



**EARLY RISK
ASSESSMENT
AND LINKAGE
ANALYSIS OF
VIOLENT
CRIME CASES**

Name: CHAMU KOTTUR



CHICAGO

COLD CASE



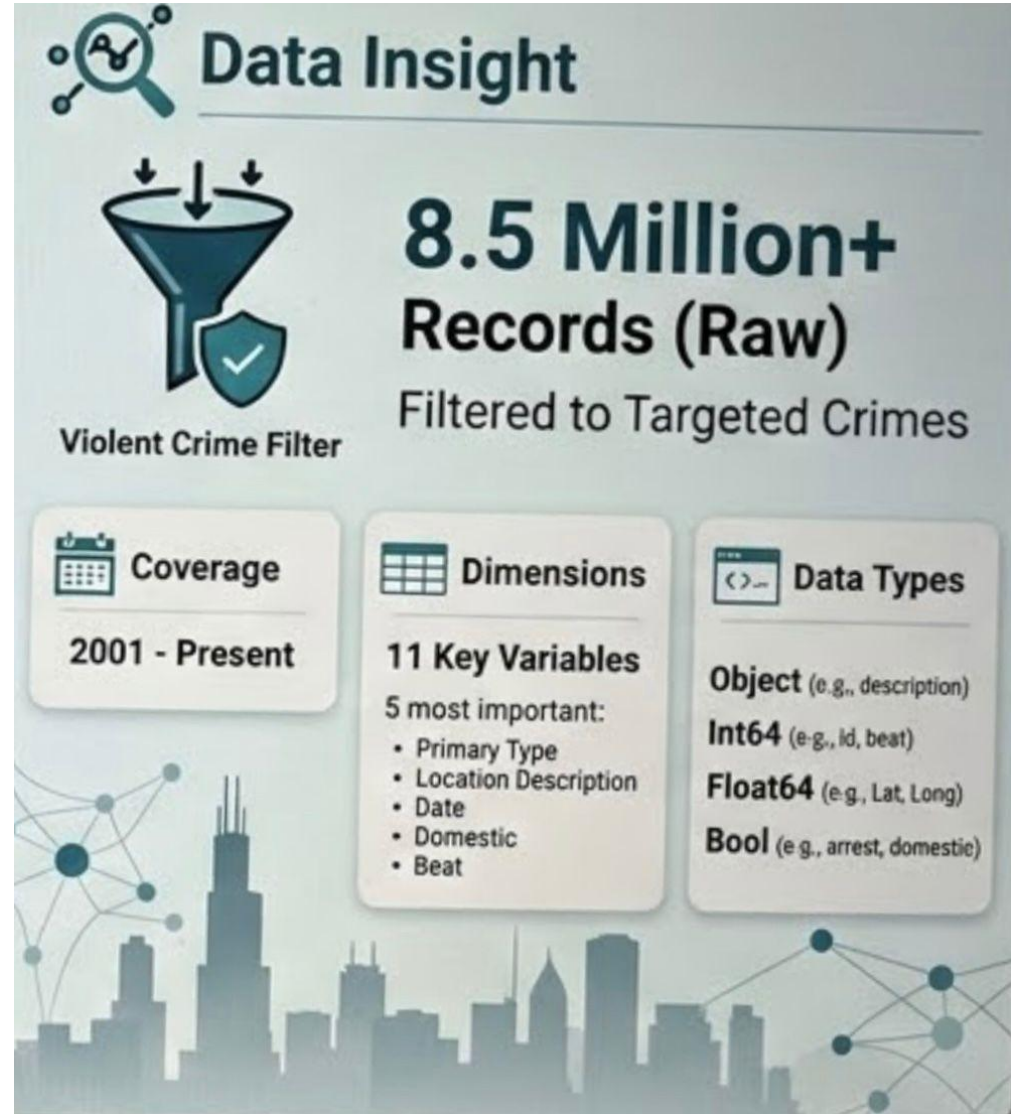
FOX 32
CHICAGO

5:42 46°

In order to address the complexity of violent crime, this project creates a case linking pipeline that turns publicly available data into useful insights and reclassifies possible cold cases.

