle each year.

Overall, the temporal analysis provided important insights into when crimes are most likely to occur, both at a city-wide and a ward-specific level. These findings can help inform resource planning, patrol scheduling, and broader crime prevention strategies throughout the year.

Predictive Modeling:

To forecast future crime counts at the ward level, we built a Random Forest regression model using historical crime data from 2003 to 2022. The model predicted the monthly count of crimes per ward, using features such as Month, Year, Ward, Season, Daypart, and crime lags (crime count one month prior and a three-month moving average).

We evaluated the model using Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE), and R^2. The model achieved strong performance, with R^2 scores around 0.96–0.97 and an average prediction accuracy of approximately 93% across multiple years.

Key Results

Crime Trend:

One of the most noticeable findings from our analysis was a steady decline in reported crime cases across Chicago from 2001 to 2023. On average, crime counts decreased by approximately 6% each year compared to the previous year. This long-term downward trend suggests that a combination of factors may be at play. Possible explanations include improvements in policing strategies, greater use of surveillance and crime prevention technologies, expanded community outreach programs, and shifts in how crimes are reported and categorized over time. While further study would be needed to pinpoint the exact causes, the overall trend points to a significant reduction in reported crime citywide over the two-decade period.

Crime Hotspots:

By grouping crime incidents by official police districts, we found that crime in Chicago is heavily concentrated in specific areas. Districts 8, 12, 6, 4, and 10 consistently recorded the highest number of reported crimes throughout the time period we analyzed. These districts are primarily located in the South and West sides of the city, which have historically been known for higher crime rates. Identifying these high-crime districts highlights important areas where law enforcement resources and prevention efforts could be most effectively focused.

Seasonality:

Our analysis of crime counts by season revealed clear seasonal trends in Chicago from 2003 to 2022. Crime rates consistently peaked during the summer months across almost every year, while winter months showed the lowest levels of crime. This seasonal pattern remained steady even as overall crime levels declined over the two-decade period. Seasonal decomposition of monthly crime data further confirmed a strong recurring seasonal cycle, highlighting how crime tends to rise and fall at predictable points each year.

Predictive model:

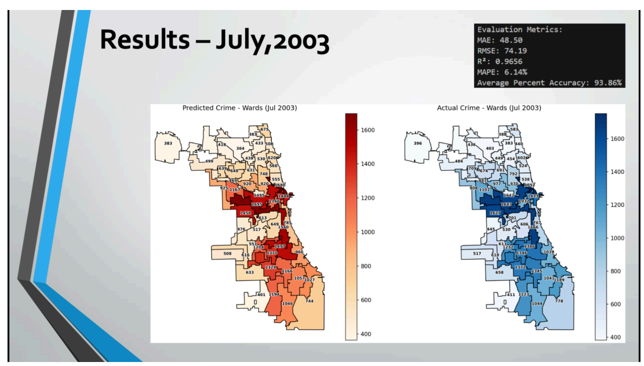

Figure 11: Model results

To visualize results, we mapped predicted crime counts against actual crime counts for individual months, helping us confirm the model captured major trends across the city. Overall, the model showed that monthly crime totals can be forecasted with a high degree of accuracy, offering a tool that could support better resource planning and policing strategies.

Applications

The results of our analysis provide several important ways that law enforcement agencies and city planners could apply data-driven strategies to improve crime prevention, resource allocation, and community safety.

By identifying crime hotspots through detailed location analysis, police departments can target their patrols and allocate officers more effectively. Rather than relying solely on traditional district or beat structures, departments can use crime density data to adjust coverage in a flexible, needs-based way. This approach ensures that areas with consistently high crime rates receive appropriate attention.

Our seasonal analysis also revealed clear patterns, with crime rates reliably peaking during the summer months. Agencies can use this insight to plan ahead, adjusting staffing levels and patrol strategies during times of higher expected criminal activity. Proactively increasing resources during high-risk seasons can help departments stay ahead of crime trends rather than responding after incidents occur.

In addition, the predictive model we developed, which forecasts monthly crime counts by ward, offers another tool for operational planning. Even in its early stages, the model shows that it is possible to use historical and seasonal patterns to anticipate future crime levels with strong accuracy. With further refinement, this type of predictive modeling could become an integral part of planning patrol routes, scheduling officers, and allocating investigative resources.

