Chicago Crime: Predicting Crime Density by Ward

By: Cole Akers & Steven Delaney

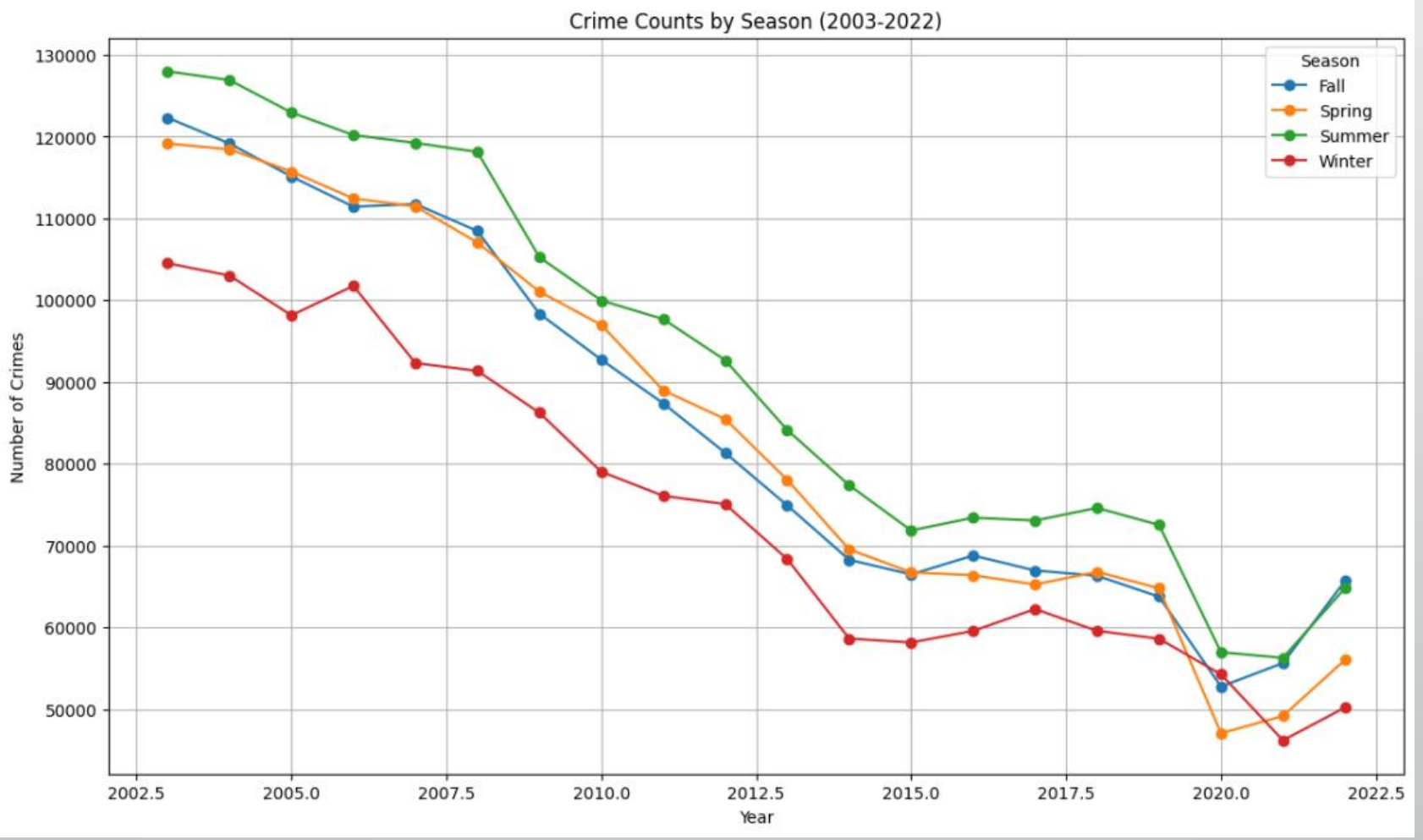
CSPB 4502: Data Mining

Chicago Crime is Notorious For Its Frequency and Violence

- Chicago is considered the leading crime capital in the United States
  - Over 600 homicides in 2023
    - https://crimelab.uchicago.edu/wp-content/uploads/sites/2/2024/04/UChicago-Crime-Lab-2023-End-of-Year-Crime-Trends.pdf
- Majority of violent crimes are carried out by gangs
  - Gangs responsible for over 60% of homicides
    - https://chicagocrimecommission.org/gang-book-2018%20Open%20link
- Gun violence is frequent
  - Over 2,700 shooting victims in 2023
    - https://**home.chicagopolice.org**/statistics-data/data-dashboards/

