Chicago Crime Analytics

A data-driven deep dive into the patterns, hotspots, and trends shaping crime in the city.

This infographic synthesizes key findings to inform strategic public safety initiatives.

Presented by

1) Opejin, Hammed
2) Diaz Mariduena, Humberto
3) Oncul, Cagatay
4) Li, Xiao

06/2025

Data-Driven Recommendations

01

Prioritize High-Impact Zones

Deploy a greater share of police and community resources to the Austin (Area 25) community, and enhance patrols against theft in the Near North Side (Area 8) and The Loop (Area 32).

02

Develop Targeted Strategies

Establish special task forces to design and implement dedicated prevention and enforcement strategies for Theft and Battery, the two most prevalent crime types.

03

Implement Dynamic Patrols

Use the crime heatmaps to guide daily patrol routes for maximum coverage of high-risk zones and increase presence during known crime peaks (e.g., holidays).

Integrate crime heatmap and cross-analysis findings into the daily police dispatch system. This will shift the policing model from reactive response to proactive prevention.

THE THEORY OF STATISTICS

Method 1: Frequency Analysis (Descriptive Statistics)

Used to answer the basic but crucial questions: **"What?"** and **"Where?"**. The core idea is simply counting occurrences for each category.

Formula: $f_i = \text{Count of occurrences for category } i$

What are the top crimes?	Where are the most dangerous areas?
As the chart shows, **THEFT** and **BATTERY** are by far the most frequent crimes in Chicago.	**Austin (Community Area 25)** "wins" by a large margin, making it the undisputed crime capital.

When single-dimension analysis isn't enough, we use cross-tabulation to uncover deeper relationships between **"specific places"** and **"specific crimes."** A heatmap is the most intuitive visualization of a cross-tabulation table.

Insight from the Heatmap: This chart clearly shows us:

- **THEFT** is rampant in **Area 8 (Near North Side)** and **Area 32 (The Loop)**, which aligns with their status as commercial hubs.
- **Austin (Area 25)** is not only high in overall crime but particularly severe in **BATTERY** and **CRIMINAL DAMAGE**.

Method 3: Kernel Density Estimation (KDE)

To identify geographical clusters of criminal activity from a macro perspective, we employ Kernel Density Estimation (KDE). This powerful statistical method transforms discrete crime location points into a smooth, continuous "hotspot" map.

The Core Mathematical Idea:

- **2D KDE Formula:** $\hat{f}(x,y) = \frac{1}{nh^2}\sum_{i=1}^{n} K\left(\frac{(x-x_i)}{h}, \frac{(y-y_i)}{h}\right)$
- **Gaussian Kernel:** $K(u,v) = \frac{1}{2\pi}e^{-\frac{u^2+v^2}{2}}$
  - *In simple terms, we place a "sphere of influence" around each crime point. The influence is strongest at the center and fades with distance. By summing up all these spheres, we generate the final heatmap.*

