

Chicago Reported Incidents Dataset

CIS 4560-01 | Group 3

Jerry Petty, Mayowa Toyinbo, Erik Pacheco, Francisco Figueroa, Eric Ventura, Gerardo Camarena



Data Description/Source

Data Name: Chicago Crime (01JAN2001-22JUL2020)

Source: Chicago P.D.

Size: 1.7GB **Format:** CSV

Dataset URL:

<https://www.kaggle.com/datasets/n3v375/chicago-crime-from-01jan2001-to-22jul2020>

Github URL:

<https://github.com/FranXavier12/Tempo-Spatial-Analysis-of-Chicago-Crime-Data>

Description: This data records incidents of crime recognized by the Chicago Police Department. It includes location, time, and the categorial reasoning for each incident.



Specifications

Cluster Version:
Hadoop Cluster 3.3.3

Number of Nodes: 3
nodes (1 Master, 2
Worker)

Memory Size:

Master Nodes: 51.2GB
RAM

Worker Nodes: 58GB
RAM each

CPU Speed: 1995.312
MHz

```
-bash-4.2$ hadoop version
Hadoop 3.3.3
Source code repository ssh://git@bitbucket.oc1.oraclecorp.com:7999/bdcs/apache_bigtop.git -r 2a01e3e243c3d99d916ecdeb76935bc93eba3db0
Compiled by root on 2023-10-05T07:47Z
Compiled with protoc 3.7.1
From source with checksum d3c98190fee32556b772294379cddb5d
This command was run using /usr/odh/2.0.7/hadoop/hadoop-common-3.3.3.jar
-bash-4.2$ yarn node -list
24/05/03 01:53:00 INFO client.DefaultNoHARMAFailoverProxyProvider: Connecting to ResourceManager at bigdaiwn0.sub03291929060.trainingvcn.o
24/05/03 01:53:01 INFO client.AHSPProxy: Connecting to Application History server at bigdaiwn0.sub03291929060.trainingvcn.oraclevcn.com/10
Total Nodes:3
Node-Id Node-State Node-Http-Address Number-of-Running-Containers
bigdaiwn2.sub03291929060.trainingvcn.oraclevcn.com:45454 RUNNING bigdaiwn2.sub03291929060.trainingvcn.oraclevcn.com:8042
bigdaiwn0.sub03291929060.trainingvcn.oraclevcn.com:45454 RUNNING bigdaiwn0.sub03291929060.trainingvcn.oraclevcn.com:8042
bigdaiwn1.sub03291929060.trainingvcn.oraclevcn.com:45454 RUNNING bigdaiwn1.sub03291929060.trainingvcn.oraclevcn.com:8042
-bash-4.2$ hdfs getconf -confKey yarn.nodemanager.resource.memory-mb
51200
-bash-4.2$ free -h
              total          used          free      shared  buff/cache   available
Mem:           58G           30G           15G           2.9G           12G           24G
Swap:           8.0G           4.8M           8.0G
-bash-4.2$ lscpu
Architecture:          x86_64
CPU op-mode(s):        32-bit, 64-bit
Byte Order:            Little Endian
CPU(s):                8
On-line CPU(s) list:   0-7
Thread(s) per core:    2
Core(s) per socket:    4
Socket(s):             1
NUMA node(s):          1
Vendor ID:             GenuineIntel
CPU family:            6
Model:                 85
Model name:            Intel(R) Xeon(R) Platinum 8167M CPU @ 2.00GHz
Stepping:              4
CPU MHz:               1995.312
BogoMIPS:              3990.62
Virtualization:        VT-x
Hypervisor vendor:     KVM
Virtualization type:   full
L1d cache:            32K
L1i cache:            32K
L2 cache:              4096K
L3 cache:              16384K
NUMA node0 CPU(s):    0-7
```

Our cluster is running Hadoop version 3.3.3, consisting of 3 nodes (1 Master and 2 Worker), with the master node having 51.2GB of RAM and each worker node having 58GB of RAM. The CPU speed is approximately 2.00 GHz.

Top 15 Crime Incident Types

This table provides insights into the top 15 crime types reported in the Chicago Police Department dataset.

