

# 2025 CHICAGO CRIME ANALYSIS

---

By: Carlos Marota, Monica  
Luevano, Ezequiel Soto,  
Tehreem Butt, Alex Jimenez

# AGENDA

Data Specification and Originality

Data Analysis using ES

Predictive Analysis using ML

Q&A



# DATA SPECIFICATION

Filtered data from 2 GB by specifying data and columns

Data size: 21.5mb

Data source: [Crimes - 2001 to Present | City of Chicago | Data Portal](#)

Abstract: This study investigates reported crime occurrences in Chicago from January to November 2025—excluding the most recent seven days—using Elasticsearch and Kibana. The analysis examines how the frequency of various crime categories has changed throughout the year, exploring crime types, geographic distribution, and monthly patterns



# HW EXPERIMENT SPECIFICATIONS

- Filtered our data from January 2025 – November 2025.
- Removed unnecessary columns such as: ward, beat, block, FBI code, community area, district, and case number
- Used the columns leftover, which were: Id, date, primary type, description, latitude, longitude, and location.

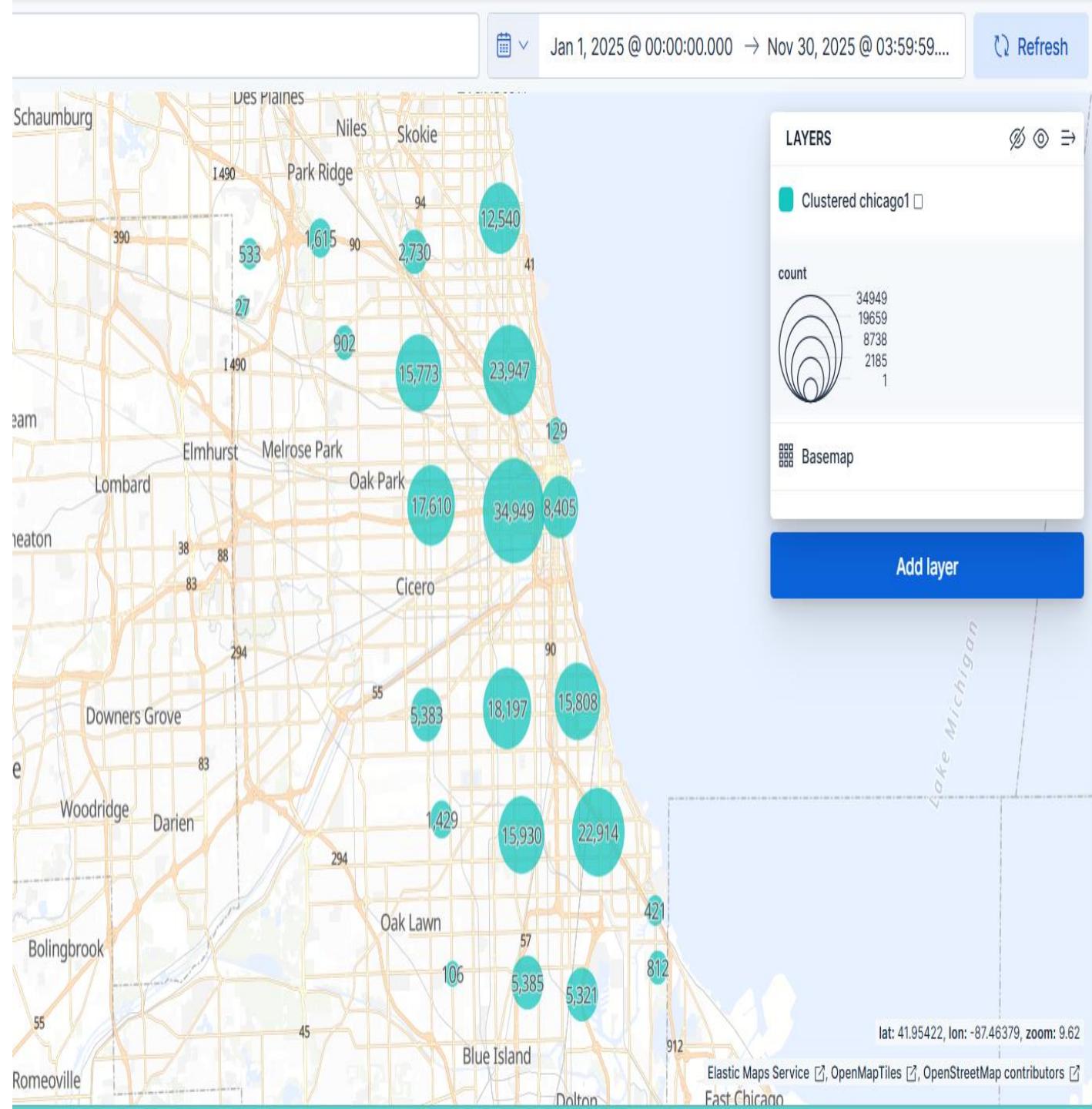


# ORIGINALITY

- This project narrows the focus to **Jan–No, 2025** to detect recent, relevant changes.
- We deliver actionable, up-to-the-minute insights rather than retrospective summaries, enabling tactical responses.
- Detect most relevant crimes of up to recent dates that may affect people more than past data.