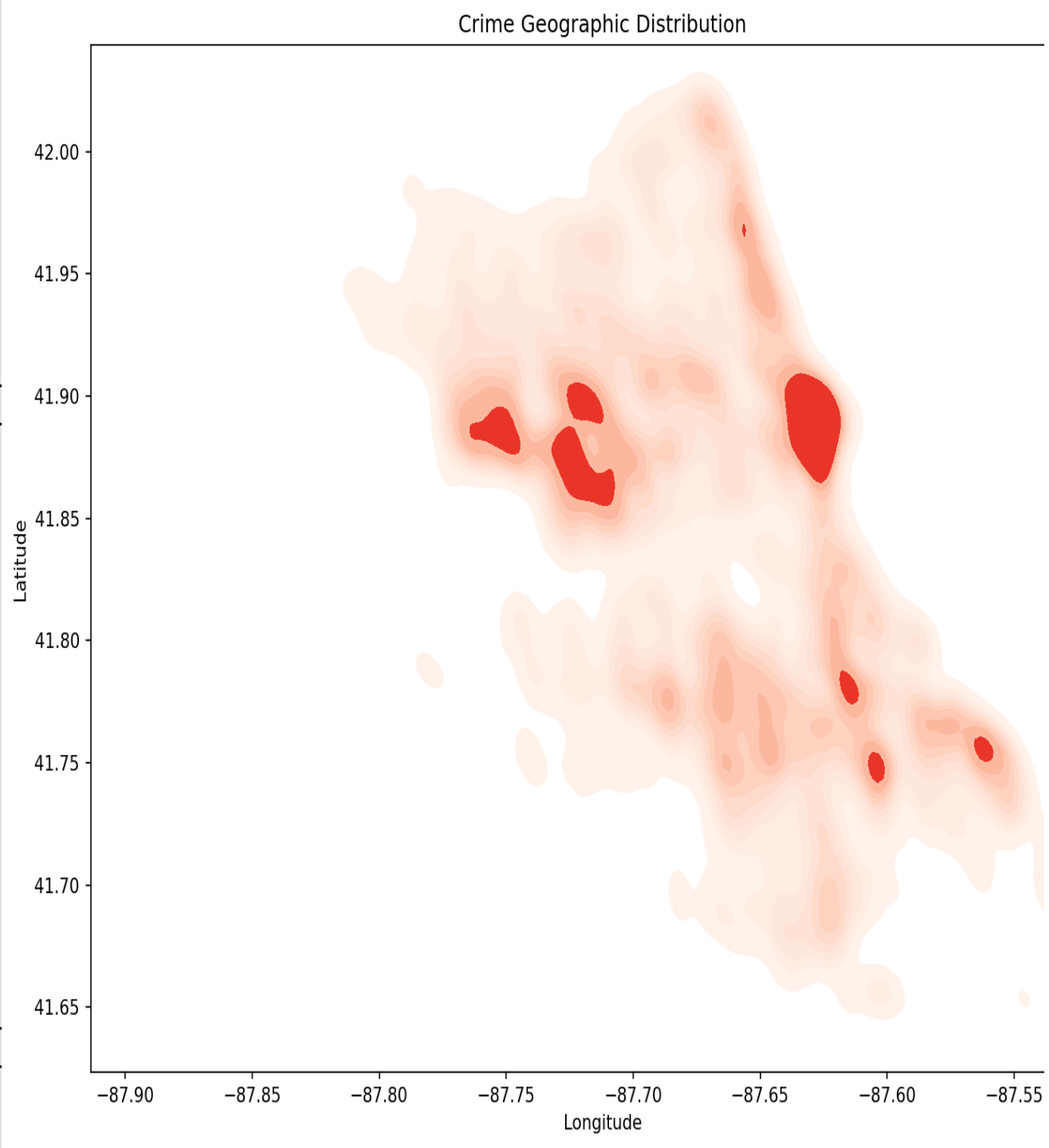
CODES PART

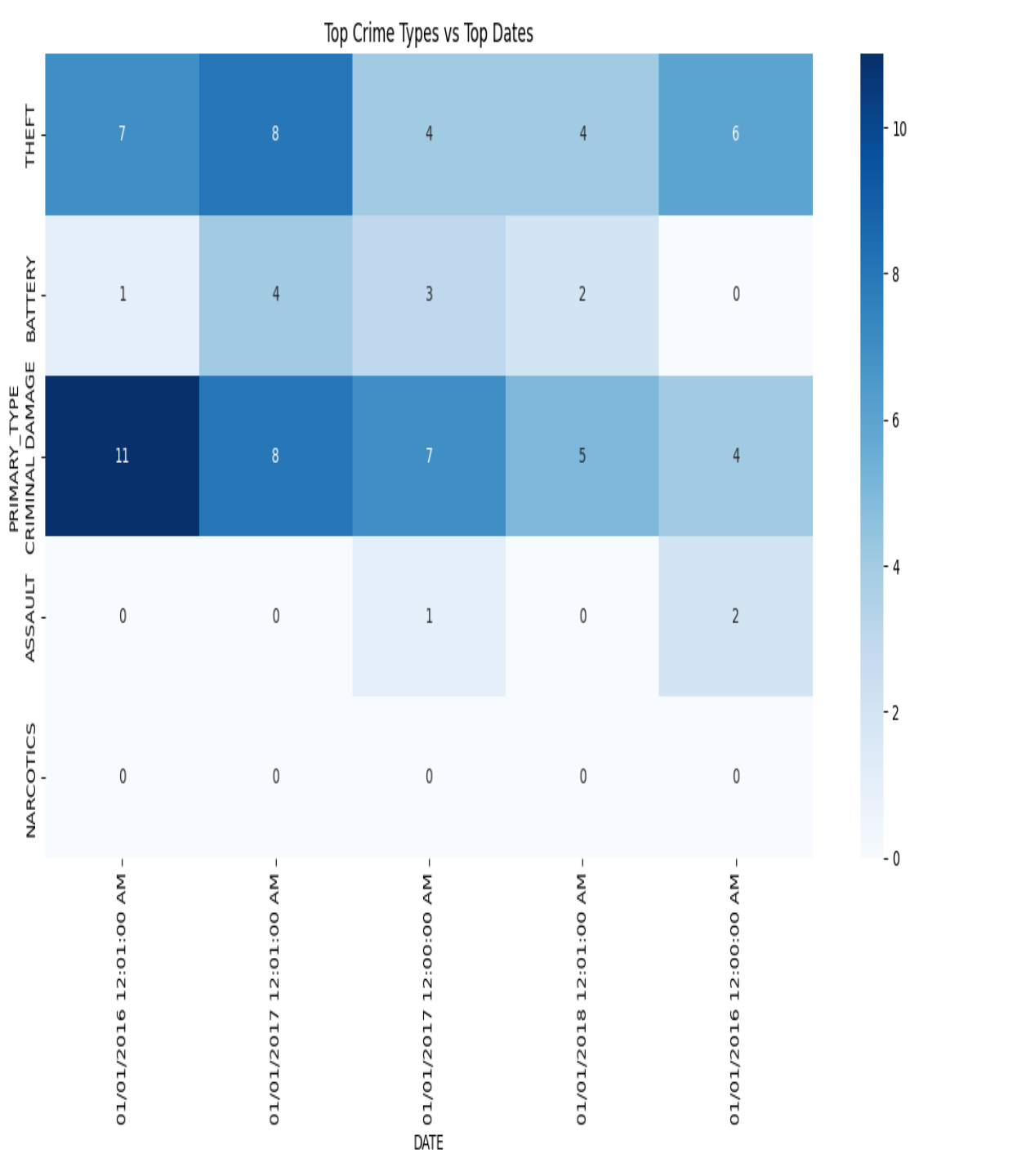
File: chicago_crime_analysis.py Page 1 of 5