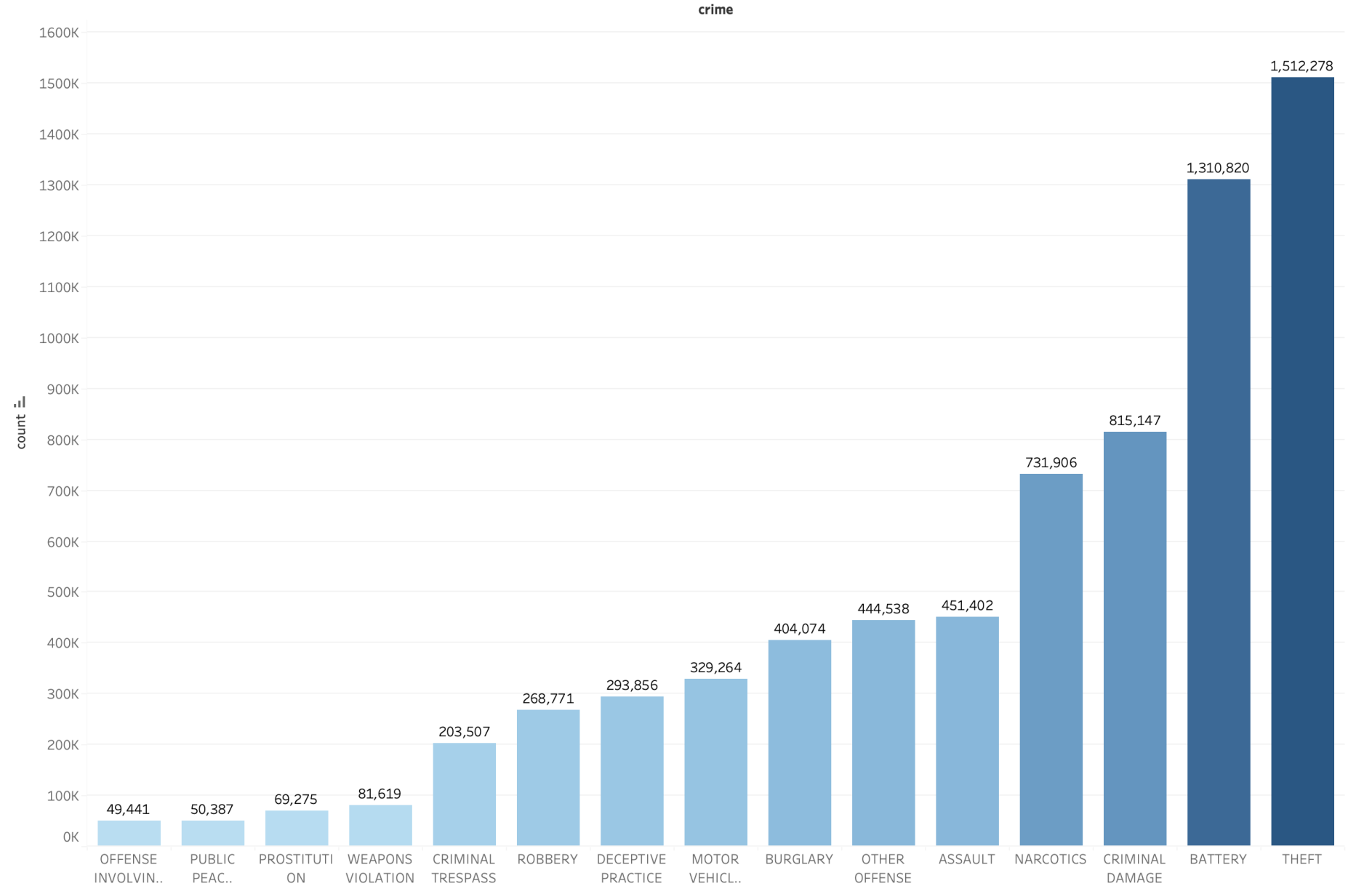
Analyzing the distribution of crime types helps identify prevalent issues in the city.