DATA SET: 20 YEARS OF CHICAGO CRIME

- THE SCOPE :
 - 8.5+ MILLION ROWS
 - 11 VARIABLES
(TIME, LOCATION, TYPE...)
 - 2001-PRESENT



NARROWING THE SCOPE:VIOLENT CRIMES



NARROWING THE SCOPE:VIOLENT CRIMES

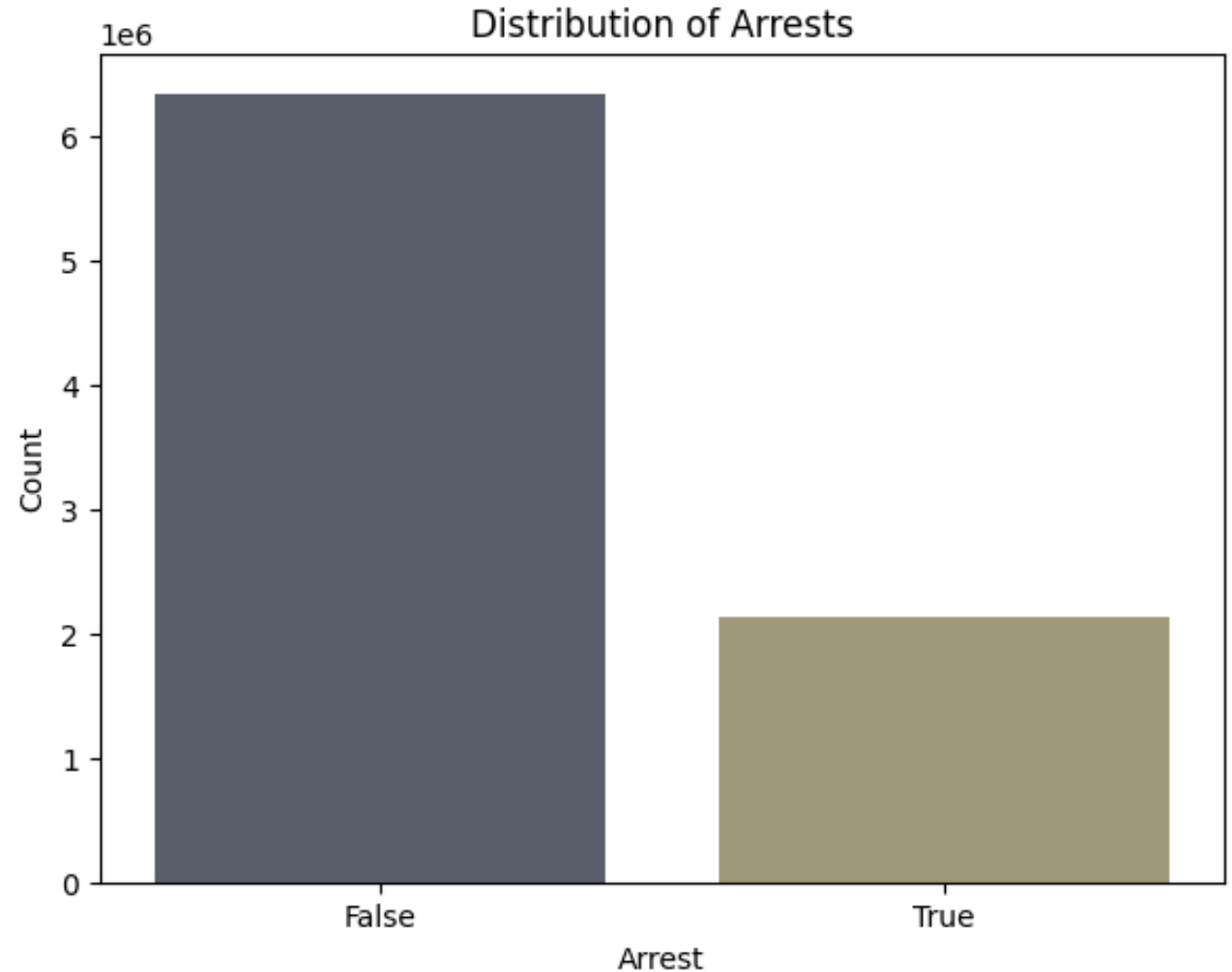
- Following the FBI standard, "violent crime" is defined as offenses involving force or the threat of force