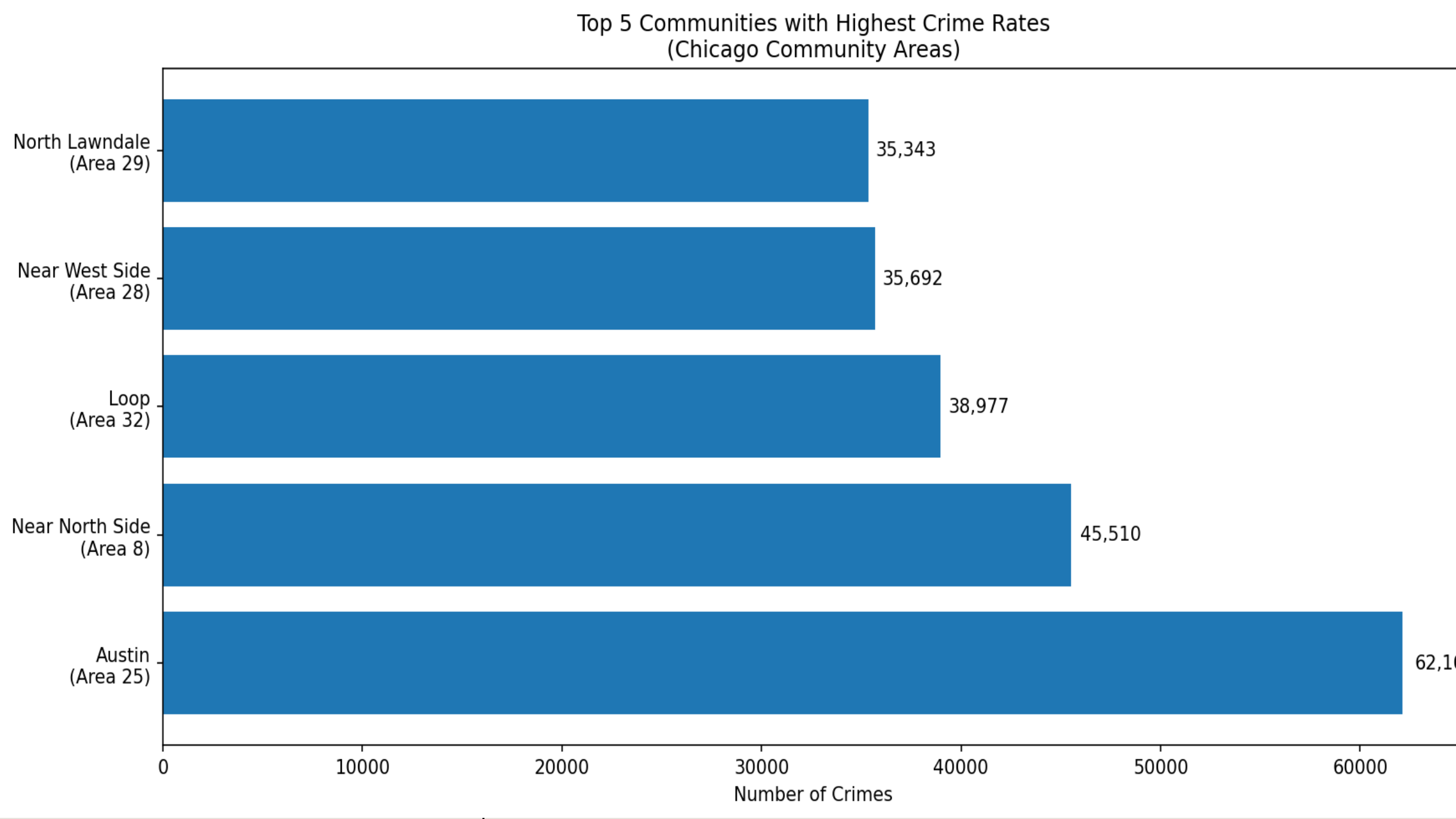
```python
import snowflake.connector
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import os

# 1. Connect Snowflake
conn = snowflake.connector.connect(
    user='lixiaowww',
    password='Sean0105@winnipeg!',
    account='gzhlbee-bv06447',
    warehouse='warehouse',
    database='CHICAGO_CRIME',
    schema='STATISTICS'
)

# 2. Analysis the dataset
sql = "SELECT * FROM CHICAGO_CRIME_COPY"
df = pd.read_sql(sql, conn)
conn.close()

print(f"Data loaded successfully! Total records: {len(df):,}")
print(f"Columns: {list(df.columns)}")

# 3. Top Crime Type
if 'PRIMARY_TYPE' in df.columns:
    top5_types = df['PRIMARY_TYPE'].value_counts().head(5)
    plt.figure(figsize=(8,5))
    sns.barplot(x=top5_types.values, y=top5_types.index)
    plt.title('Top 5 Crime Types')
    plt.xlabel('Count')
    plt.tight_layout()
    plt.savefig('top5_crime_types.png')
    plt.close()
    print("Top 5 crime types chart created")
else:
    print("PRIMARY_TYPE column not found")
    top5_types = pd.Series()

# 4. Top 5 criminal communities
# Chicago Community Area mapping
community_areas = {
    1: "Rogers Park", 2: "West Ridge", 3: "Uptown", 4: "Lincoln Square", 5: "North Center",
    6: "Lake View", 7: "Lincoln Park", 8: "Near North Side", 9: "Edison Park", 10: "Norwood
Park",
    11: "Jefferson Park", 12: "Forest Glen", 13: "North Park", 14: "Albany Park", 15: "Portage
Park",
    16: "Irving Park", 17: "Dunning", 18: "Montclare", 19: "Belmont Cragin", 20: "Hermosa",
    21: "Avondale", 22: "Logan Square", 23: "Humboldt Park", 24: "West Town", 25: "Austin",
    26: "West Garfield Park", 27: "East Garfield Park", 28: "Near West Side", 29: "North
Lawndale", 30: "South Lawndale",
    31: "Lower West Side", 32: "Loop", 33: "Near South Side", 34: "Armour Square", 35:
"Douglas",
    36: "Oakland", 37: "Fuller Park", 38: "Grand Boulevard", 39: "Kenwood", 40: "Washington
Park",
    41: "Hyde Park", 42: "Woodlawn", 43: "South Shore", 44: "Chatham", 45: "Avalon Park",
    46: "South Chicago", 47: "Burnside", 48: "Calumet Heights", 49: "Roseland", 50: "Pullman",
    51: "South Deering", 52: "East Side", 53: "West Pullman", 54: "Riverdale", 55: "Hegewisch",
    56: "Garfield Ridge", 57: "Archer Heights", 58: "Brighton Park", 59: "McKinley Park", 60:
"Bridgeport",
```