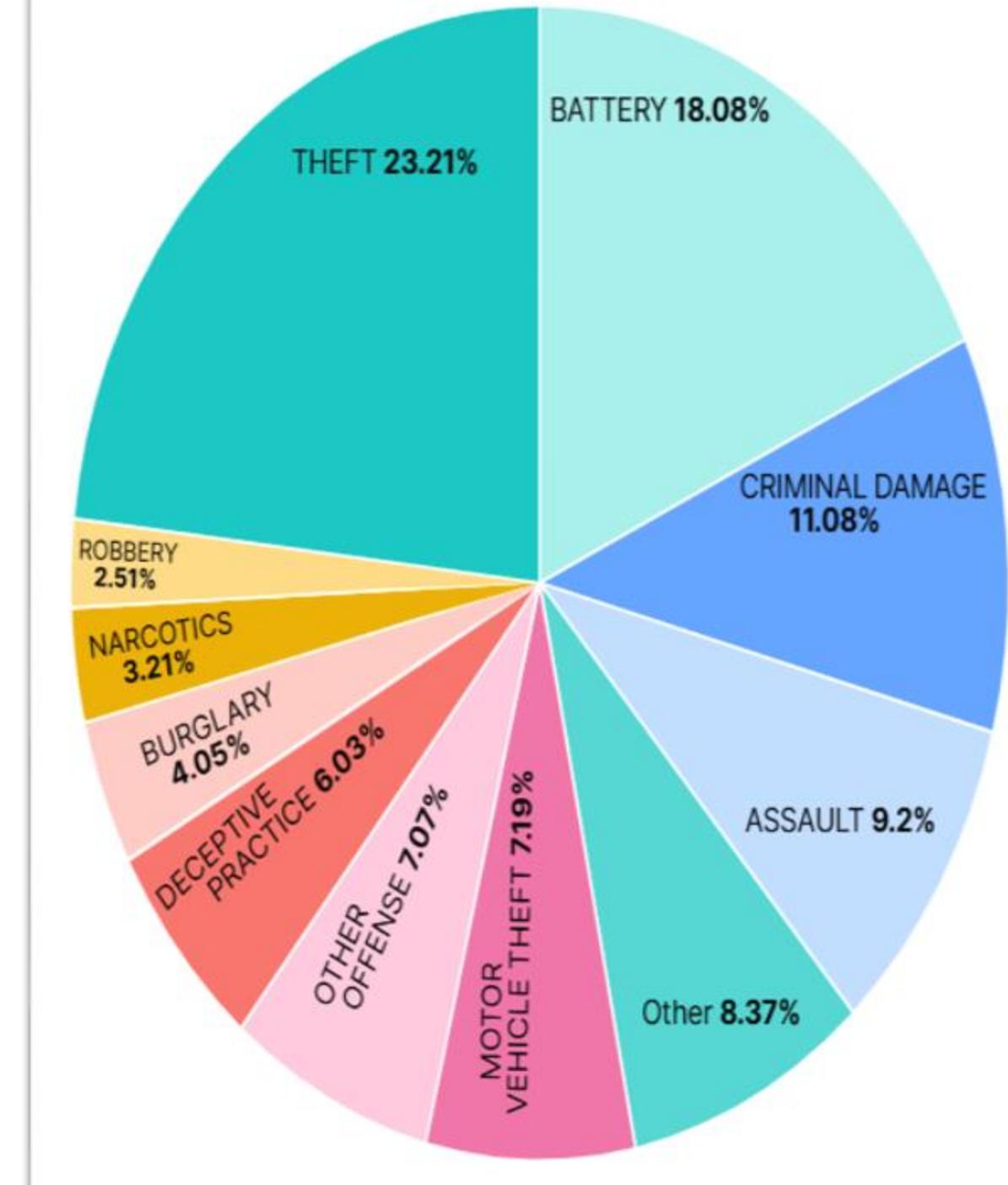
# DATA ANALYSIS USING ES

Cluster of the most committed crimes throughout Chicago from Jan 2025 to Nov 2025.