Importantly, the findings from this project also provide a data-driven foundation that districts can use when applying for additional funding or grants. Demonstrating clear, statistically-backed needs based on crime patterns and seasonal surges can strengthen proposals and ensure that limited resources are distributed to areas where they are needed most.

Finally, by understanding how specific types of crimes correlate with different neighborhoods and socio-economic factors, agencies and policymakers can design more targeted intervention programs that go beyond traditional policing, addressing the root causes of crime where possible.

Overall, our analysis highlights how data can be leveraged to make more informed, proactive decisions at every level of law enforcement and city management.

Evaluation Methods

We evaluated the performance of our predictive model using several standard metrics. Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) measured the average differences between predicted and actual crime counts. The R^2 score, at 0.9656, indicated that the model explained a large portion of the variance in the data. Mean Absolute Percentage Error (MAPE) was low at 6.14%, and the model achieved an overall average prediction accuracy of approximately 93.86%. These results suggest that the model provides reliable forecasts based on historical crime trends.

```
Evaluation Metrics:
MAE: 48.50
RMSE: 74.19
R²: 0.9656
MAPE: 6.14%
Average Percent Accuracy: 93.86%
```

Figure 12: Evaluation of model

Visualization

Throughout this project, we created a range of visualizations designed to highlight key trends and insights in Chicago crime data.

To analyze geographic patterns, we created crime density heatmaps (**Figure 4**) and applied clustering methods to identify hotspots across the city (**Figure 5** and **Figure 6**). These visualizations made it clear where crime was most concentrated, allowing us to move beyond predefined district boundaries and look at natural crime patterns. We also summarized overall district-level crime totals in **Figure 12**, highlighting which police districts consistently experienced the highest volume of crime.

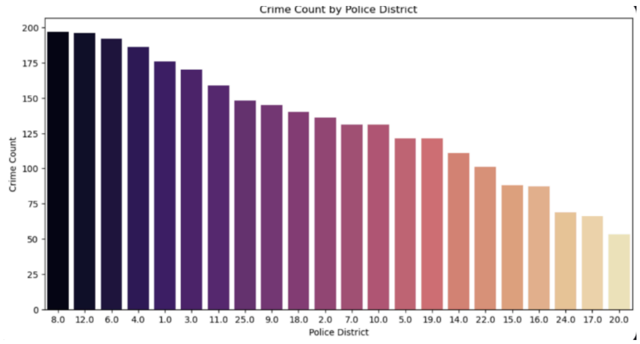

Figure 13: Crime count by police district

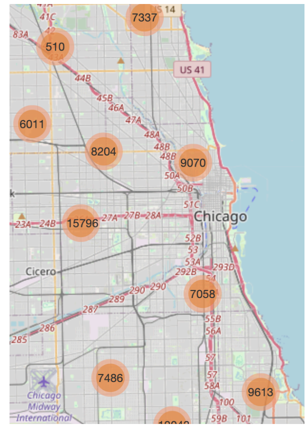

Figure 14: html map of crime hotspots

For temporal analysis, we built seasonal line graphs (**Figure 8**) showing trends across Winter, Spring, Summer, and Fall, and developed a normalized crime density heatmap (**Figure 9**) to visualize how crime activity varied by hour throughout the day. In addition, we created a heatmap showing the peak crime hour for each ward and season (**Figure 10**), offering a highly localized view of when crimes are most likely to occur.

To confirm long-term patterns, we applied seasonal decomposition (STL) to monthly crime data, clearly separating out the trend, seasonality, and residual noise components. This decomposition confirmed that crime exhibits strong and regular seasonal patterns year after year (shown in **Figure 8**).

In addition to static visualizations, we also developed an interactive map using Folium that clusters crime incidents based on their geographic location. Figure 13 shows a screenshot of the interactive crime cluster map generated from a sample of 100,000 crime incidents across Chicago. Each cluster represents a concentration of crimes in a given area, providing a dynamic way to explore crime density at different scales.

Here is a link to a video demonstrating the interactive component:
https://youtu.be/h9Q-A-mcynQ

Overall, the visualizations served not just to display data, but to reveal meaningful patterns and insights that supported our analysis, forecasting, and recommendations for resource planning.