

CHICAGO: EMPOWERING A SAFER COMMUNITY THROUGH DATA-DRIVEN CRIME INSIGHTS



X



Oral Presentation

Presented by Team 51

March 6, 2024

Presentation Outline

01 Problem Statement and Overview

02 Data

03 Modeling

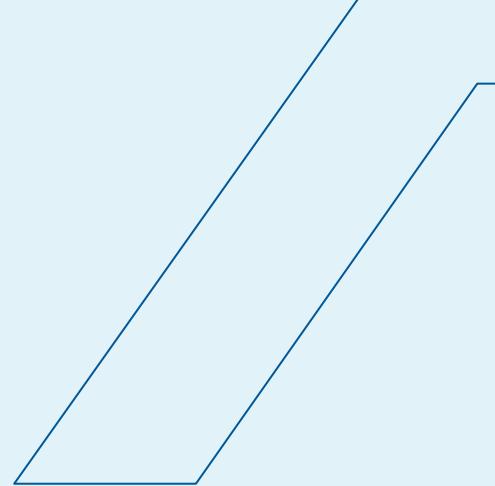
04 Information Access

05 Conclusions

06 Recommendations

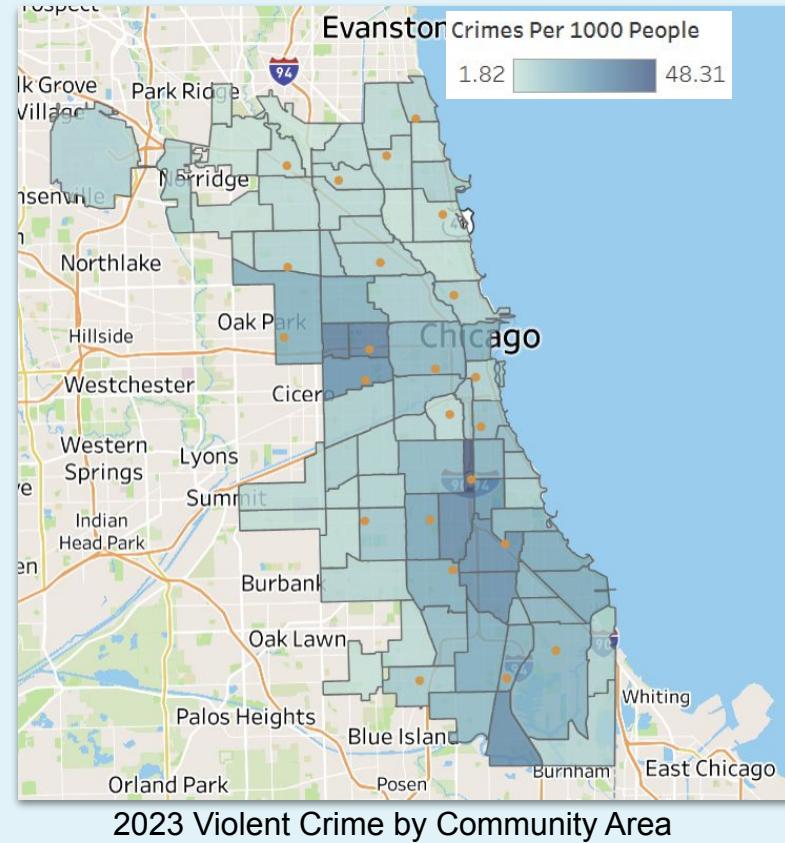
07 Demo

Problem Statement and Overview



Problem Statement

- Chicago had the 6th highest homicide rate among top cities in 2023
 - 2023 violent crime was at an 11-year high
 - 73 of 77 Community Areas exceed the national violent crime rate
 - 51 of 77 Community Areas exceed the property crime rate
 - Chicago's 2023 automobile theft was at a 20 year high



Overview

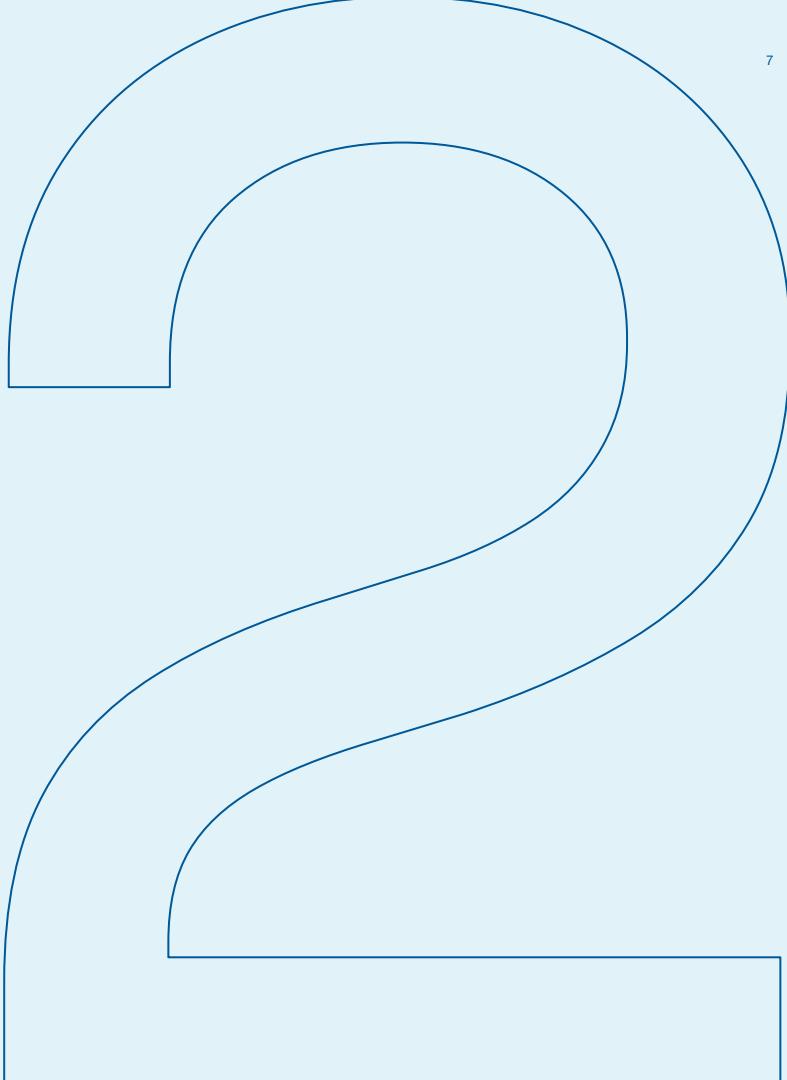
**One of Chicago's main objectives
for 2024 is crime prevention,
including data-driven policing**



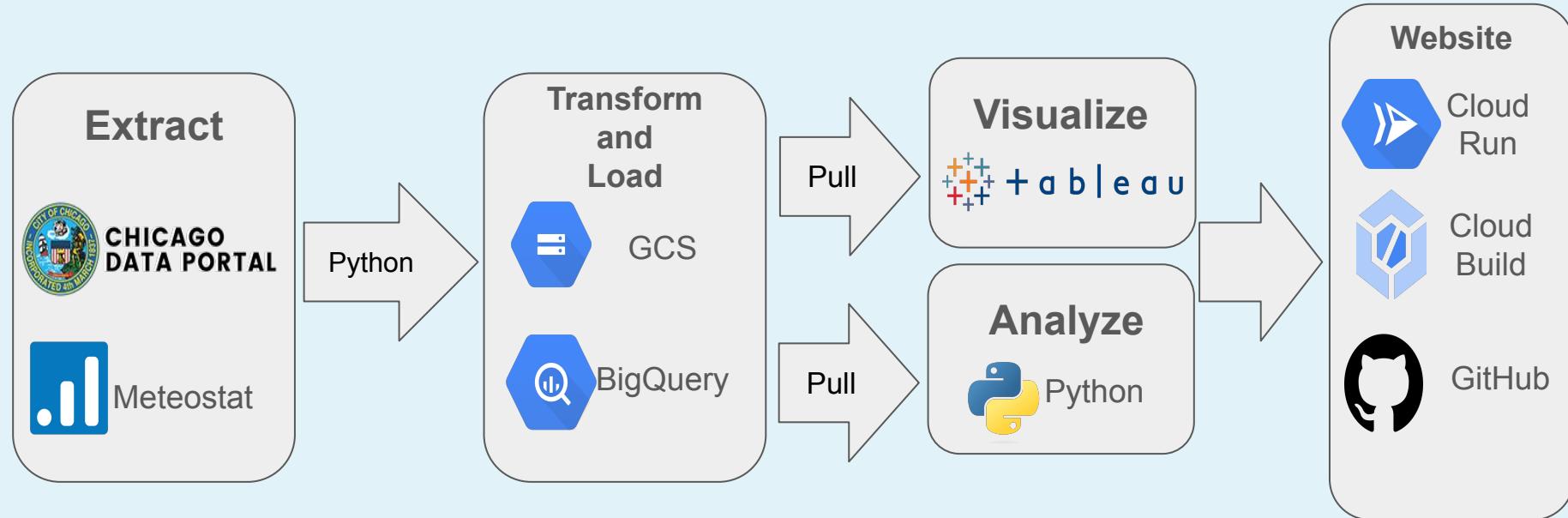
Team 51 Developed and Delivered Robust Solutions to Support Chicago in Reaching its Crime Reduction Goals