Despite Its prevalence, crime has been steadily decreasing in Chicago year over year

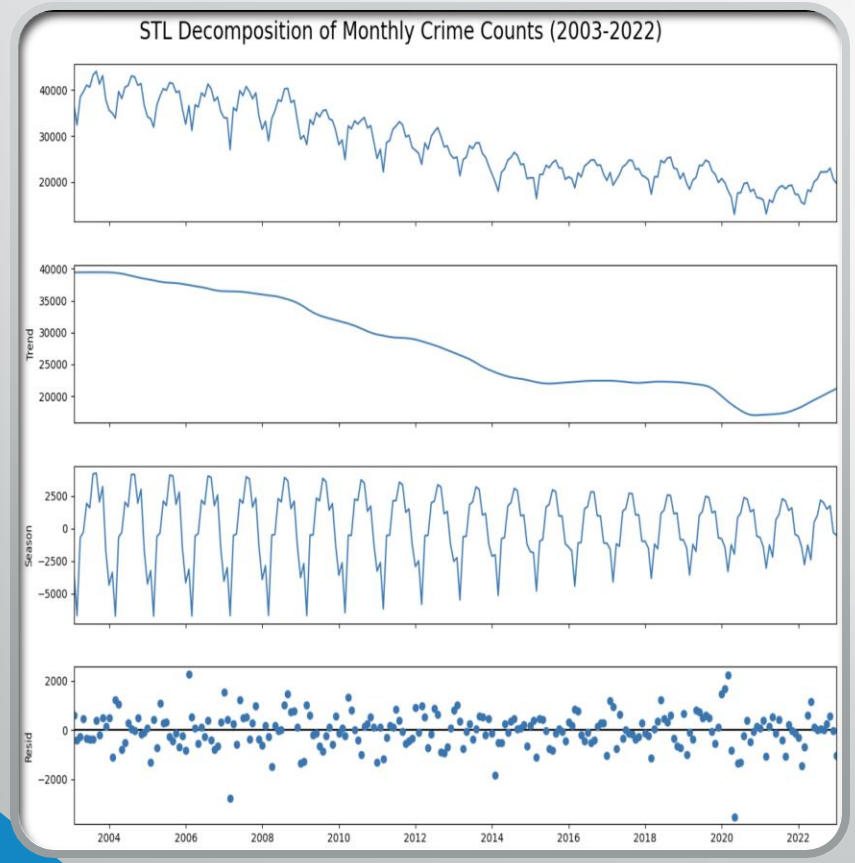
- The number of reported cases has dropped an average of 6% every year compared to the previous year.



Why is crime declining? Analysts Speculate

- Strategic Policing Reforms (CompStat, Data-Driven Deployment)Allocation of resources
  - Bureau of Justic Assistance (2013), COMPSTAT: Its Origins, Evolution, and Future in Law Enforcement Agencies
- Demographic Shifts and Aging Population
  - Levitt, S. (2004). Understanding Why Crime Fell in the 1990s: Four Factors that Explain the Decline. Journal of Economic Perspectives.
- Decline in Lead Exposure
  - Reyes, J. (2007). Environmental Policy as Social Policy? The Impact of Childhood Lead Exposure on Crime. The B.E. Journal of Economic Analysis & Policy.
- Technological Advances (Cameras, ShotSpotter, Surveillance)
- Community-Based Violence Prevention Programs
  - Skogan, W.G., et al. (2009). Evaluation of CeaseFire-Chicago. Northwestern University Institute for Policy Research.
  - Cure Violence Global impact reports (2013–2020)