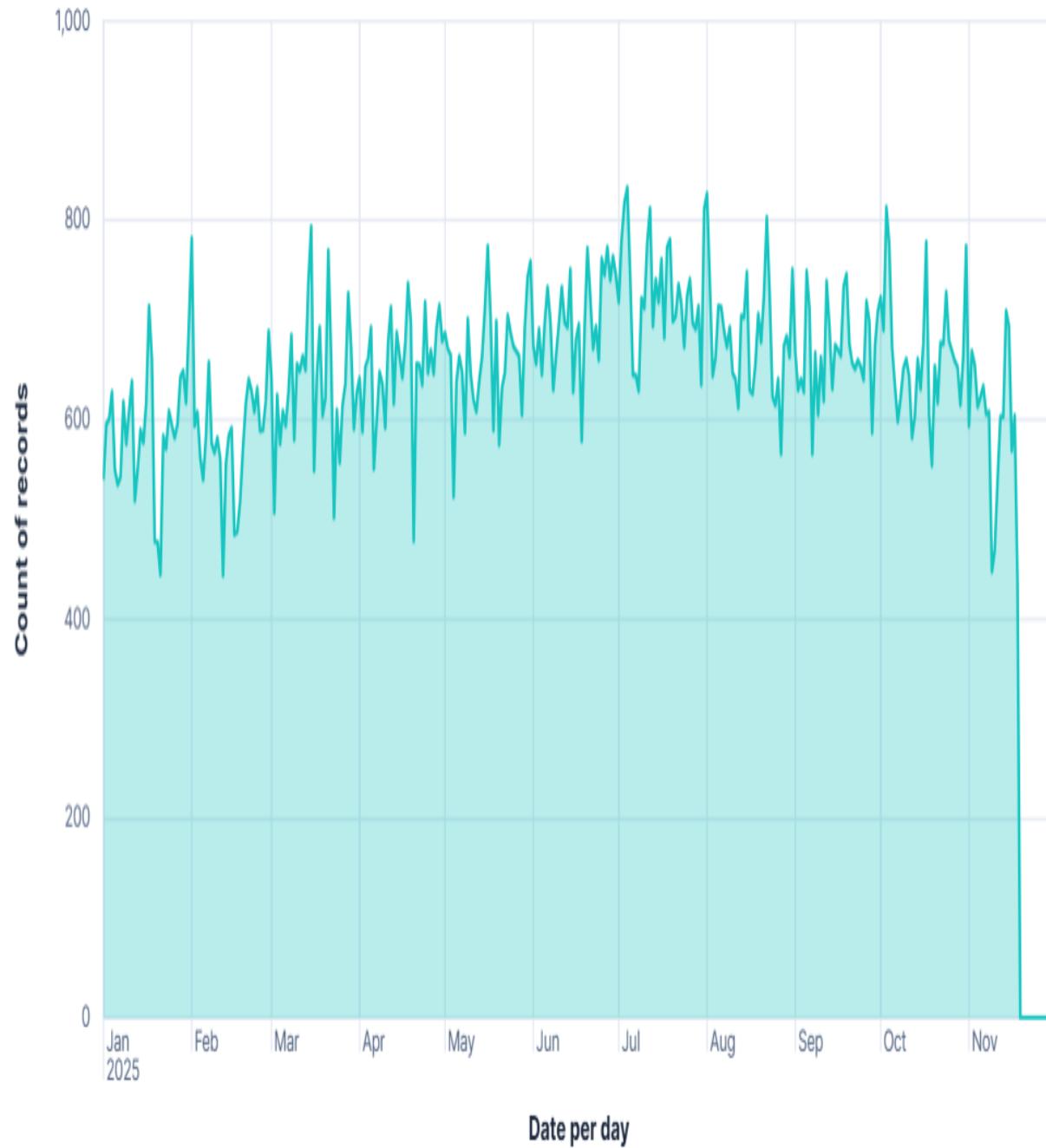
# DATA ANALYSIS USING ES

Using a pie chart, we can compare the top 10 crimes in Chicago from January 2025 to November 2025. The percentile gives us a clear comparison of the number of crimes committed.



# DATA ANALYSIS USING ES

Area chart to show  
the difference  
between the  
amount of crime  
per day from  
January 2025 to  
November 2025.



# PREDICTIVE ANALYSIS USING MACHINE LEARNING

Model evaluation Job status stopped docs evaluated 210866 Classification evaluation docs

Normalized confusion matrix for entire dataset

Predicted class		Normalized confusion matrix for entire dataset						
		ASSAULT	BATTERY	BURGLARY	CRIMINAL DAMAGE	DECEPTIVE PRACTICE	MOTOR VEHICLE THEFT	4 more
Actual class	ASSAULT	92%	6%	0%	0%	0%	0%	***
	BATTERY	40%	56%	0%	0%	0%	0%	***
	BURGLARY	0%	0%	98%	0%	0%	0%	***
	CRIMINAL DAMAGE	0%	0%	0%	100%	0%	0%	***
	DECEPTIVE PRACTICE	0%	0%	0%	0%	85%	0%	***