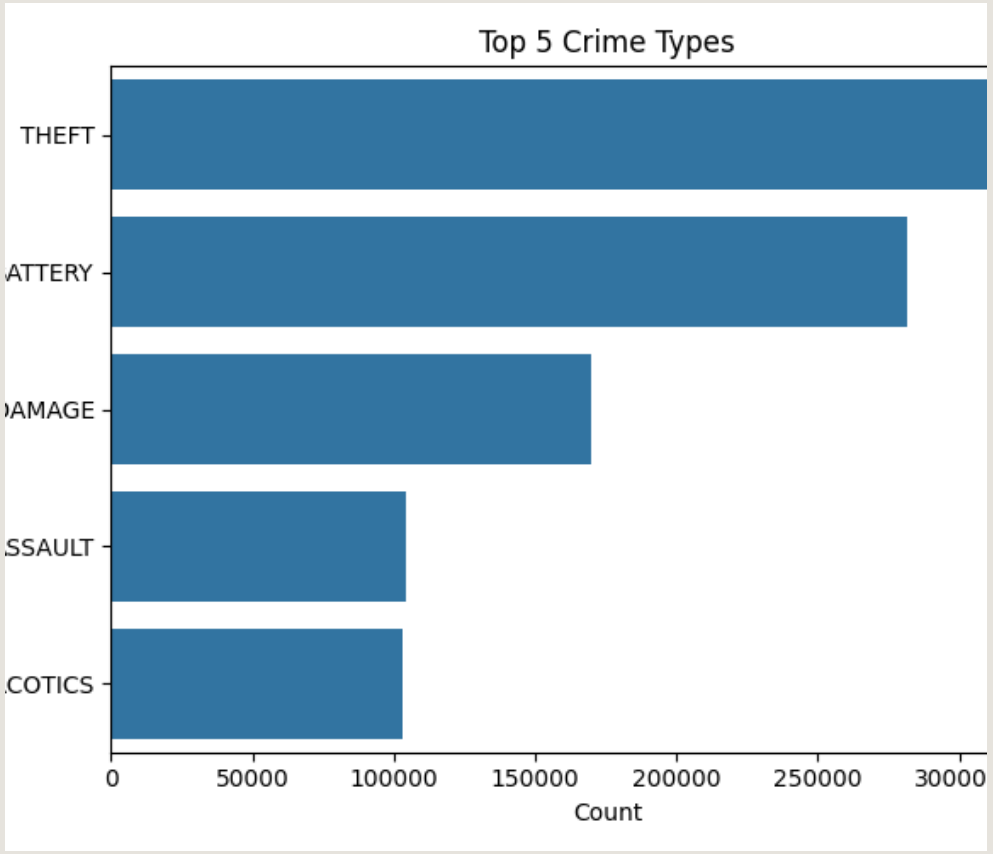
```
chicago_crime/
├── data/               # Contains raw data
├── src/                # Source code
│   └── main.py         # Main script
├── notebooks/          # Jupyter notebooks
├── results/            # Analysis results and charts
├── docs/               # Documentation
└── requirements.txt    # Project dependencies
```