While crime is decreasing it's still occurring. Luckily, it's predictable



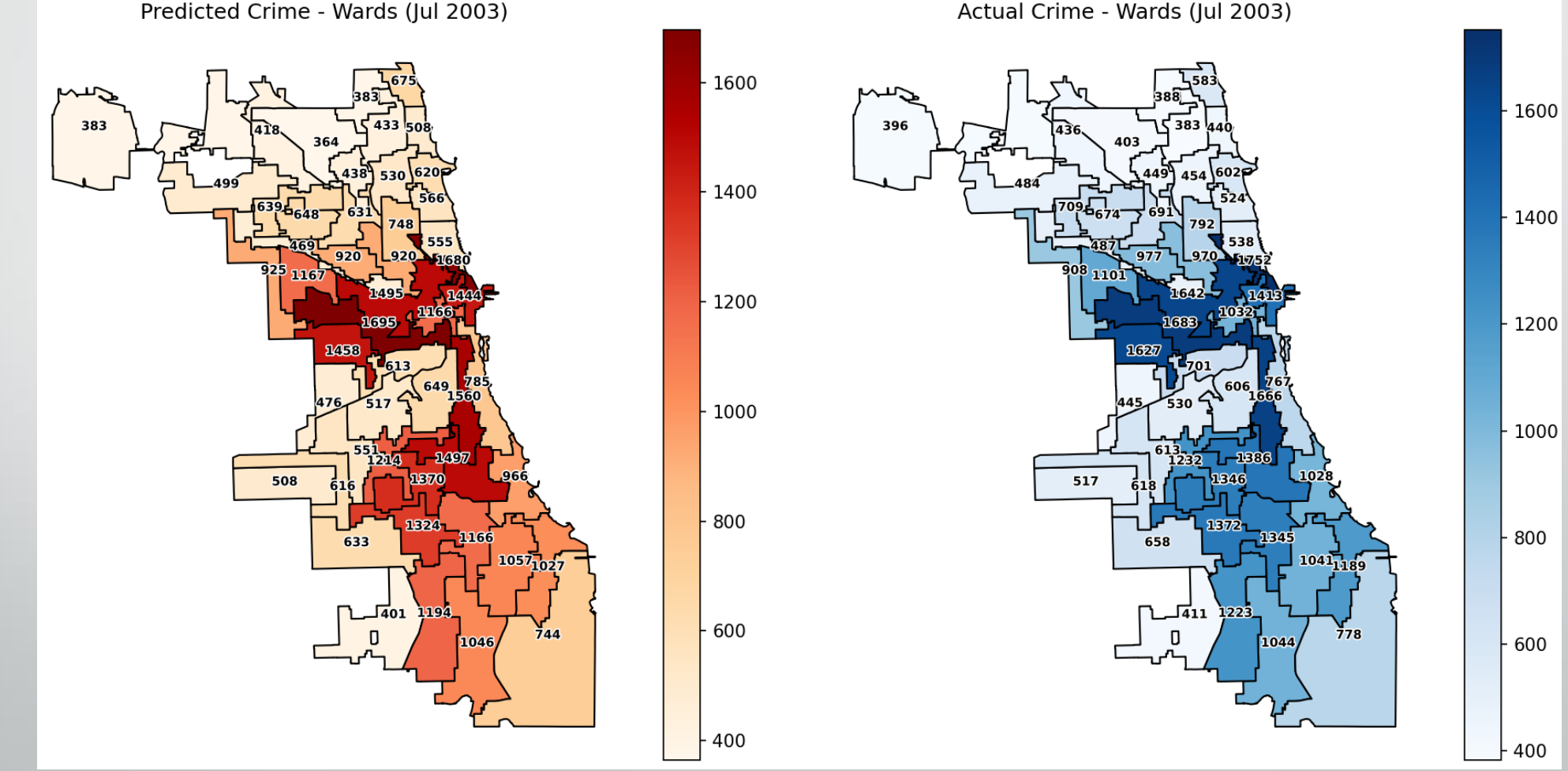
- Seasonal patterns and consistent geographic hotspots suggest crime is not random.
- We've developed a model to predict crime density across all wards in Chicago for any given year.
- The goal of this model is to enable pro-active intervention by identifying high risk areas before crime spikes occur
- Room for improvement: can't predict random spikes from unexpected events

Methodology

- Cleaned data, removed outliers, removed null values, eliminated redundant/unnecessary columns
- Final data set was composed of data from 2003 – 2022 which consisted of roughly 6M records.
- Exploratory analysis
  - Time-Series
  - Density
- Random Forest: Decision Tree Prediction Averaging
  - Dependent: Count of Crime by Ward
  - Independent Variables: Month, Year, Ward, Season, Daypart, Crime Lag 1 Month, Crime Lag 3 Month Average