	MOTOR VEHICLE THEFT	0%	0%	0%	0%	0%	97%	***

Evaluation quality metrics

0.864 0.573 Overall accuracy Mean recall

Per class recall and accuracy

Receiver operating characteristic (ROC) curve

True Positive Rate (TPR) (a.k.a. Recall)

False Positive Rate (FPR)

Primary Type

- ASSAULT (AUC: 0.96245)
- BATTERY (AUC: 0.9600)
- BURGLARY (AUC: 0.99995)
- CRIMINAL DAMAGE (AUC: 0.99...)
- DECEPTIVE PRACTICE (AUC: 0.85...)
- MOTOR VEHICLE THEFT (AUC: 0.99995)
- NARCOTICS (AUC: 0.96056)
- OTHER OFFENSE (AUC: 0.99995)
- ROBBERY (AUC: 0.99995)
- THEFT (AUC: 0.99995)
- Baseline

# PREDICTIVE ANALYSIS USING MACHINE LEARNING

Evaluation quality metrics

0.864

Overall accuracy ⓘ

0.573

Mean recall ⓘ

Accuracy of the Crime Prediction and Total Feature Importance.

▼ Total feature importance

Total feature importance values indicate how significantly a field affects the predictions across all the training data.

ⓘ Feature importance docs



# Q&A

---

+

•

○

THANK YOU