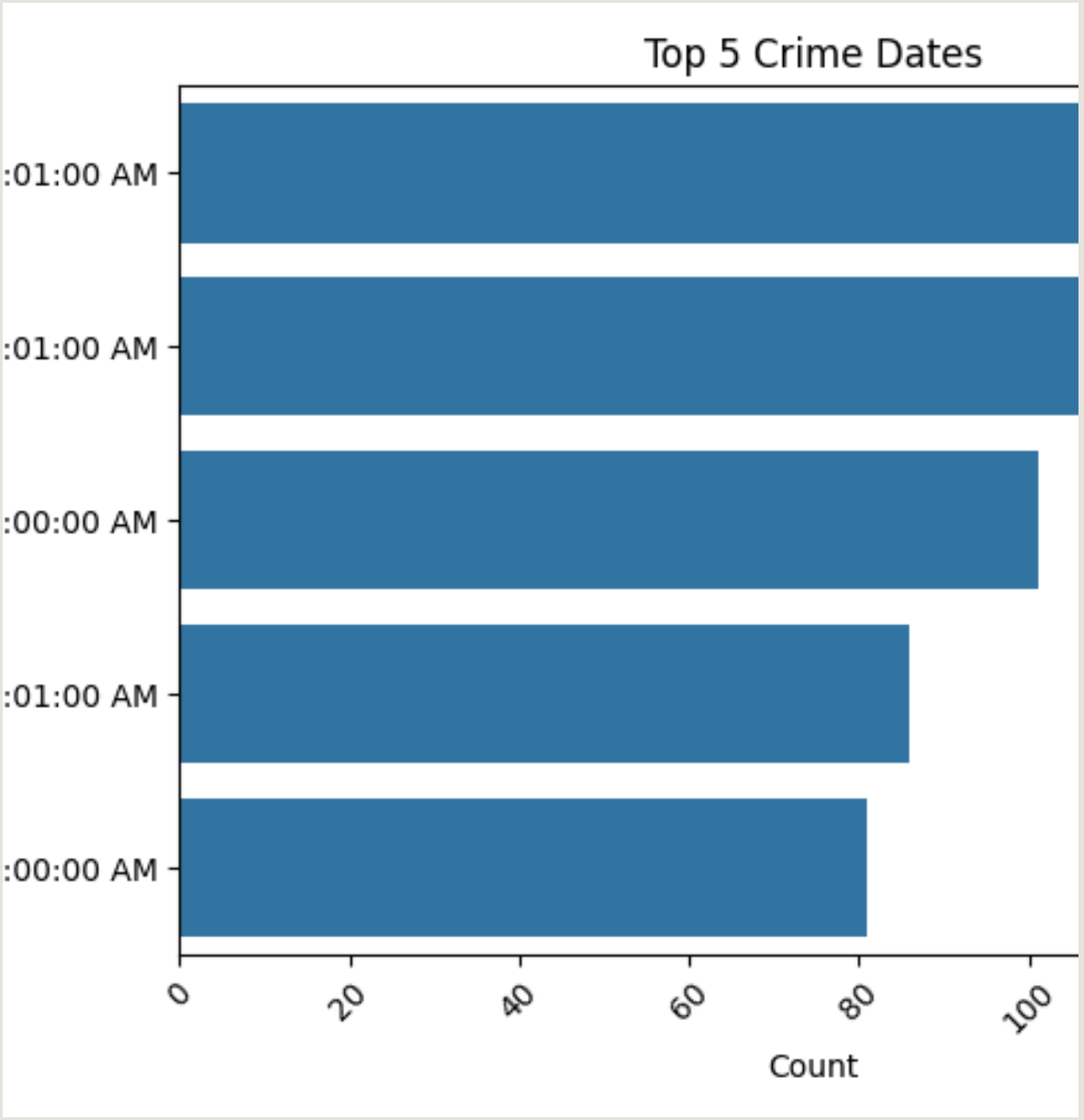
```sql
-- Create file format for CSV
CREATE OR REPLACE FILE FORMAT csv_format
TYPE = 'CSV'
FIELD_DELIMITER = ','
SKIP_HEADER = 1
FIELD_OPTIONALLY_ENCLOSED_BY = '"'
ESCAPE_UNENCLOSED_FIELD = '\';


-- Upload chicago_crime_part_001.csv
-- Step 1: Upload file to stage
PUT 'file://C:/AS/chicago_crime/data/split_files/chicago_crime_part_001.csv' @~/staged;

-- Step 2: Copy data from stage to table
COPY INTO CHICAGO_CRIME
FROM @~/staged/chicago_crime_part_001.csv
FILE_FORMAT = csv_format
```


Crime Geographic Distribution
Latitude
Longitude
42.00
41.95
41.90
41.85
41.80
41.75
41.70
41.65
−87.90
−87.85
−87.80
−87.75
−87.70
−87.65
−87.60
−87.55


Top Crime Types vs Top Dates
PRIMARY_TYPE
THEFT
BATTERY
CRIMINAL DAMAGE
ASSAULT
NARCOTICS
7
8
4
4
6
1
4
3
2
0
11
8
7
5
4
0
0
1
0
2
0
0
0
0
0
01/01/2016 12:01:00 AM
01/01/2017 12:01:00 AM
01/01/2017 12:00:00 AM
01/01/2018 12:01:00 AM
01/01/2016 12:00:00 AM
DATE
10
8
6
4
2
0


Top 5 Communities with Highest Crime Rates
(Chicago Community Areas)
North Lawndale
(Area 29)
35,343
Near West Side
(Area 28)
35,692
Loop
(Area 32)
38,977
Near North Side
(Area 8)
45,510
Austin
(Area 25)
0
10000
20000
30000
40000
50000
60000
Number of Crimes


Top 5 Crime Types
THEFT
ATTERY
AMAGE
SSAULT
COTICS
0
50000
100000
150000
200000
250000
30000
Count


Top 5 Crime Dates
:01:00 AM
:01:00 AM
:00:00 AM
:01:00 AM
:00:00 AM
0
20
40
60
80
100
Count