Categories of Violent Crimes

- HOMICIDE
- ASSAULT
- BATTERY
- KIDNAPPING
- SEXUAL ASSAULT
- HUMAN TRAFFICKING
- ROBBERY
- ETC.,

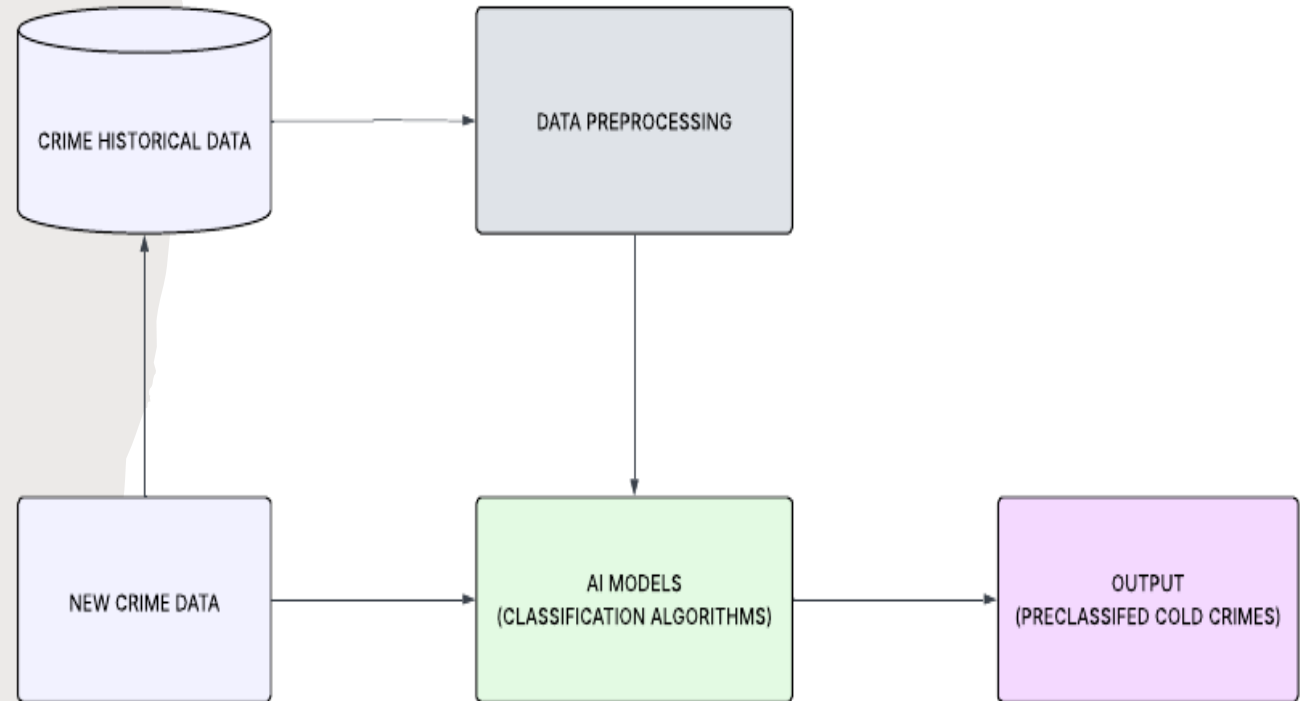
COLD CASE RISK ASSESSMENT

- **DEFINITION:** A CRIMINAL INVESTIGATION THAT REMAINS UNSOLVED AFTER 12 MONTHS OR MORE
- **WHY THIS MATTERS:**
 - ALLOCATION OF RESOURCES
 - PUBLIC SAFETY
 - JUSTICE AND CLOSURE FOR VICTIMS