-  Understanding of the Top Crime Drivers and Crime Hotspots
-  Analytics and Data-science Models
-  Conclusions and Recommendations
-  Information Access Tools: Dashboards, Mobile App, Chatbot, Website

Data



Data & Architecture Diagram



Exploratory Data Analysis

Crime from Jan 2019 - Jan 2024

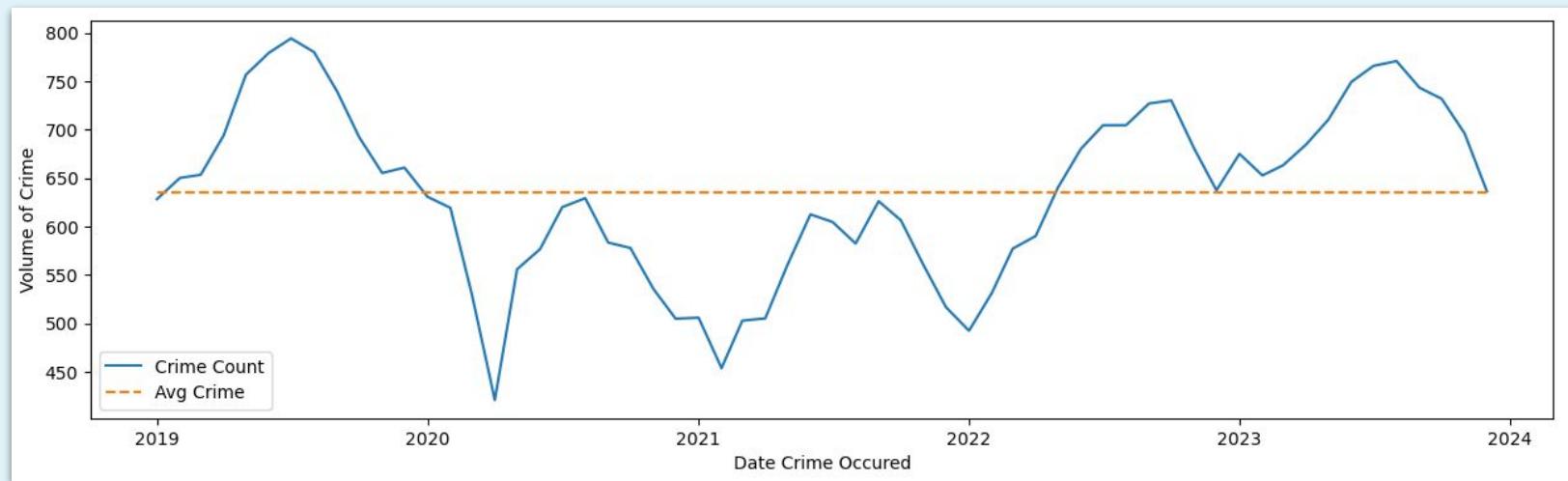


Figure 1. Number of Crimes Committed in Chicago from 2019-2023

2023 Crime Rates - By Category (Per 1,000)

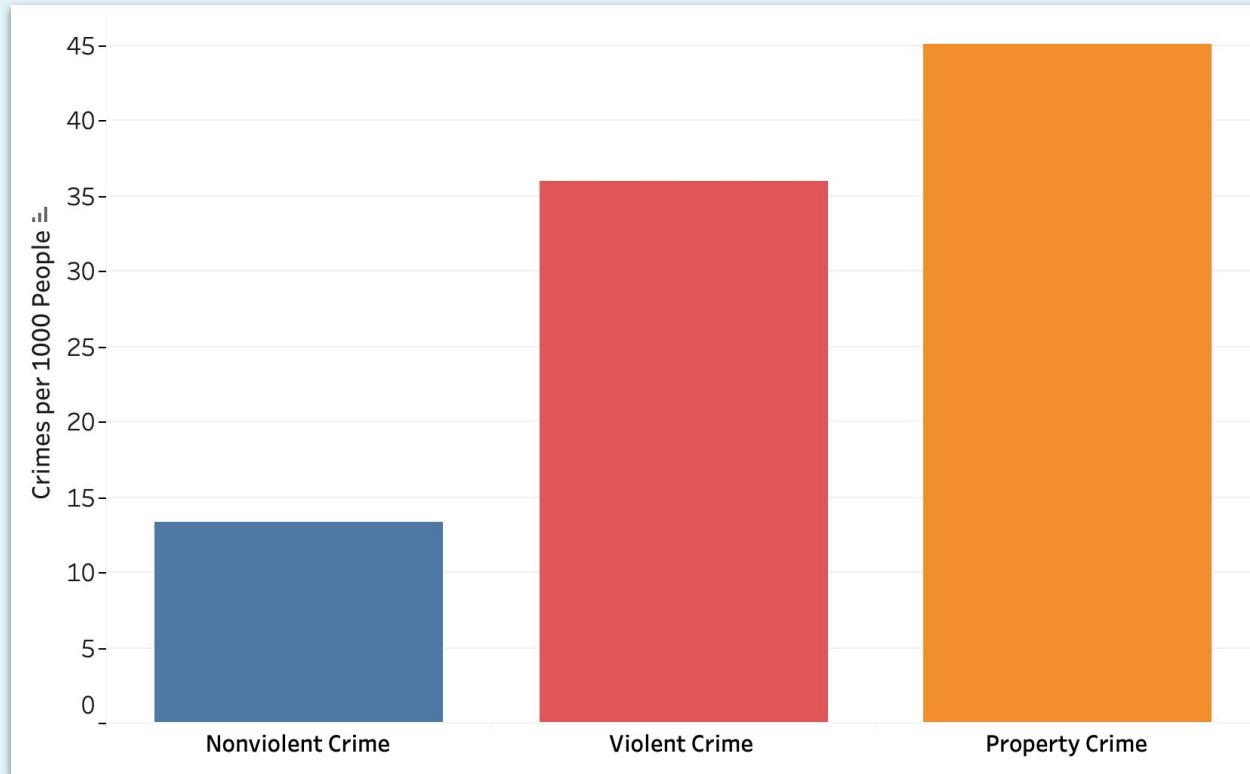


Figure 2. Crime Rates Per 1000 People by Crime Type for 2023

2023 Crime Rates - By Category and Month (Per 1,000)

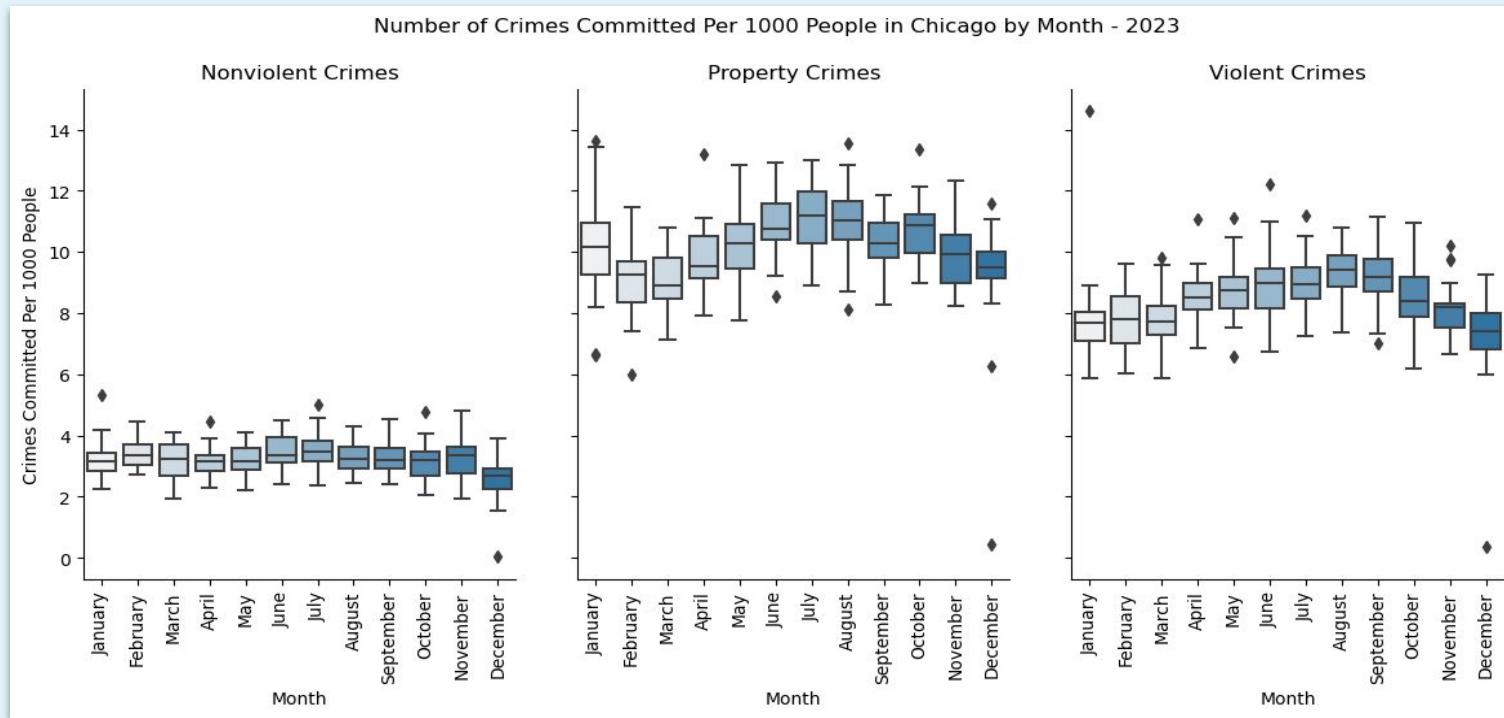