Crime Landscape Overview

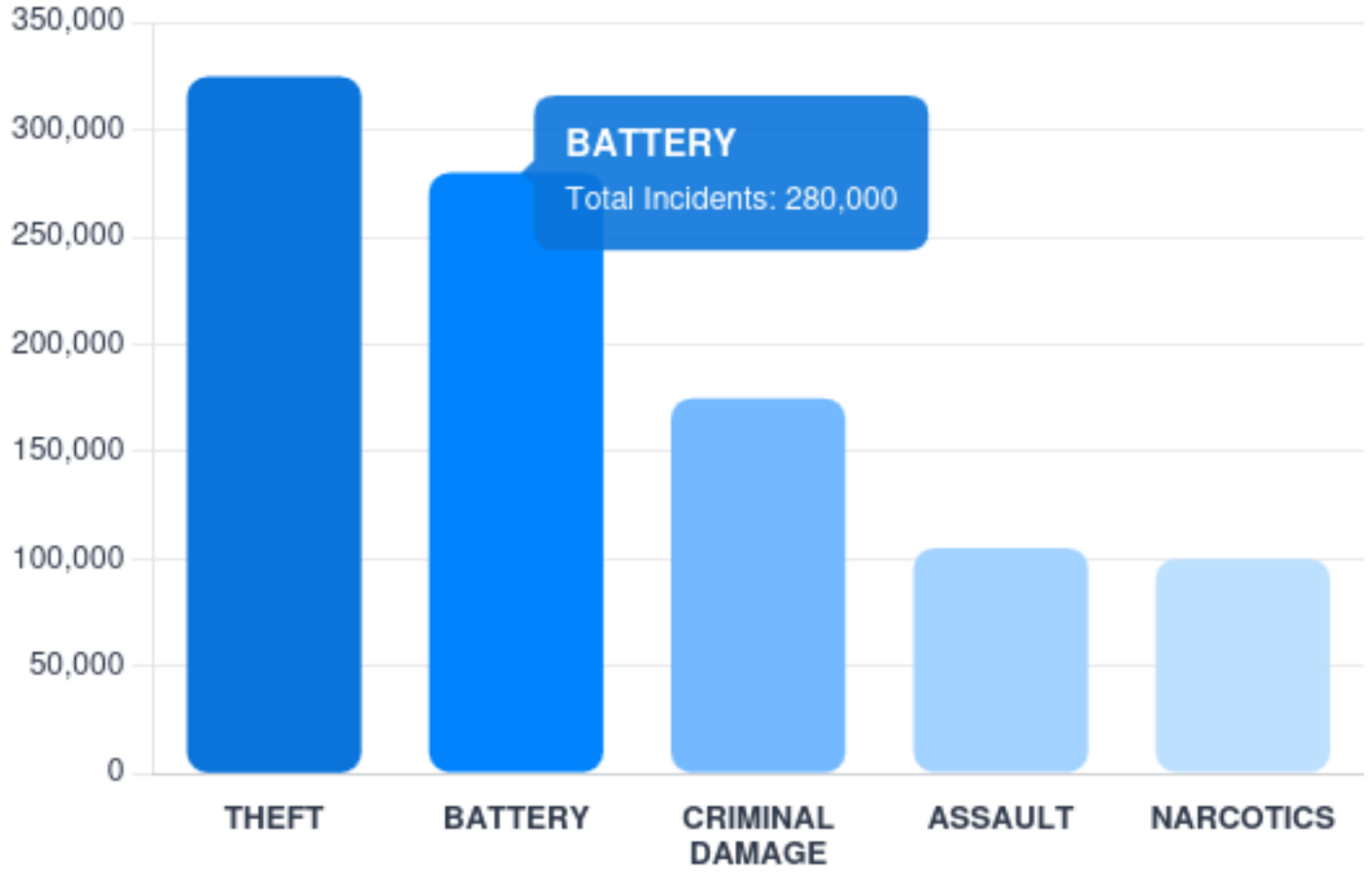
To understand the scope of Chicago's public safety challenges, we first look at the most prevalent types of crime. The data reveals that a few specific categories account for a vast majority of all reported incidents. This initial analysis helps prioritize resources towards the most significant and frequent threats to the community.

The Two Leading Crimes

Theft and Battery are the most reported crimes, forming the primary public safety concern city-wide.