Sheet 1

crime	
OFFENSE INVOLVING CHILDREN	49,441
PUBLIC PEACE VIOLATION	50,387
PROSTITUTION	69,275
WEAPONS VIOLATION	81,619
CRIMINAL TRESPASS	203,507
ROBBERY	268,771
DECEPTIVE PRACTICE	293,856
MOTOR VEHICLE THEFT	329,264
BURGLARY	404,074
OTHER OFFENSE	444,538
ASSAULT	451,402
NARCOTICS	731,906
CRIMINAL DAMAGE	815,147
BATTERY	1,310,820
THEFT	1,512,278

Top 15 Causes of Incidents Visualization

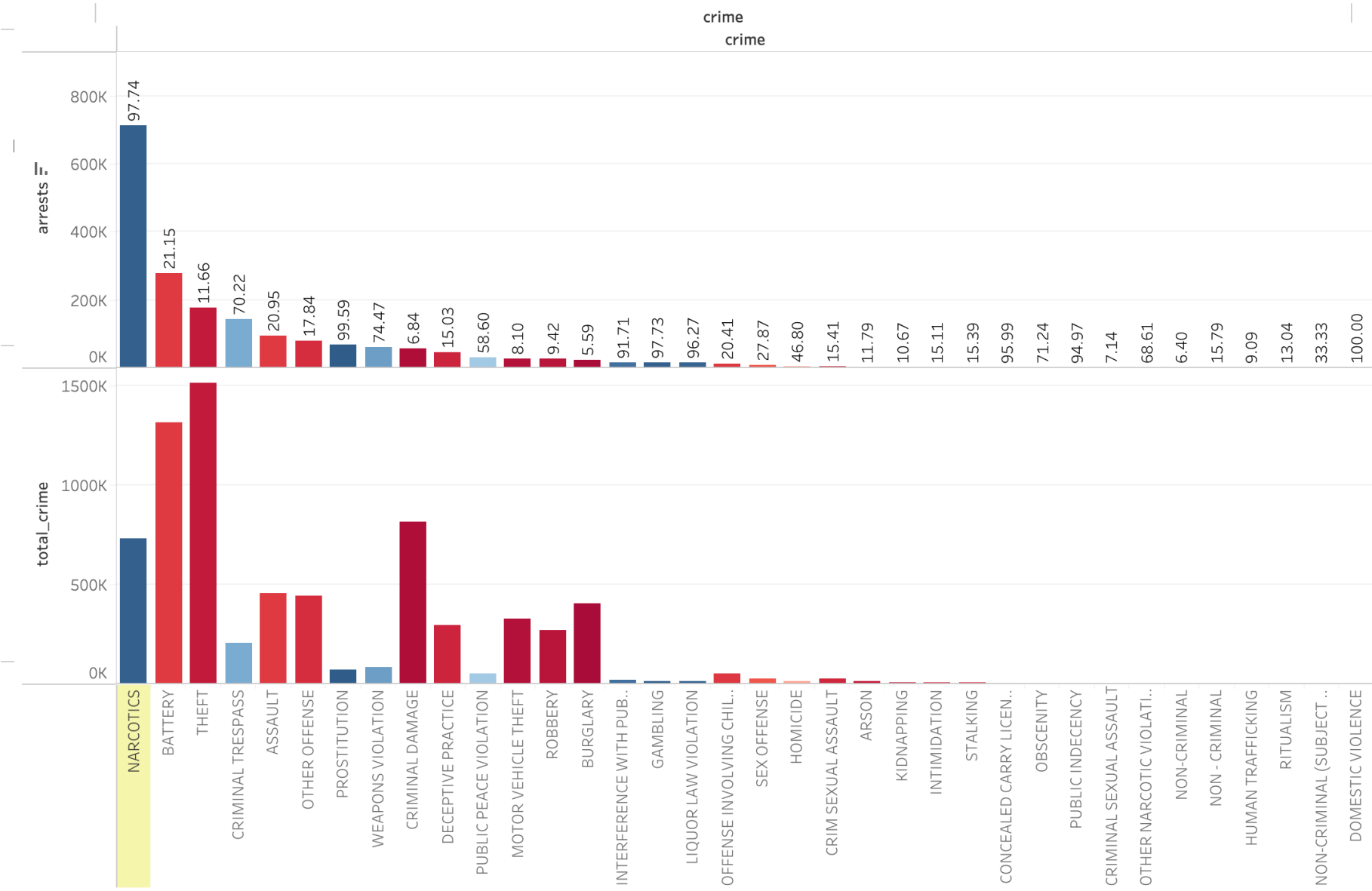
Sheet 1



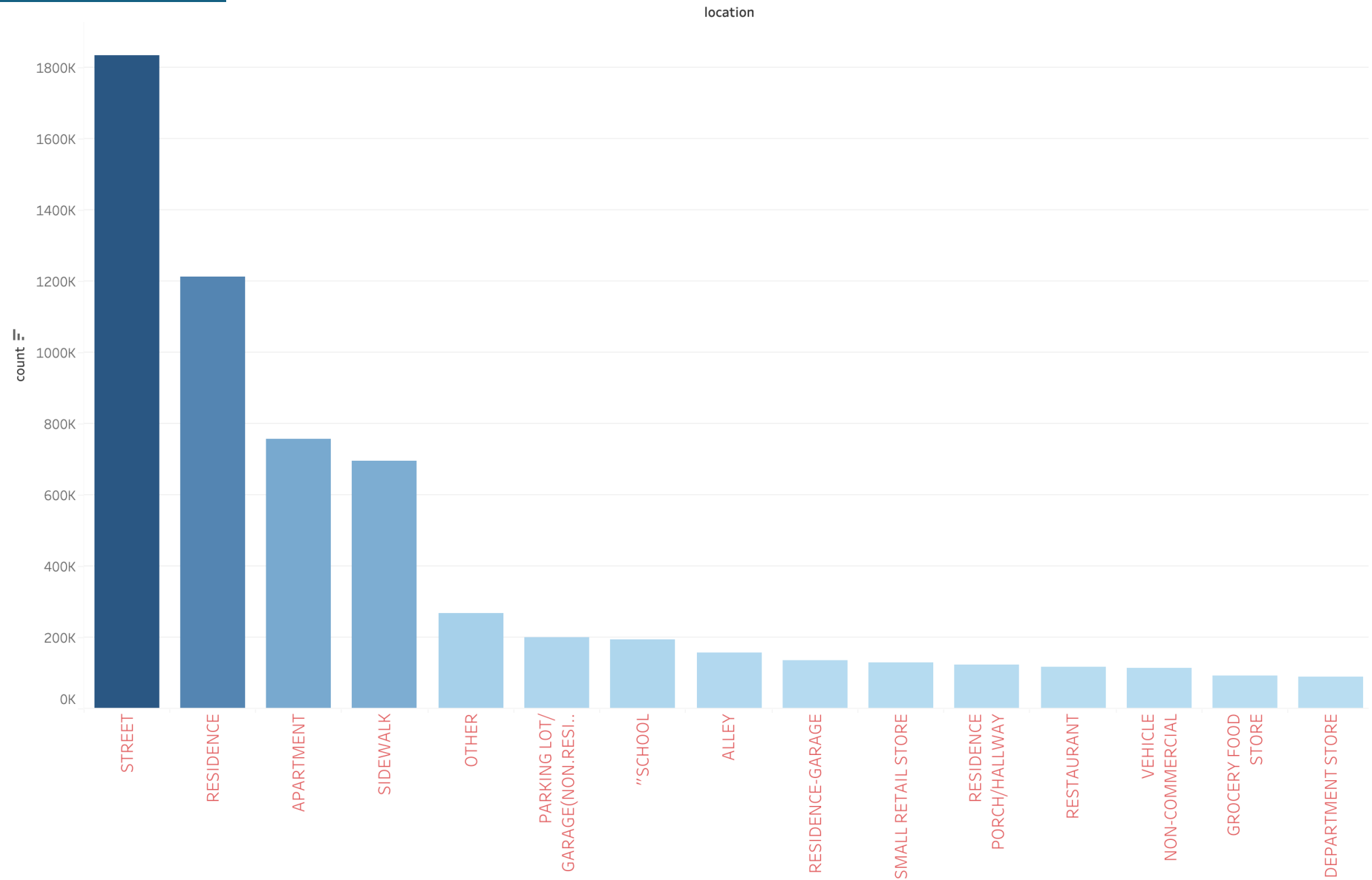
Arrests Analysis by Crime Type

This bar chart compares the crimes reported to arrests made. This allows us to understand the types of crimes that result in arrests.