Figure 4. Chicago Monthly Crime Rate Compared to Average Temperature for 2023

2023 Crime Rates - By Community Area (Per 1,000)

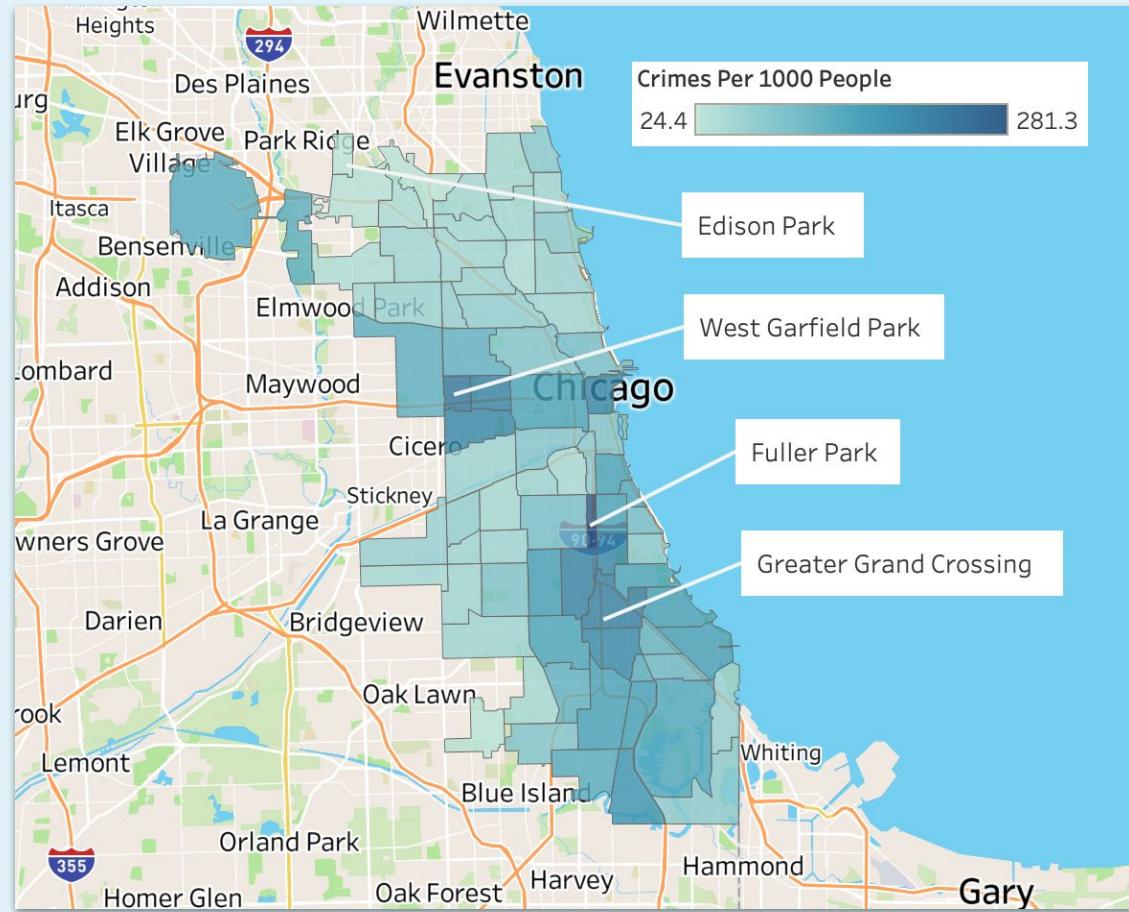
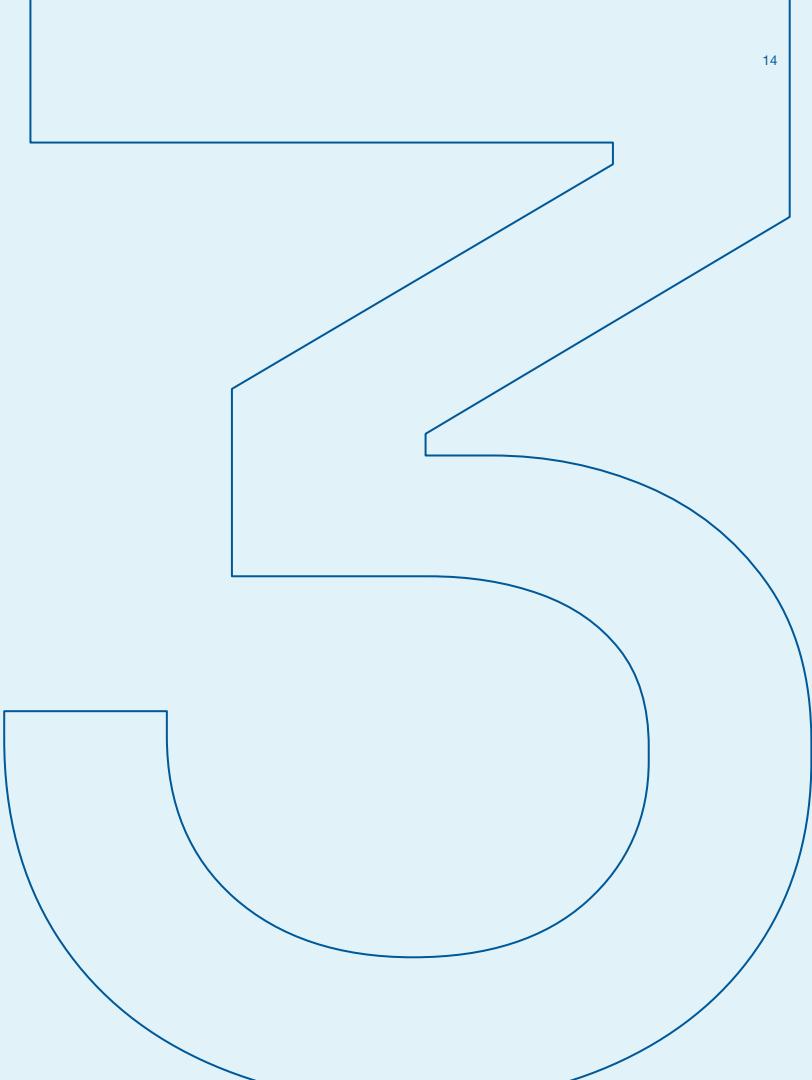


Figure 5. Crime Per 1000 People by Chicago Community Area in 2023

Modeling



Supervised Learning Methods

5 Supervised Methods from Team 51

Supervised methods are used to train predictive models, and are the types of methods Team 51 is using to predict crime types and crime counts. They classify outcomes based on input data by learning the relationship between inputs and corresponding outputs from labeled training examples.