METHODOLOGY

- **CLASSIFICATION MODELS:**
 - LOGISTIC REGRESSION
 - RANDOM FOREST
 - HIST GRADIENT BOOSTING
- **TARGET VARIABLE (COLD_CASE):**
 - 1 IF ARREST == FALSE AND OLDER THAN 12 MONTHS ,
 - 0 IF ARREST == TRUE
- **CLASS IMBALANCE OPTIMIZATION:**
 - classification threshold ($T = 0.35$)





Target Case
Profile:

```
target_case = {  
    'id': 10000000,  
    'date': '2026-02-11 00:00:00',  
    'primary_type': 'OFFENSE INVOLVING CHILDREN',  
    'description': 'CHILD ABDUCTION',  
    'location_description': 'ALLEY',  
    'latitude': 41.794173,  
    'longitude': -87.703576,  
    'arrest': False,  
    'domestic': False,  
    'beat': 923,  
    'district': 9  
}
```

FINDINGS

- **TARGET CASE INFERENCE:**

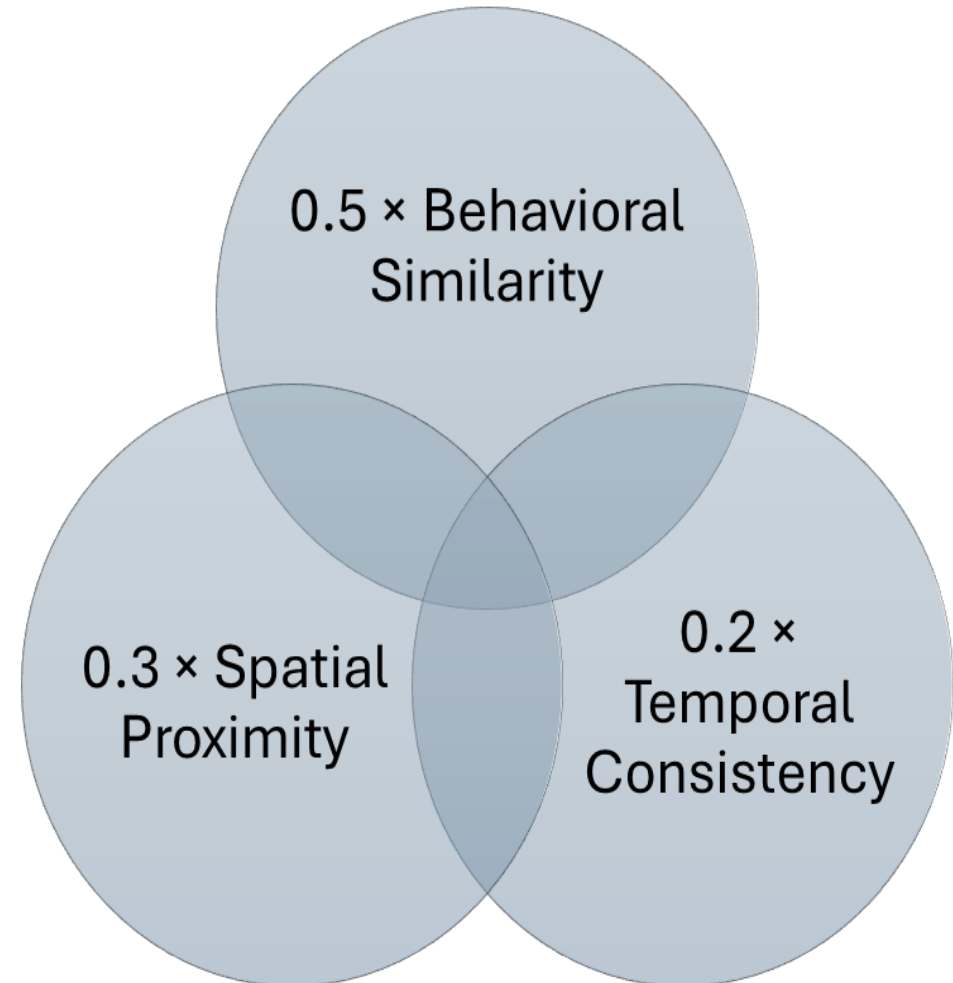
Successfully classified a high-probability case with