1st THEFT

2nd BATTERY

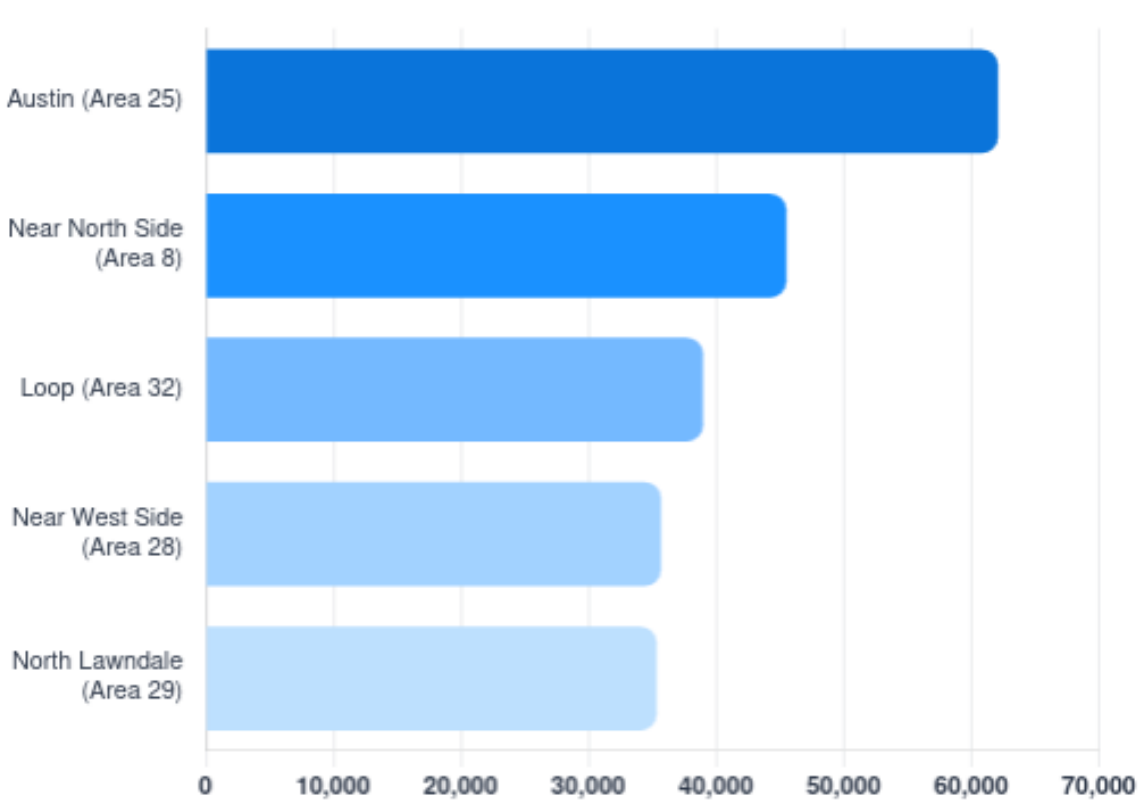


Geographical Hotspots: Where Crime Concentrates

Crime is not uniformly distributed. Our analysis pinpoints specific community areas that bear a disproportionate burden of criminal activity. Identifying these hotspots is the first step toward targeted, location-based interventions and resource allocation.

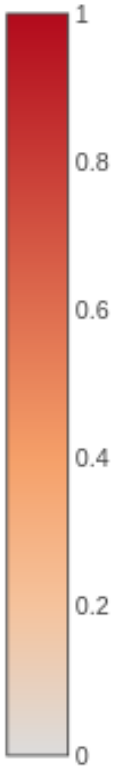
Top 5 High-Crime Community Areas

The data clearly shows that Austin (Area 25) experiences a significantly higher volume of crime than any other community, making it a critical area of focus.