Types of unsupervised models used:

- Linear Regression
- Logistic Regression
- Decision Trees
- Random Forest
- Gradient Boosting

Crime Type Prediction

Accuracy and RMSE for Supervised Learning - Crime Type (Bar)

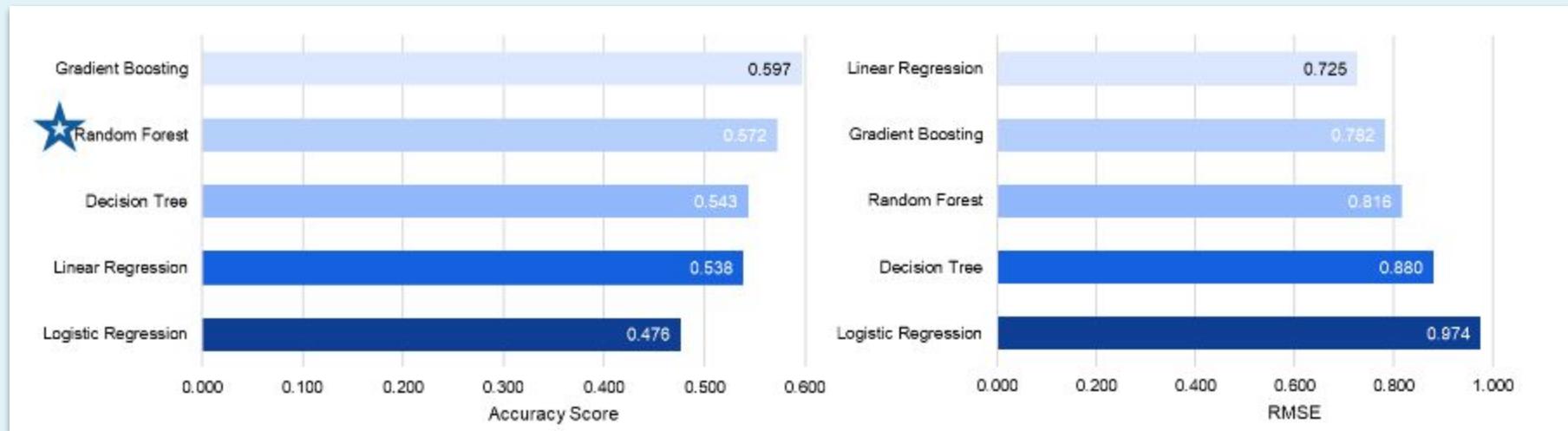


Figure 7: Performance for Supervised Models Based on Accuracy Score (left) and RMSE (right) for Test Data

Feature Importance - Crime Type

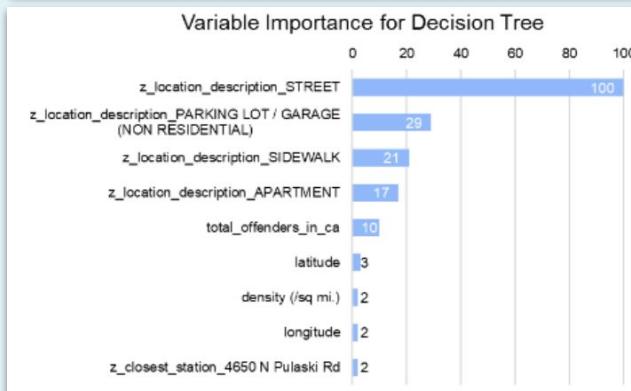
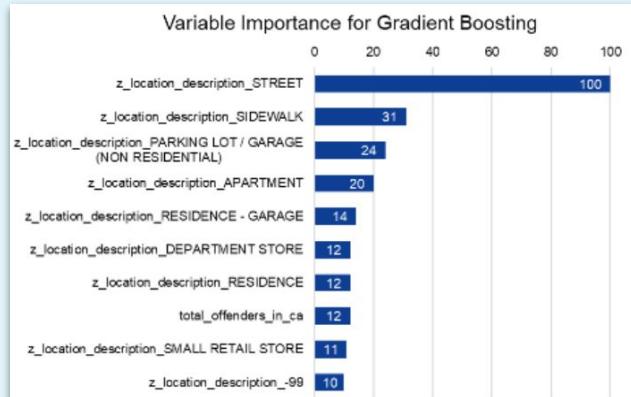


Figure 8. Feature (variable) Importance of Predicting Crime for Each Model

Crime Count Prediction

Accuracy and RMSE for Supervised Learning - Crime Count

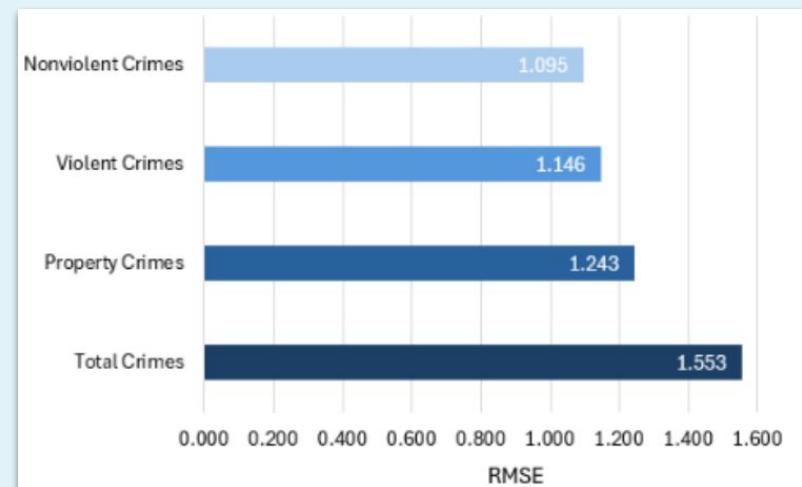
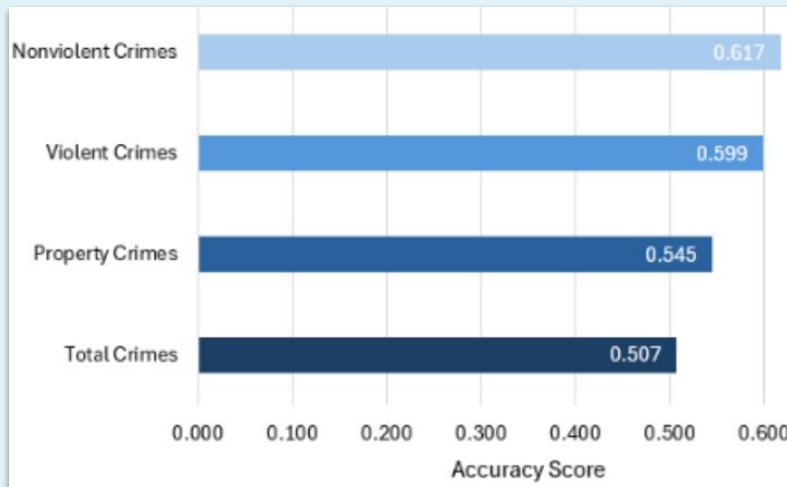


Figure 9. Crime Count Trend Prediction Performance of Random Forest Model Based on Accuracy Score (top) and RMSE (bottom) for Categorized Crime Type with Test Data