- $P(\{\text{Cold Case}\}) = 0.877$,
- cold_case_prediction: 1,
- threshold: 0.35

MODEL	PR-AUC	PRECISION	RECALL
HISTGRADIENT BOOSTING	0.7961	0.7417	0.9801
RANDOM FOREST	0.7933	0.7463	0.9344
LOGISTIC REGRESSION	0.7794	0.7579	0.8574

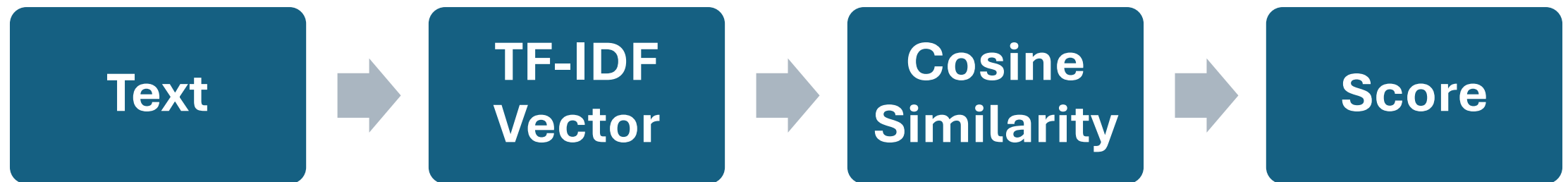
Case Linkage Analysis

- **DEFINITION:** IDENTIFYING POTENTIAL CONNECTIONS BETWEEN CRIMES.
- **WHY IT HELPS:**
 - IDENTIFIES CRIME PATTERNS
 - HELPS NARROW SUSPECTS
 - LINKS CASES ACROSS JURISDICTIONS



Formula

- **Linkage Score** = $(0.5 \times \text{Behavioral Similarity}) + (0.3 \times \text{Spatial Proximity}) + (0.2 \times \text{Temporal Consistency})$
- **Behavioral Similarity** : I combined the 'description' and 'location_description' columns into a single 'behavioral_signature' text for both the target case and historical cases then used Tf-idf and then calculate the cosine_similarity to derive a 'Behavioral Similarity' score.