The ratio of ‘arrests made’ to ‘crimes reported’ is also shown in percentages.

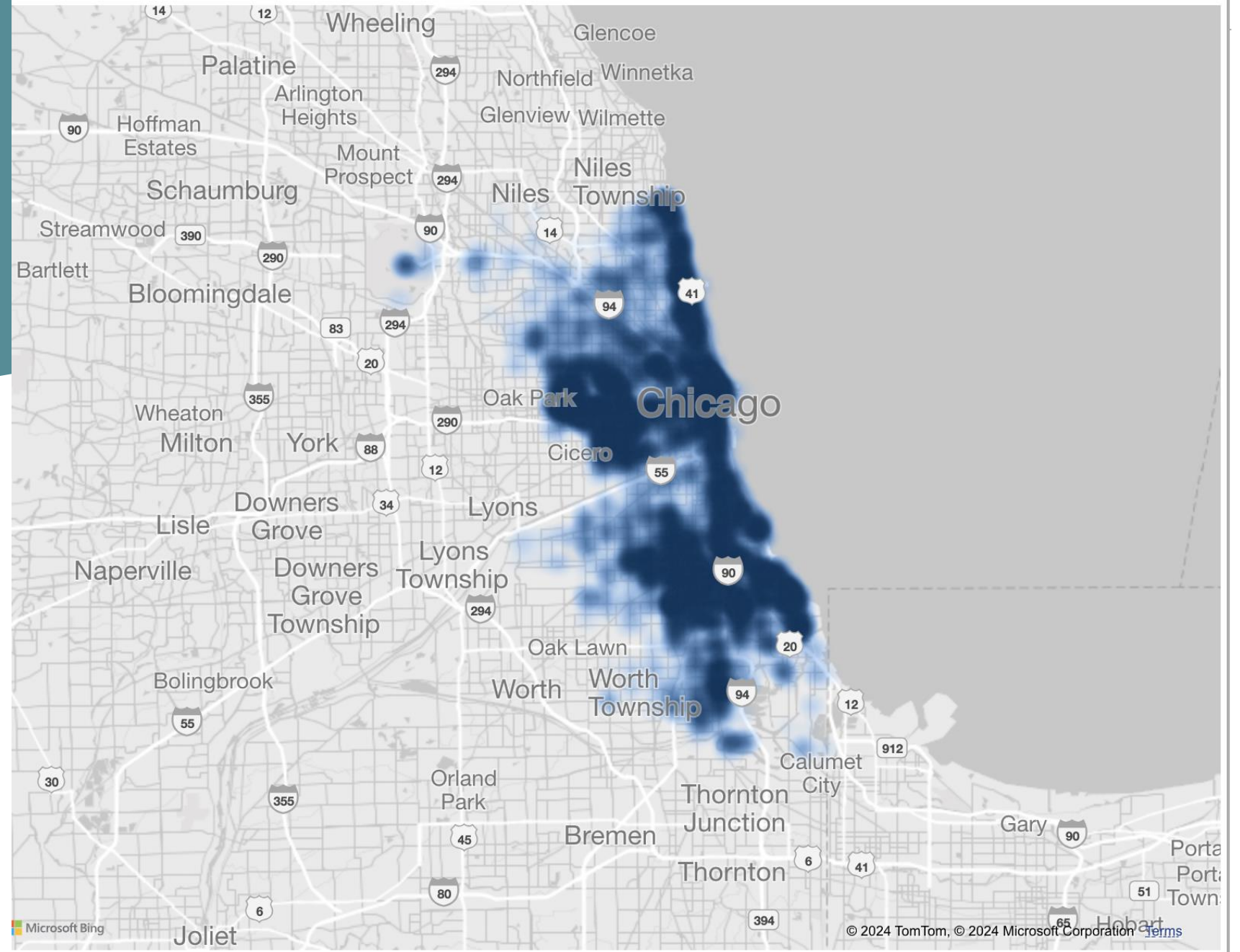


Crime by Location