Feature Importance - Crime Count

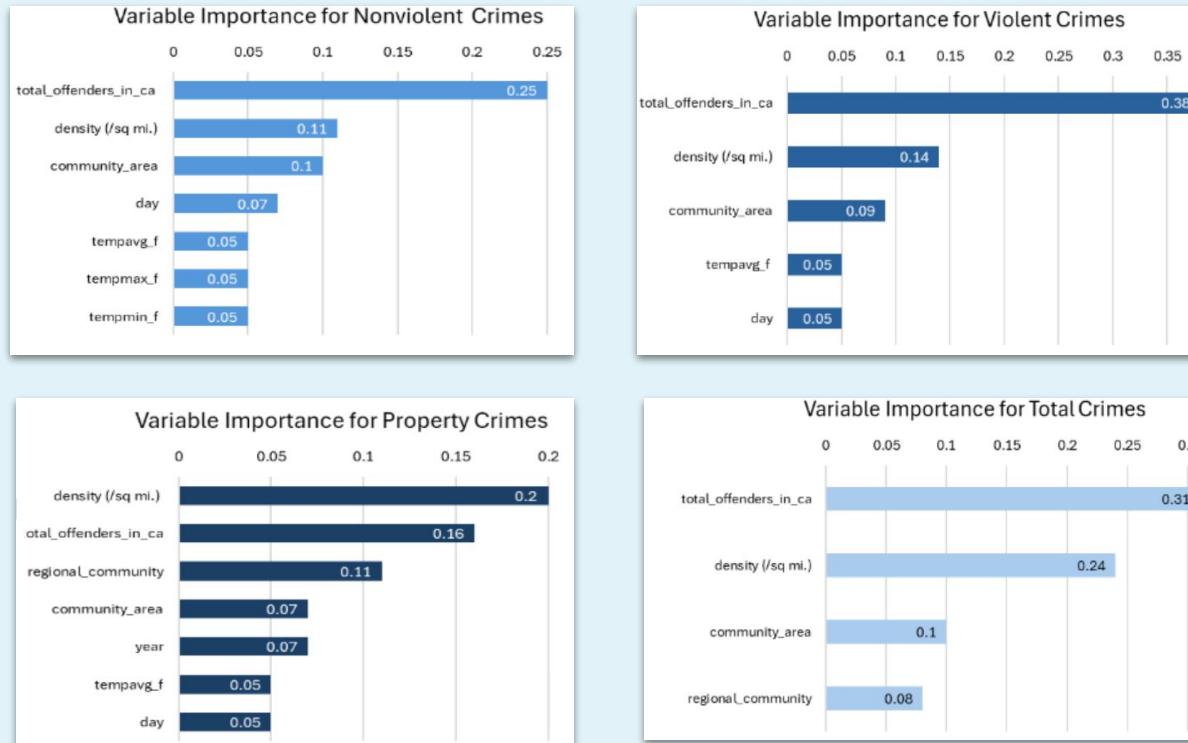


Figure 10. Feature (variable) Importance of Predicting Trends in the Number of Crimes for Total and Categorized Crimes

Crime Trends



Figure 11. Crime Trend Predictions Based on Test Data (20%) For Overall (top left), Non-Violent (top right), Violent (bottom left) and Property (bottom-right) Crimes

Crime Count Forecasting

Tableau Forecast - Crime Count



Figure 12. Total Crimes Per Month Prediction for 2024 (Using Tableau Forecasting)

Tableau Forecast - Crime Count by Type

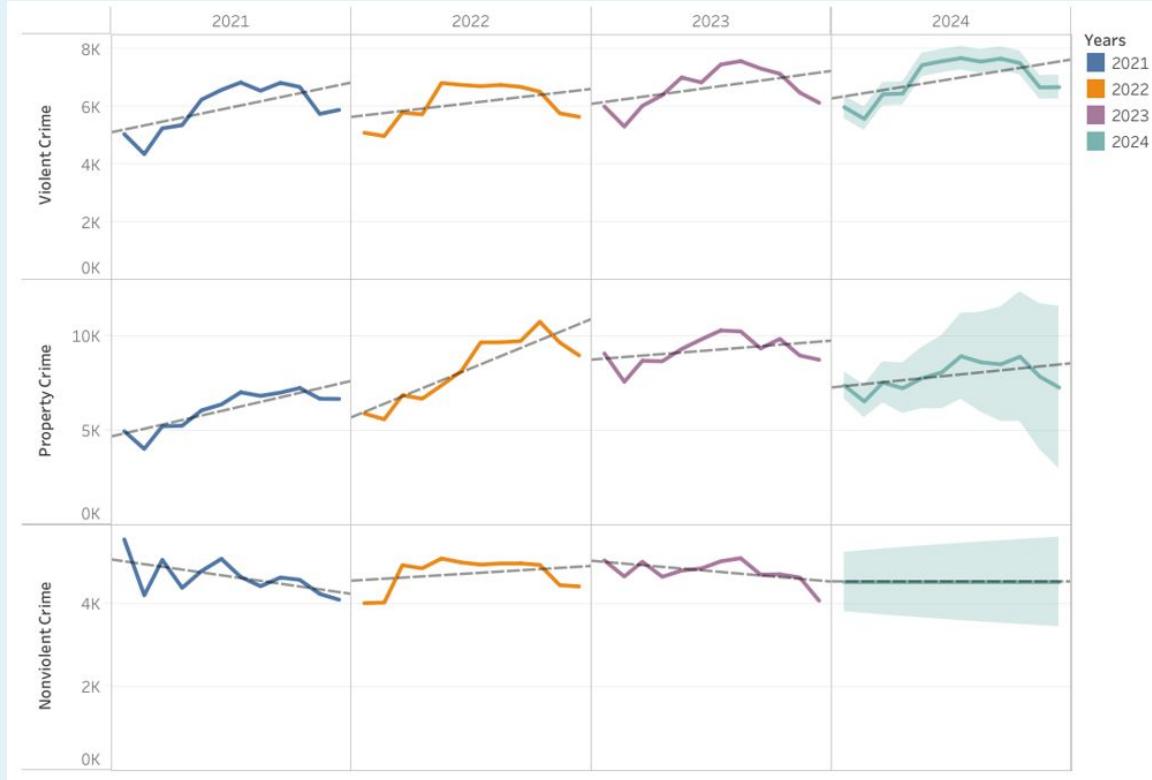


Figure 13. Crimes Per Year Prediction for 2024 for Each Categorized Crime Type (Using Tableau Forecasting)

Unsupervised Learning Methods

3 Unsupervised Methods from Team 51

Unsupervised methods are intended to identify relationships, patterns and clustered groupings in data that does not have a predefined target variable. Unsupervised methods are being used for segmentation to identify crime profile clusters that share similar characteristics.

Types of unsupervised models used:

- K-Means
- Clustering models
- Principal Component Analysis (PCA)

Hierarchical Clustering

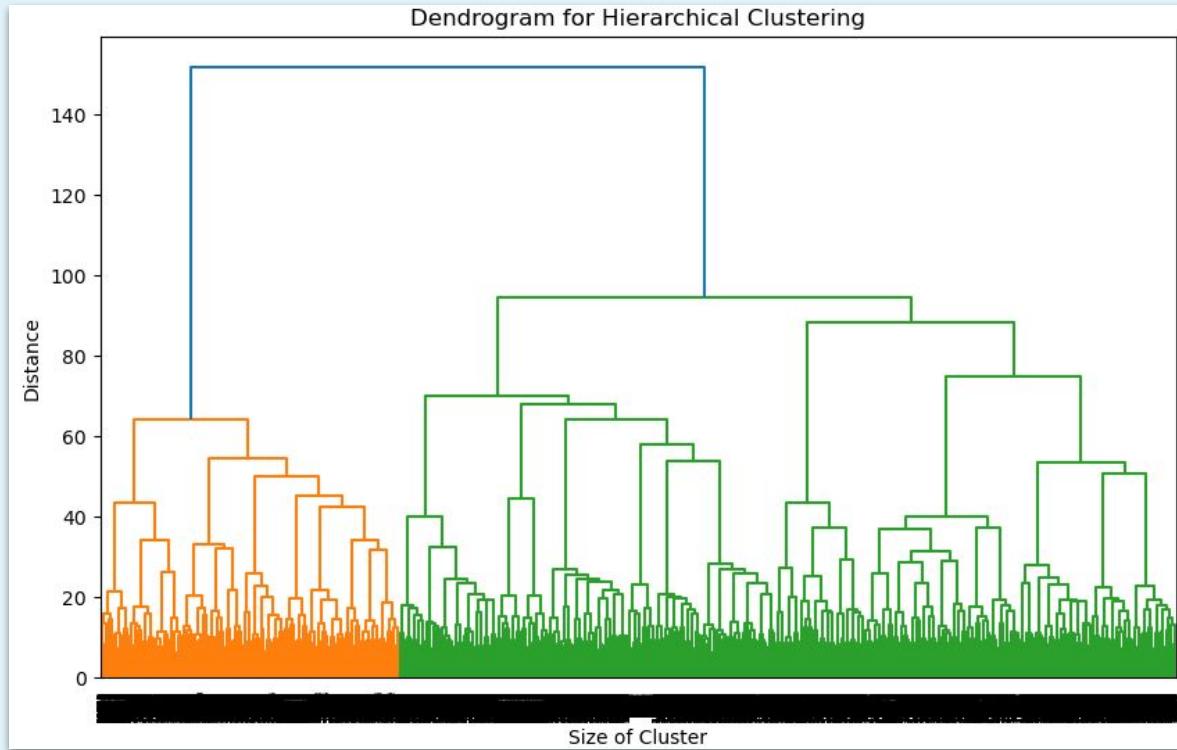


Figure 14. Visual representation of the structure of hierarchical clusterings with two distinct clusters