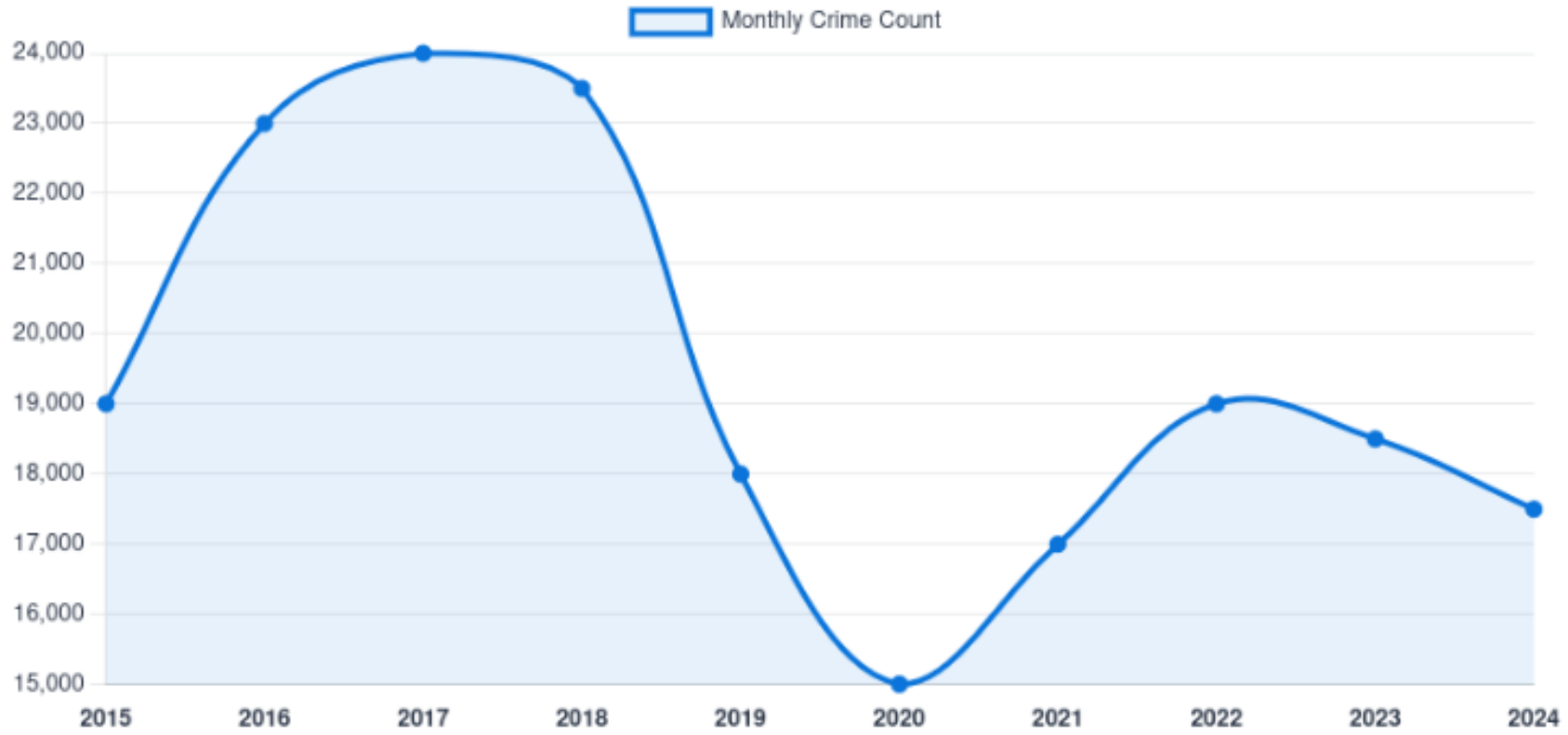
Crime Density Map

This geographic heatmap visualizes crime concentration across the city. The intense red areas represent "hotspots" where incidents are most densely clustered.



Temporal Patterns: When Crime Occurs

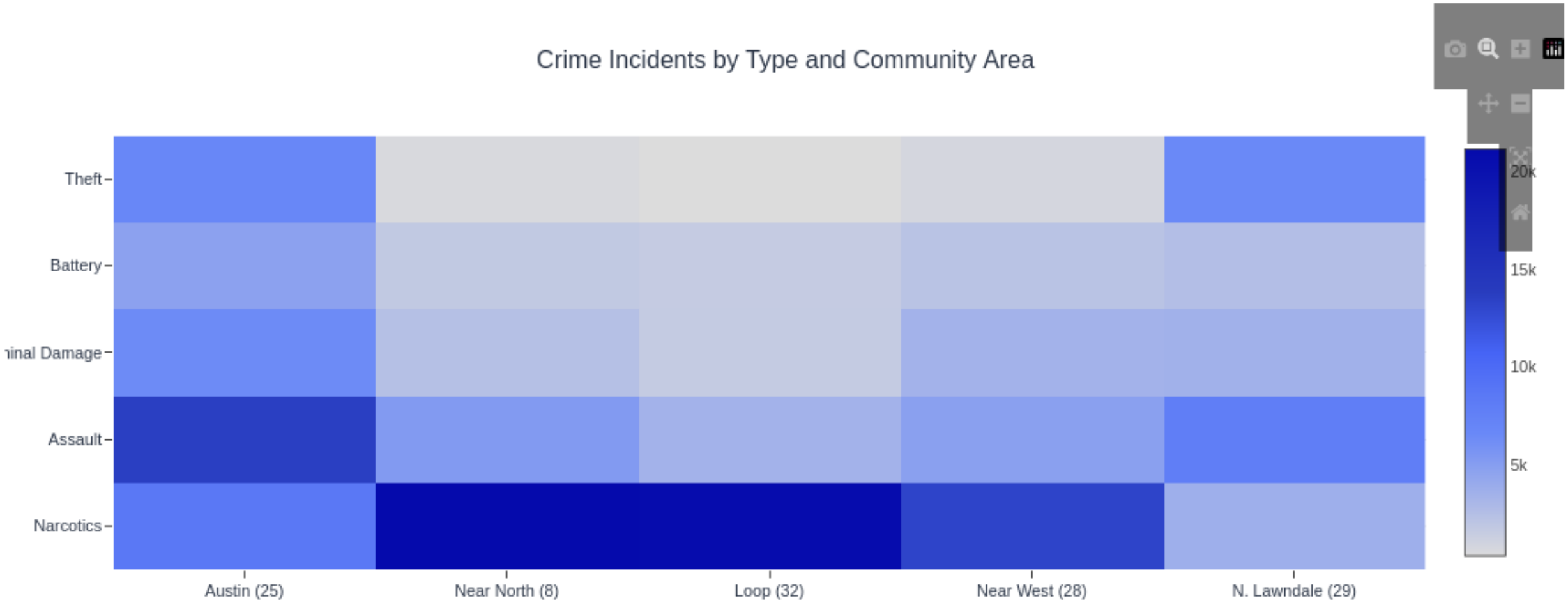
Understanding the rhythm of crime is as important as knowing its location. This section explores long-term trends and specific high-risk dates, revealing cyclical patterns and anomalies that can inform patrol scheduling and public awareness campaigns.



The line chart above illustrates the fluctuation in total reported crimes per month, highlighting periods of significant increase or decrease in criminal activity over the years.

Cross-Analysis: Connecting Crime Type to Location

Deeper insights emerge when we analyze the relationship between *what* crime occurs and *where*. This heatmap reveals that different communities face unique challenges, requiring tailored, rather than one-size-fits-all, policing strategies.



CONCLUTION

Based on the data analysis, we propose the following recommendations:

- **Focus on High-Risk Areas:** Prioritize the deployment of more police and resources to the Austin (25) community, and also focus on other high-crime-rate areas such as the Near North Side and the Loop.
- **Target Major Crime Types:** Develop specialized enforcement and prevention plans for high-incidence crimes such as theft and battery.
- **Strengthen Patrols at Specific Times:** Significantly increase police deployment and street patrols on identified high-crime dates, such as New Year's Day.
- **Implement Precision Prevention:** Utilize crime heatmaps and cross-analysis results to implement targeted prevention strategies for specific communities and crime types to improve efficiency.