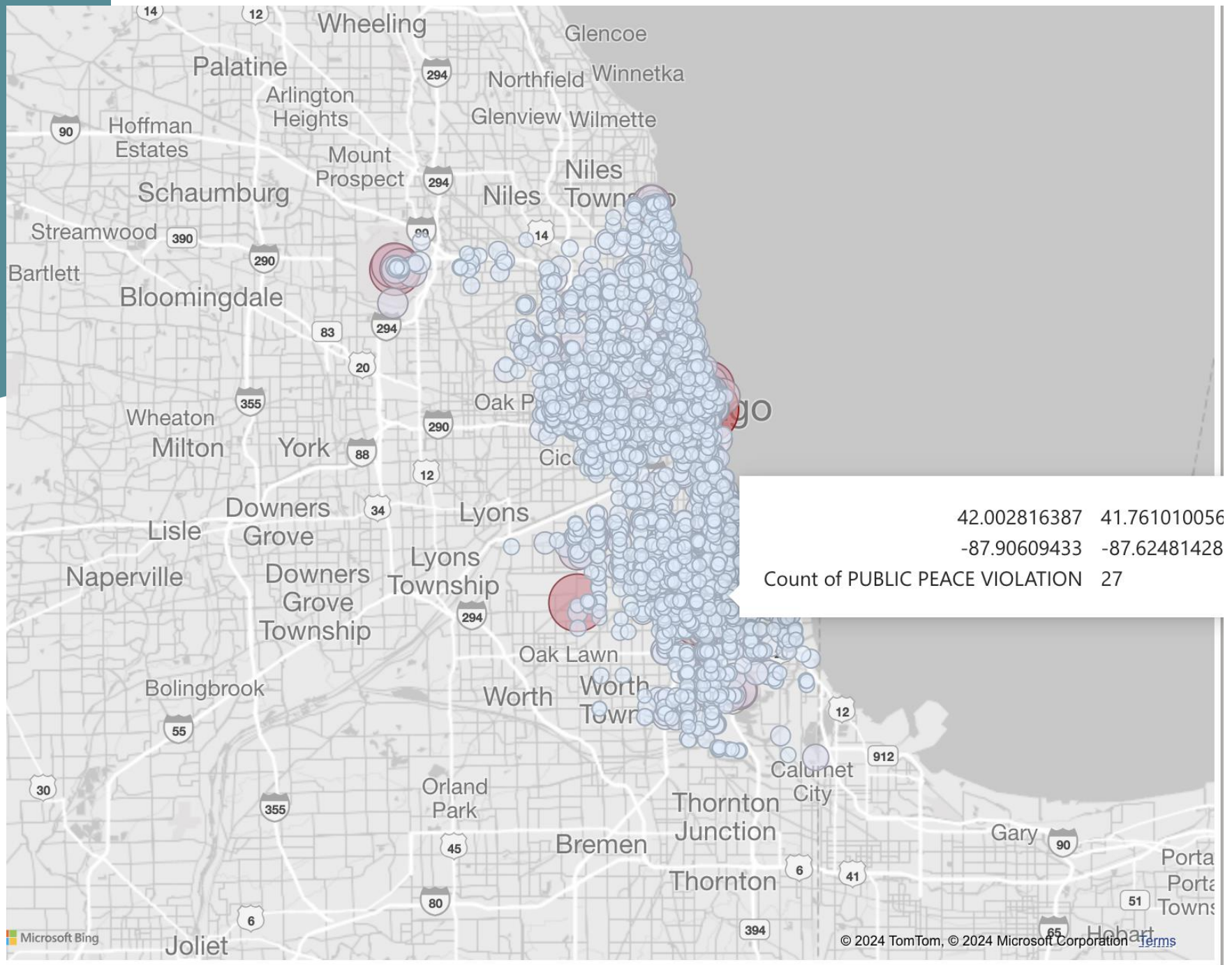
Geo-Spatial Analysis

Heatmap



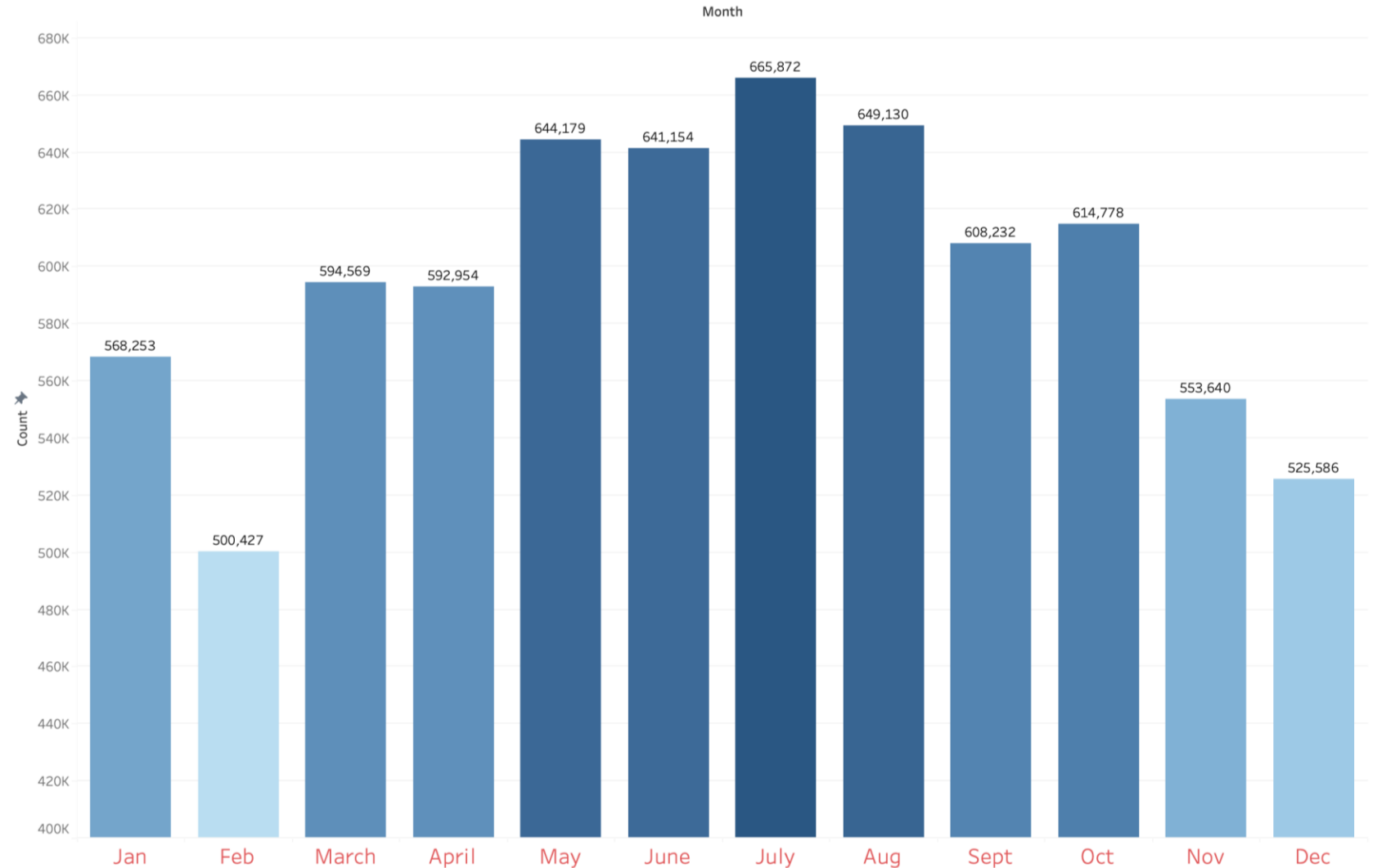
Geo-Spatial Analysis

Bubbles



Temporal Analysis Crime Count by Month

This Chart compares the reported crimes each month, providing insights into the unresolved issues across Chicago.