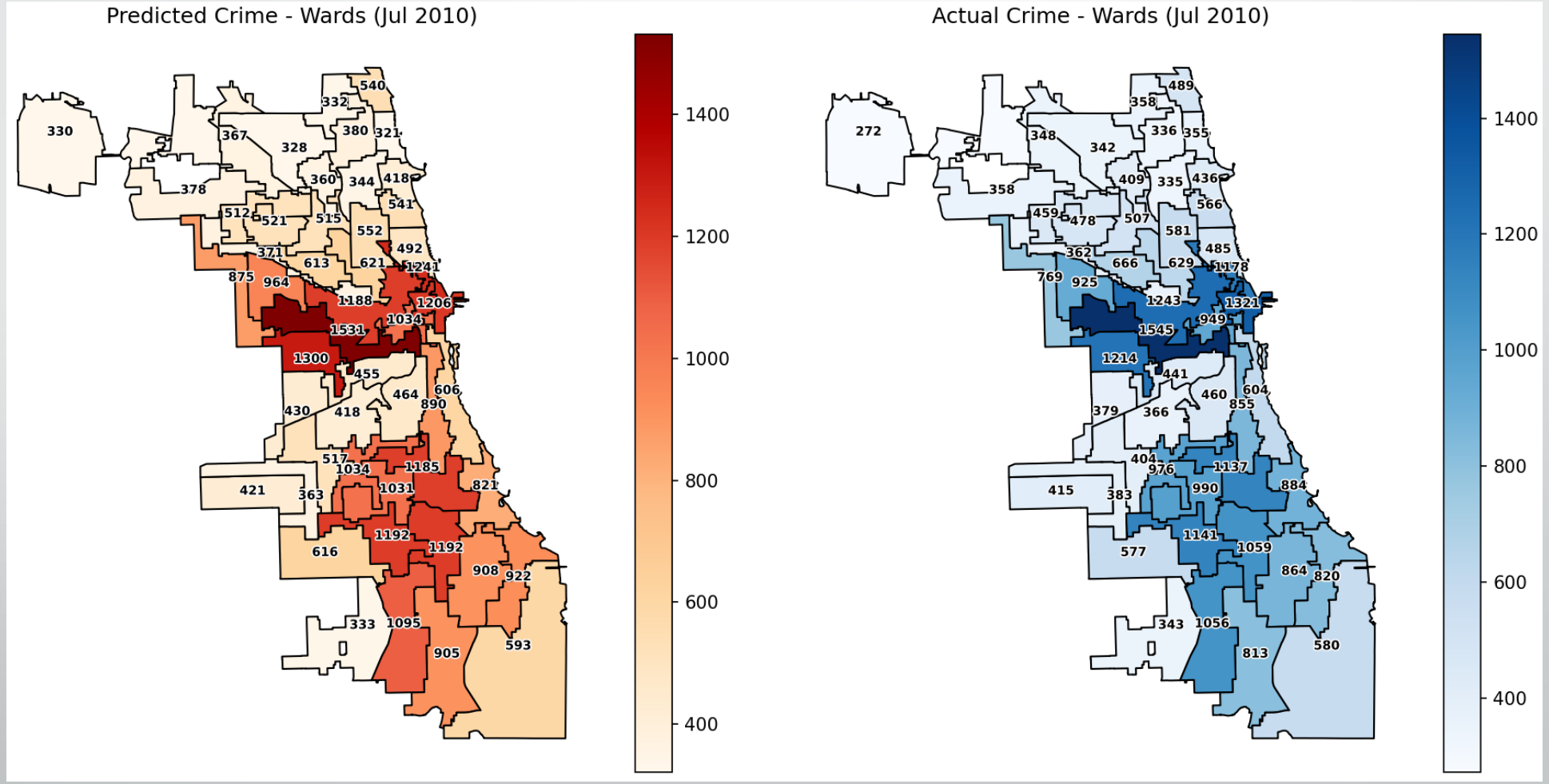
Results – July,2003

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



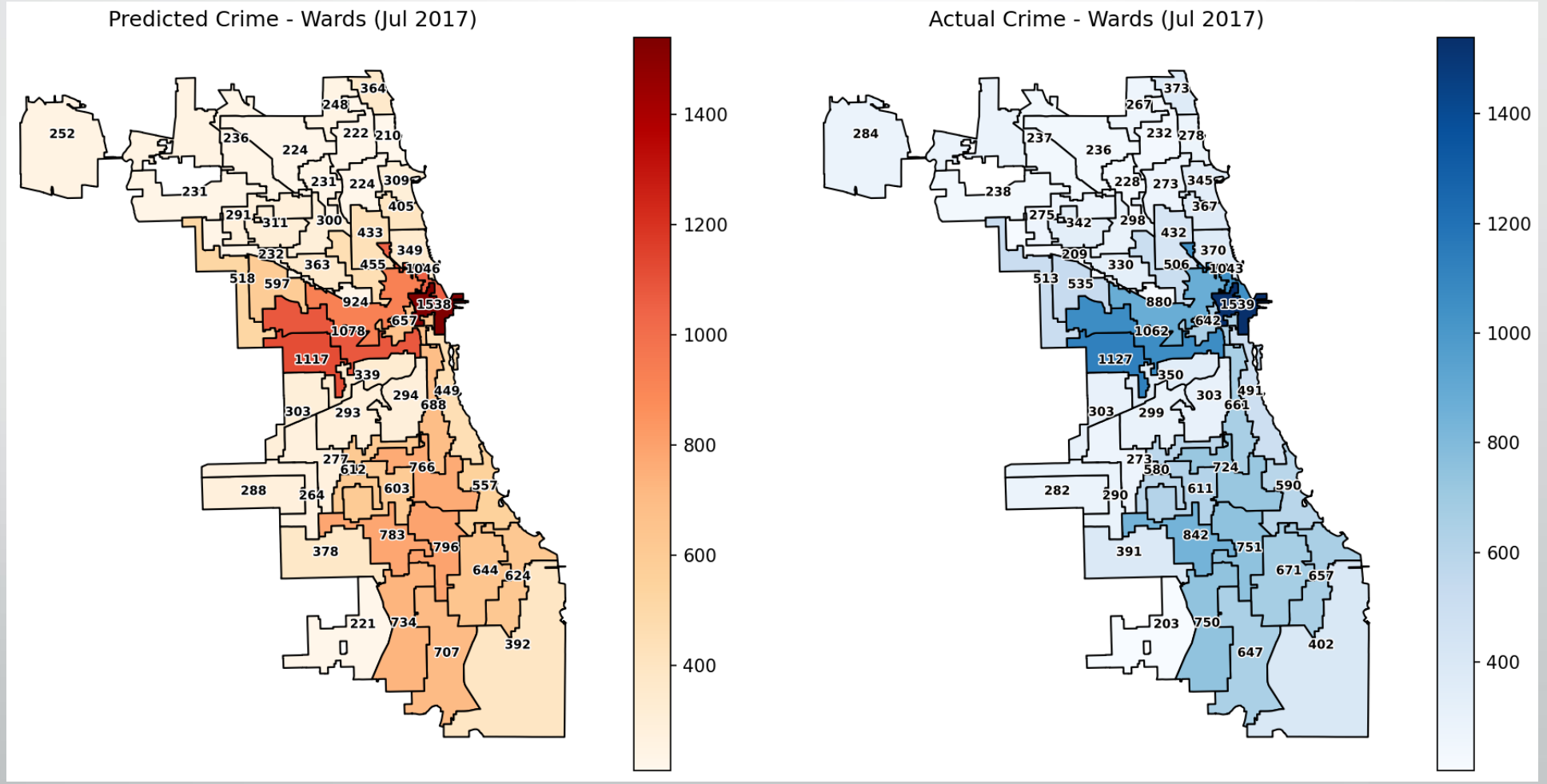
Results – July,2010

```
Evaluation Metrics:
MAE: 39.16
RMSE: 51.26
R²: 0.9724
MAPE: 6.96%
Average Percent Accuracy: 93.04%
```