K-Means

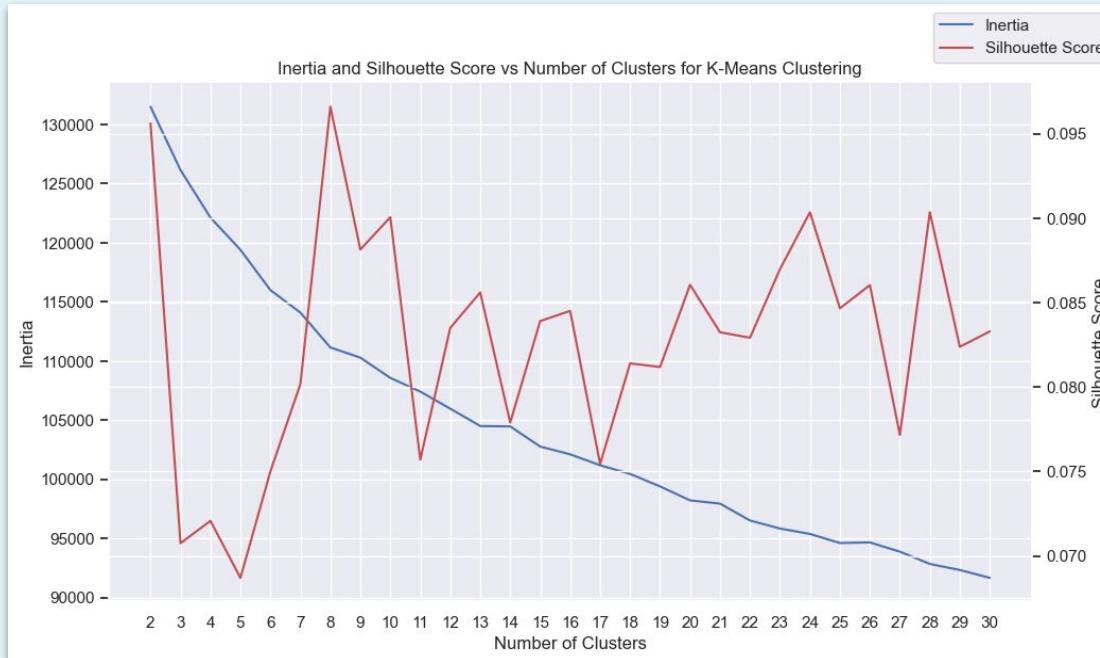


Figure 15. Line graph to depict the optimal number of clusters to use

Map of Chicago Based on Crime Profile Clusters

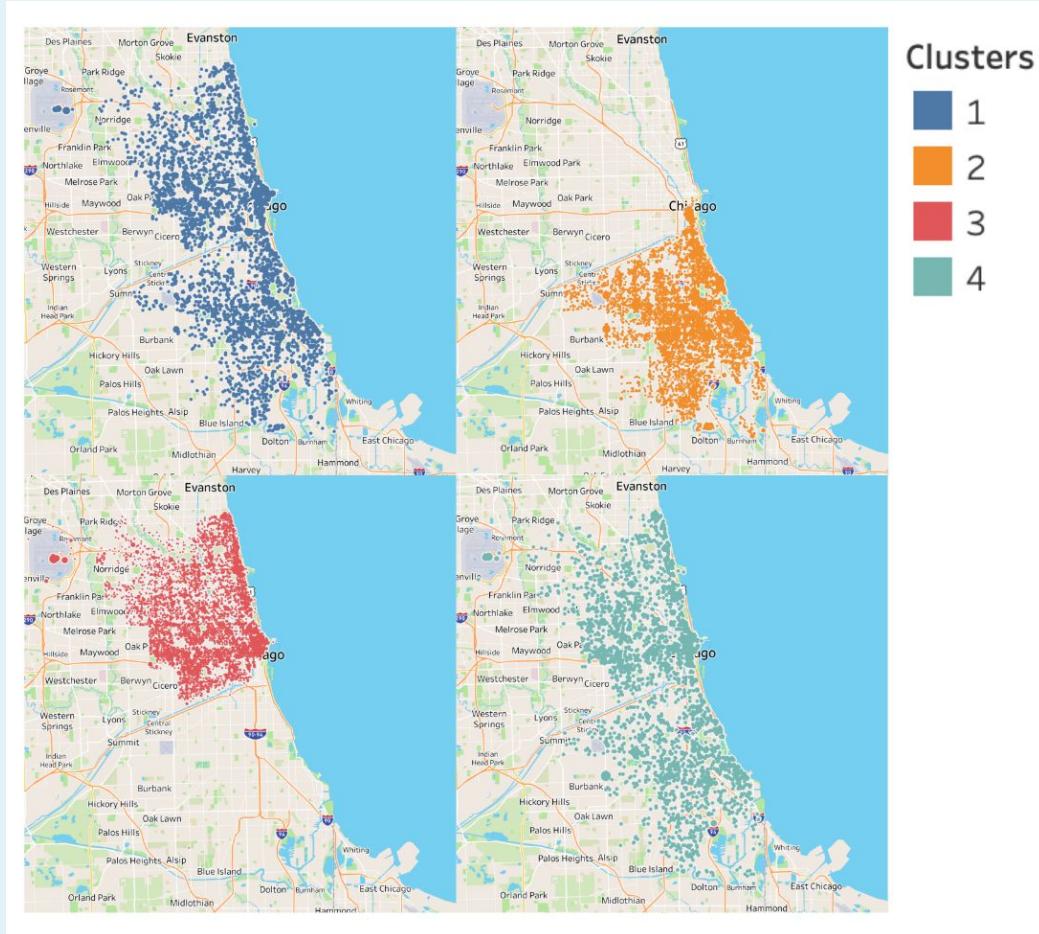


Figure 16. Crime Profile Clusters Separated by Cluster Size

Principal Component Analysis

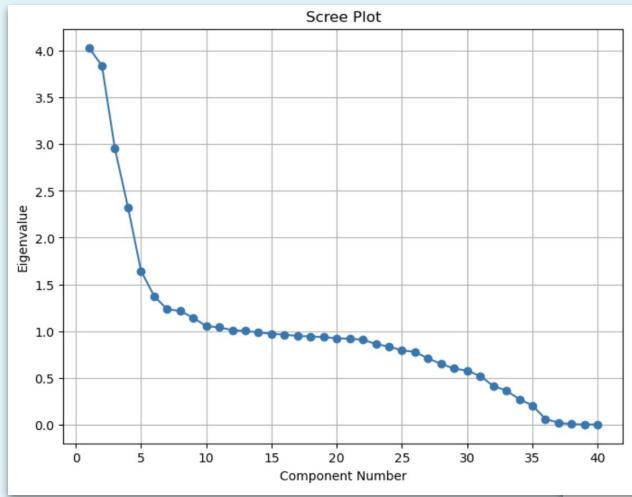
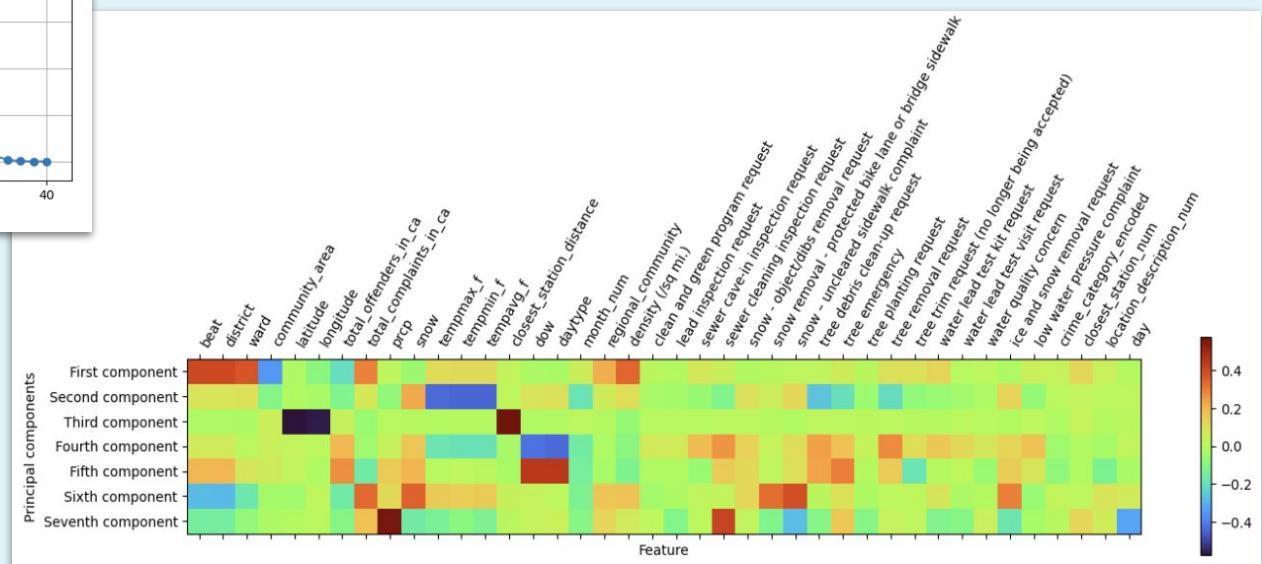


Figure 17. Scree Plot to Determine the Number of Principal Components (PCs)

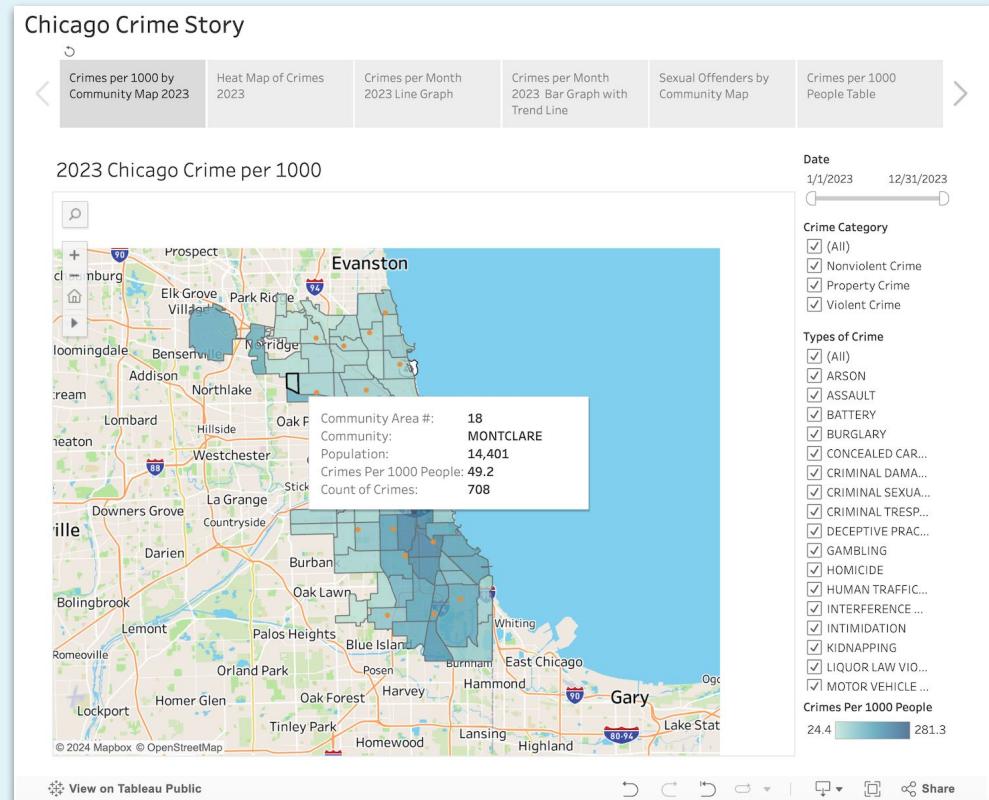
Figure 18. Heatmap of Principal Components by each variable



Information Access

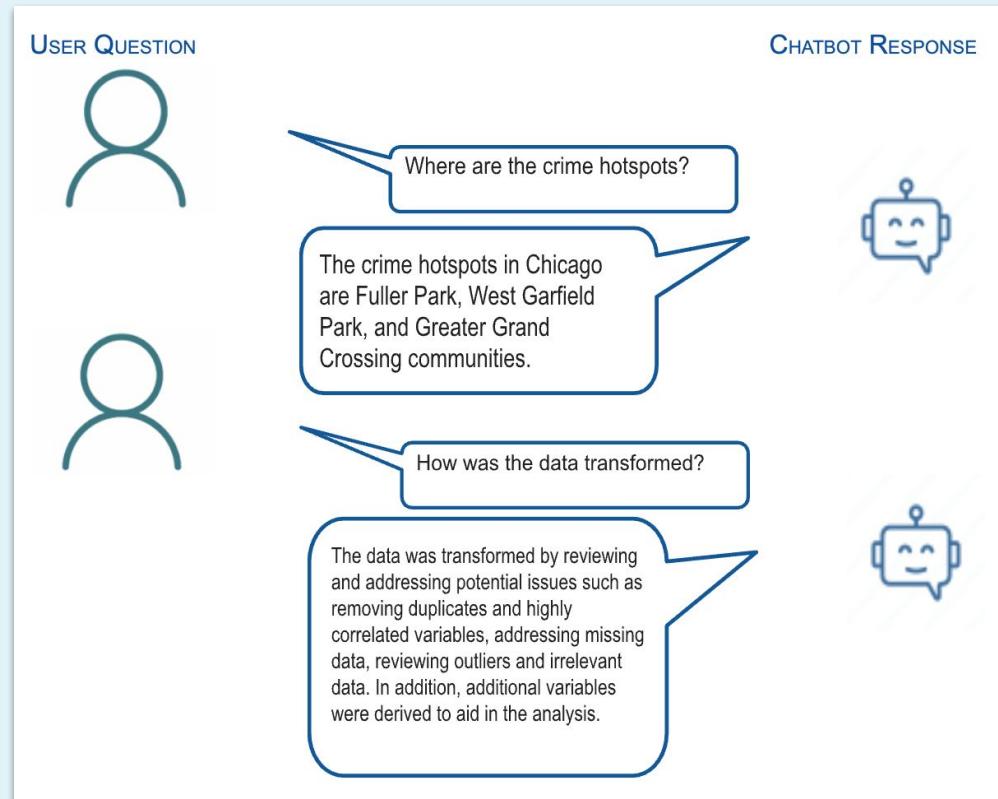
Dashboards

- 3 Dashboards
 - Historical
 - Model Performance
 - Tableau Forecasting
- Provides insights at the community level
- Monitor and Predict



Chatbot

- Quick access to information on the project
- Answers questions based on the report
- Developed with Python and power of Open AI's ChatGPT 3.5



Website

- Houses dashboards and chatbots in one easy to access location
- Editable and expandable
- Updated through GitHub & Google Cloud Build for quick updates
- 100% uptime through updates

The screenshot shows a website interface with a dark blue sidebar on the left and a light gray main content area on the right.