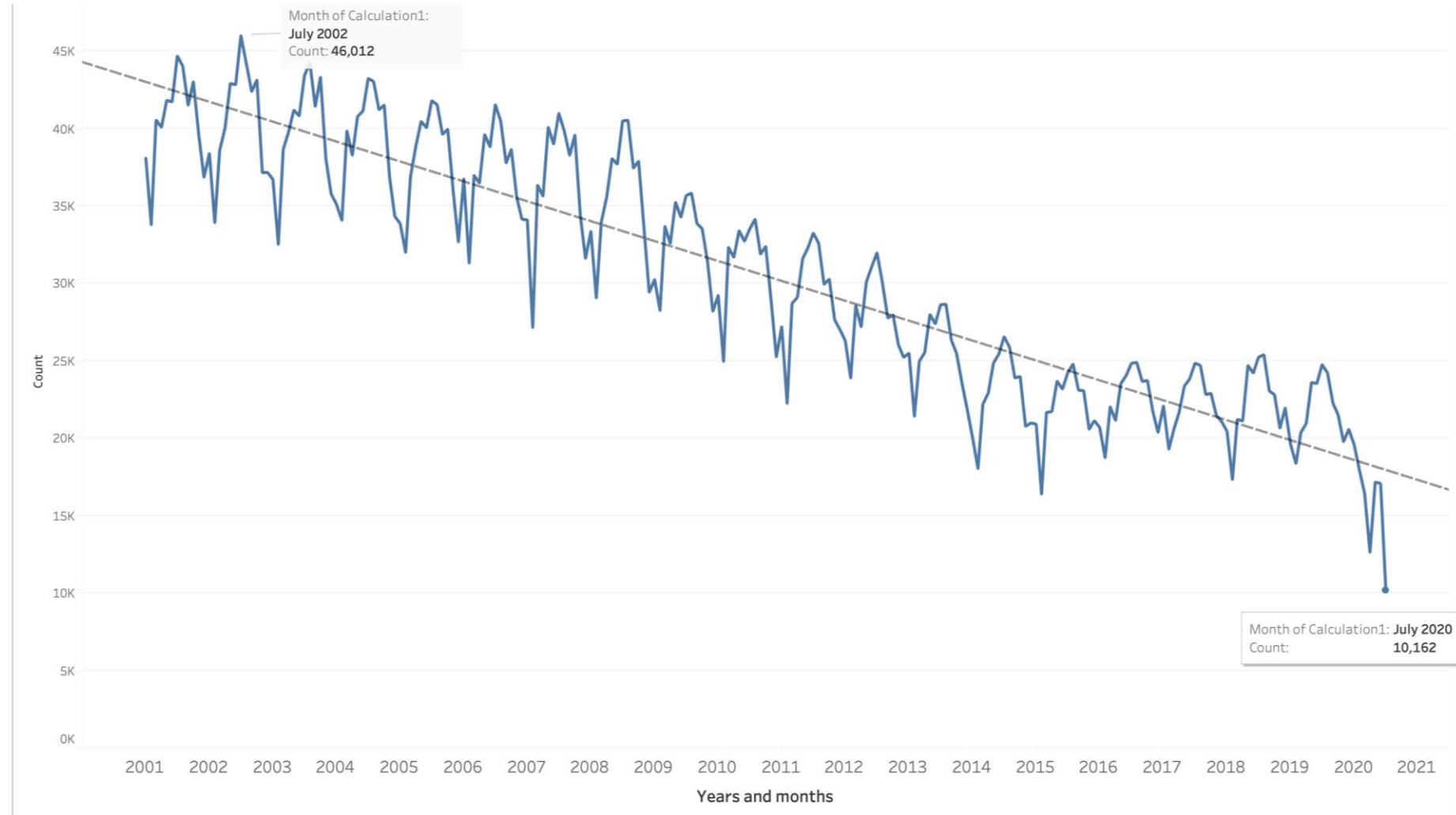


Temporal Analysis Crime Count by Year

This slide presents a temporal analysis of crime in Chicago from 2001 to 2020.

A peak in crime incidents is observed in July 2002, with a count of 46,012 reported cases.

Despite fluctuations, there is an overall trend of declining crime rates over the observed period.



Implementation Flow Chart



Download
the Chicago PD
Dataset from
Kaggle.com

Step One



Upload the
Dataset using
HDFS

Step Two



Sort the Data
and
Create Tables in
Beeline/Hive

Step Three



Use Tools to do
Data
Analysis and
find **Meaning**

Step Four

Thank you

Jerry Petty, Mayowa Toyinbo, Erik Pacheco, Francisco Figueroa, Eric Ventura, Gerardo Camarena

CIS 4560-01 | Group 3