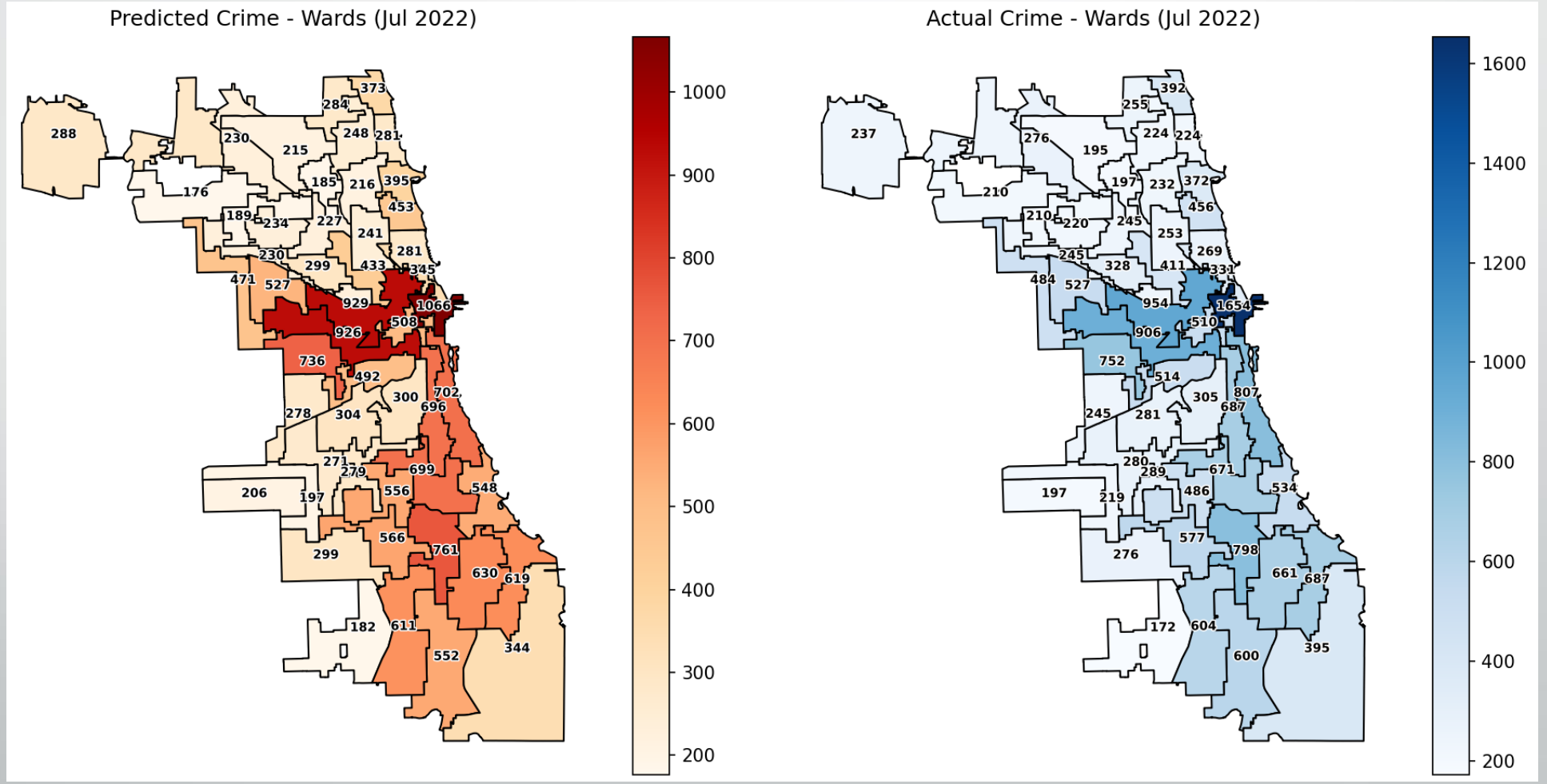
Results – July,2017

```
Evaluation Metrics:
MAE: 29.35
RMSE: 43.12
R²: 0.9723
MAPE: 7.12%
Average Percent Accuracy: 92.88%
```



Results – July,2022

```
Evaluation Metrics:
MAE: 35.28
RMSE: 54.79
R²: 0.9362
MAPE: 9.46%
Average Percent Accuracy: 90.54%
```



Application

- This work represents an initial step toward building a comprehensive predictive analytics framework.
- Allows Law Enforcement to anticipate hotspots and strategically allocate resources.
- Provides valuable insights for resource allocation.
- Data driven results that justify distribution of funding.
- Potential Benefits include:
  - Faster response times
  - More efficient use of police resources
  - Allows for a preventative approach versus responsive approach
  - Appropriate funding for crime spot areas

Conclusion & Retrospective

- Results show that while the number of crimes have consistently decreased city-wide, hot spots have remained consistent over the years. This fact is likely known by Chicago Law Enforcement but with data science it can be proven definitively.
- Our initial goal with this project was to build a model that predicts the primary type based on factors like ward, time of day, location description and seasonality. However, this seems impractical in application. This drove our decision to predict count of crime by ward.
- Given more time we make a more robust model and sharpening the accuracy
  - Indication of likely crime type by area
  - Indication of likely time (daypart) of occurrence of crime by location and type
  - Predicting how the crime landscape might change in reaction to law enforcement