Sidebar (Left):

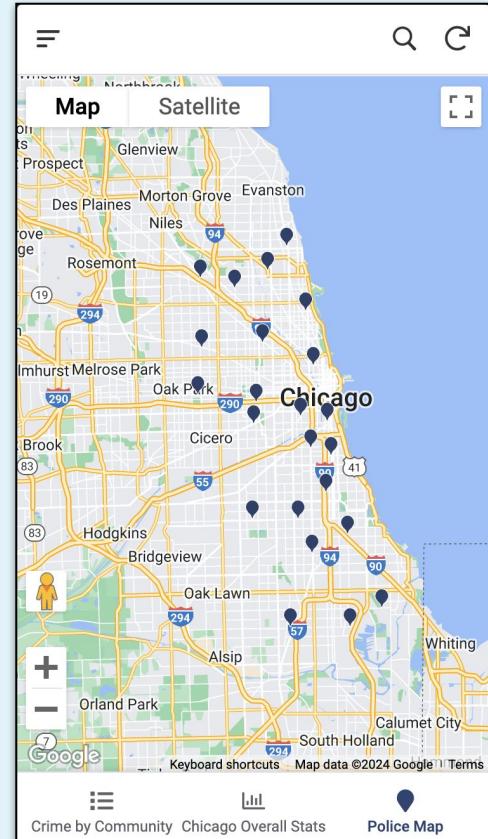
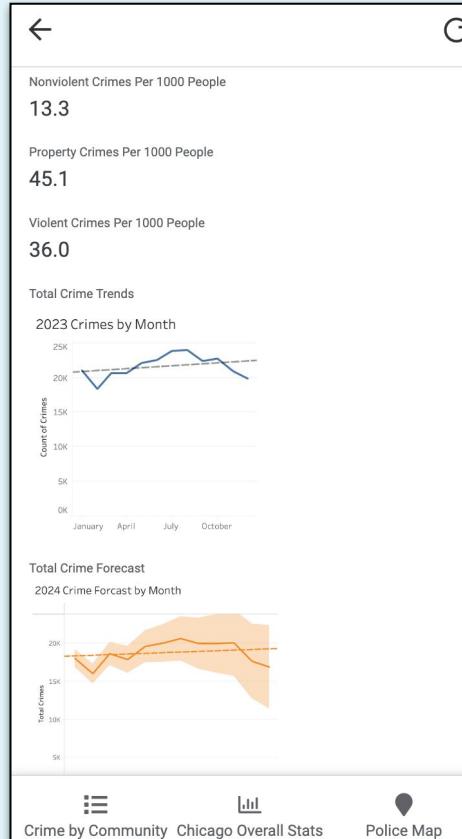
- TEAM 51** logo
- Team 51 x Chicago** header
- HOME** section with [Home Page](#)
- DASHBOARDS** section with [General Dashboards](#), [Tableau Forecast](#), and [Model](#)
- TOOLS** section with [Chatbot](#)

Main Content Area (Right):

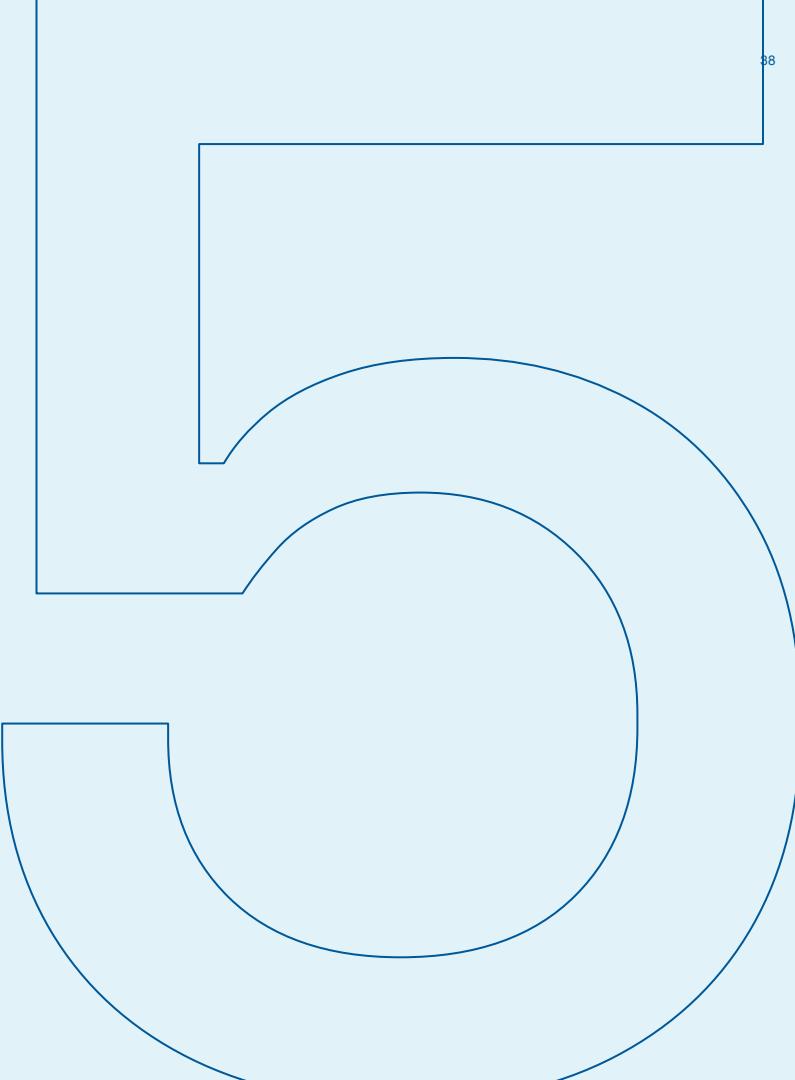
- Form Elements** header with a search icon.
- / ChatBot** breadcrumb.
- ChatBot Input** section with an input field containing "Who is on the team" and a blue **Query** button.
- Input Question** label above the input field.
- Response** section below the input field, displaying the names: Nicole Crothers, Booga Gilbertson, Fatima Gowher, TK Kim, Matt Riegsecker.

Mobile App

- Available to Chicago officials, police, and residents
 - Provides historical crime data and forecasting at the city and community level
 - Map of police stations with phone numbers and directions



Conclusions



Conclusions: Insights from Team 51's analysis reveal the characteristics of crime in Chicago



Crime rates are dominated by property and violent crime



Crime is the highest in the South and Far Southeast side of Chicago. Fuller Park's crime rate is more than 10x that of Edison Park (lowest crime rate)



Crime rates are higher in the warmer months, May through September



Longitude/latitude, police station proximity, weather, and crime location description (e.g. "street") are the top factors that predict crime type



Population density and number of sex offenders are the top two predictive factors for crime rates



Snow removal, tree removal, lead, and sewer cleaning environmental complaints are associated with crime occurrences

Conclusions: Analytic models provide robust data-driven crime insights; technology tools make information accessible



- Random Forest models were deployed that effectively predict crime type and crime rate, are highly interpretable, and provide informative predictive variables
- A Tableau exponential smoothing model was deployed that provides indicative 2024 crime trend forecasts
- K-Means and PCA modeling identified associations between environmental complaints and crime occurrences that supervised modeling did not emphasize



- Dashboards, mobile app, chatbot, and a website provide ready access to crime information

Recommendations

Recommendations Developed For the City of Chicago

01 Evaluate and align crime response resources in the crime hotspot community areas to address staffing, shift design and police support with crime activity and volume

02 Evaluate and promote crime awareness and prevention strategies targeted to the top factors associated with crime:

- Geographic location
- Weather
- Crime location

03 Deploy and promote information access tools

- Chicago crime story dashboards
- Mobile app
- Chatbot
- Website

04 Review the mitigation of 311 environmental complaints to assure that potential crime factors are also considered as part of mitigation

Recommendations For Future Project Phases, and Market Expansion

Future Project Phases:

Crime prediction models and information access tools should be maintained and augmented with the newest machine-learning enhancements as more data is gathered along with resultant insights about crime mitigation strategies.

Future Market Expansion:

Team 51's **innovations and investment from the Chicago project are scalable to other markets** with minimal additional investment and would **increase revenue** from the intellectual property that has already been developed. Team 51 has already assessed publicly available data for Los Angeles and New York City. **We recommend that we engage Los Angeles and New York City officials to present and market our solution.**



Fatima Gower
Data Scientist, Data
Exploration, Visualization



Nicole Crothers
Data Scientist, Data
Exploration, Predictive Models



TK Kim
Data Scientist, Data
Engineer, Cloud Platforms



Booga Gilbertson
Data Scientist, NLP,
Communications



Matt Riegsecker
Data Scientist, Project
Manager, Data Analysis

PARTNERING WITH YOU TO EMPOWER A SAFER COMMUNITY THROUGH
DATA-DRIVEN CRIME INSIGHTS