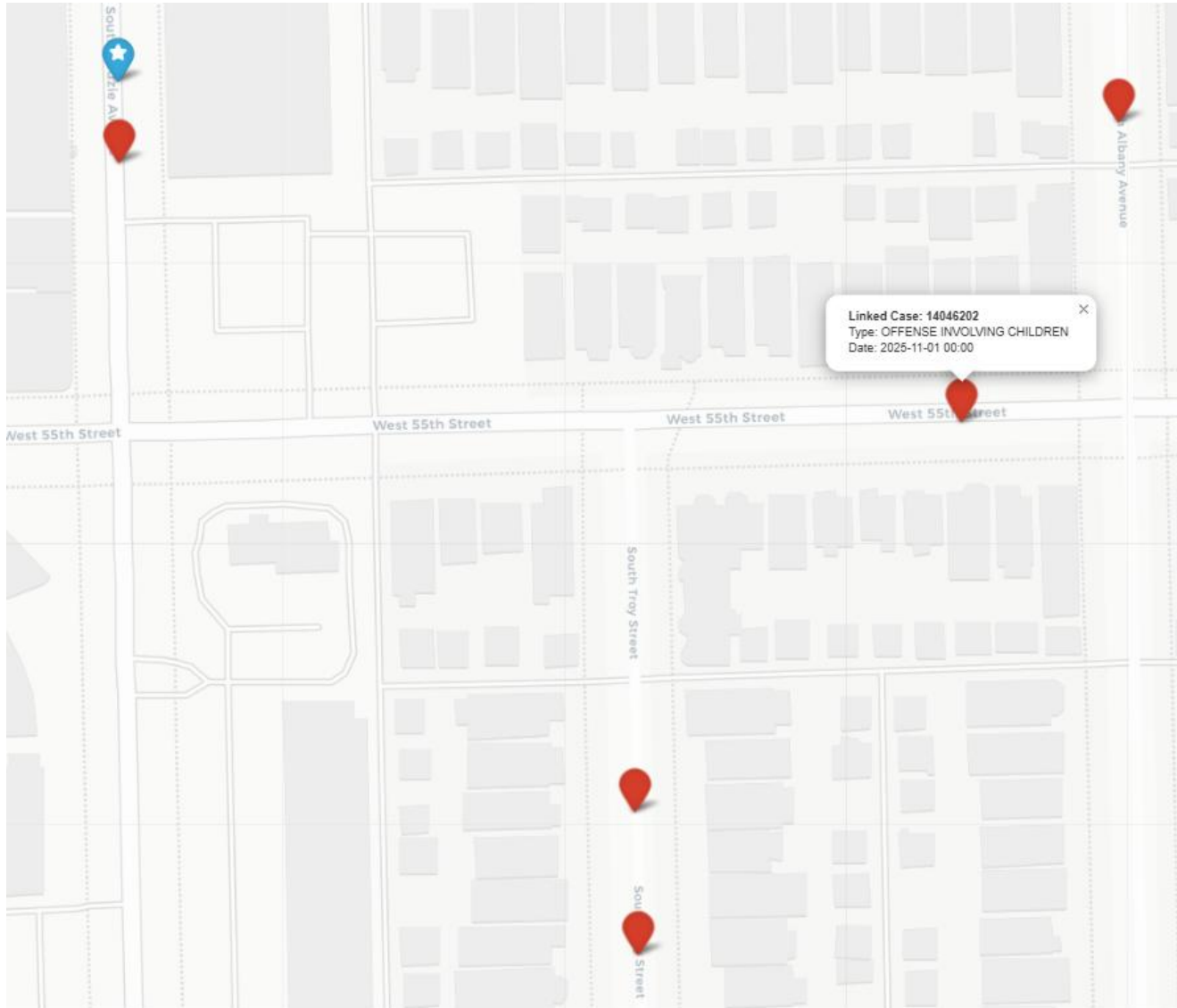




Target Case
Profile:

```
target_case = {  
    'id': 10000000,  
    'date': '2026-02-11 00:00:00',  
    'primary_type': 'OFFENSE INVOLVING CHILDREN',  
    'description': 'CHILD ABDUCTION',  
    'location_description': 'ALLEY',  
    'latitude': 41.794173,  
    'longitude': -87.703576,  
    'arrest': False,  
    'domestic': False,  
    'beat': 923,  
    'district': 9  
}
```

LINKAGE VISUALIZATION



- Rendered via Folium
- **CartoDB Positron** tileset

Case ID	Relation	Primary Type	Date
10000000	Target	OFFENSE INVOLVING CHILDREN	2026-02-11
13552396	Linked	BATTERY	2024-08-04
12986119	Linked	OFFENSE INVOLVING CHILDREN	2023-01-01
14046202	Linked	OFFENSE INVOLVING CHILDREN	2025-11-01

Linkage Visualization

