

Unveiling Insights of Chicago Crimes

Venkata Sai Charan(OO18458)

Bhaavani Madabathula(KN44173)

PROJECT MOTIVATION

We aim to analyze Chicago Crime data from 2001 to 2024 to determine the number and types of the crimes committed, and to report the arrest rates for each crime type.

Here we can know the number of crimes occurred in a day or in a month or in a year, so that this is helpful to reduce the crime-rate in Chicago.

Problem Statement

We want to find out what are the prevalent types of crimes in Chicago, and how do their seriousness and punishment correlate? Furthermore, what geographic hotspots exhibit higher frequencies of these crimes?

Additionally, we want to find out the score accuracy of Arrest rate using different hyperparameters and models.

Dataset

- ❖ The dataset includes details on criminal activity in a certain location.
- ❖ The data is for 24 years that is from 2001 - 2024.
- ❖ The data includes information on case number, date, location, type of offense, and whether an arrest was made or not .
- ❖ The dataset is used to produce visualizations that reveal patterns and trends in crime.
- ❖ The insights of the dataset are clearly plotted.

https://data.cityofchicago.org/Public-Safety/Crimes-2001-to-Present/ijzp-q8t2/about_data

Variables we plan to use throughout analysis:

Date - Date on which crime occurred.

Primary Type - Primary type of crime.

IUCR – Four-digit Illinois Uniform Crime Reporting (IUCR) codes.

Description - Short description of the type of crime.

Location description - Description of where crime occurred.

District - District code where crime occurred.

Community - Community area code where crime occurred.

Longitude & Latitude - Exact coordinates of crime occurrence.

FBI Code - numeric code indicating FBI crime categorization.

Year - Year of crime.

Arrest - Indicates whether arrest was made or not

Exploratory Data Analysis (EDA)

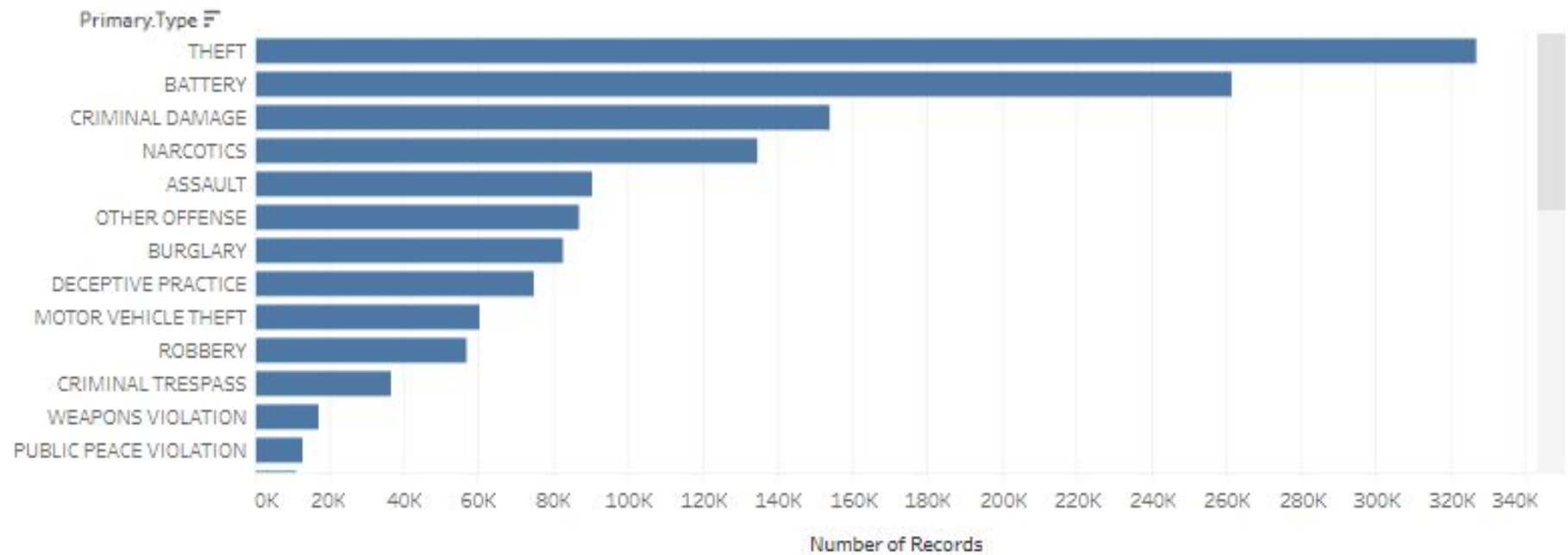
Explore our Tableau Story :

Delve into the data visualization journey through Chicago crime statistics from January 2012 to December 2016.

<https://public.tableau.com/app/profile/venkata.sai.charan.lysetty/viz/UnvielingInsightsofChicagoCrimes/Story1?publish=yes>

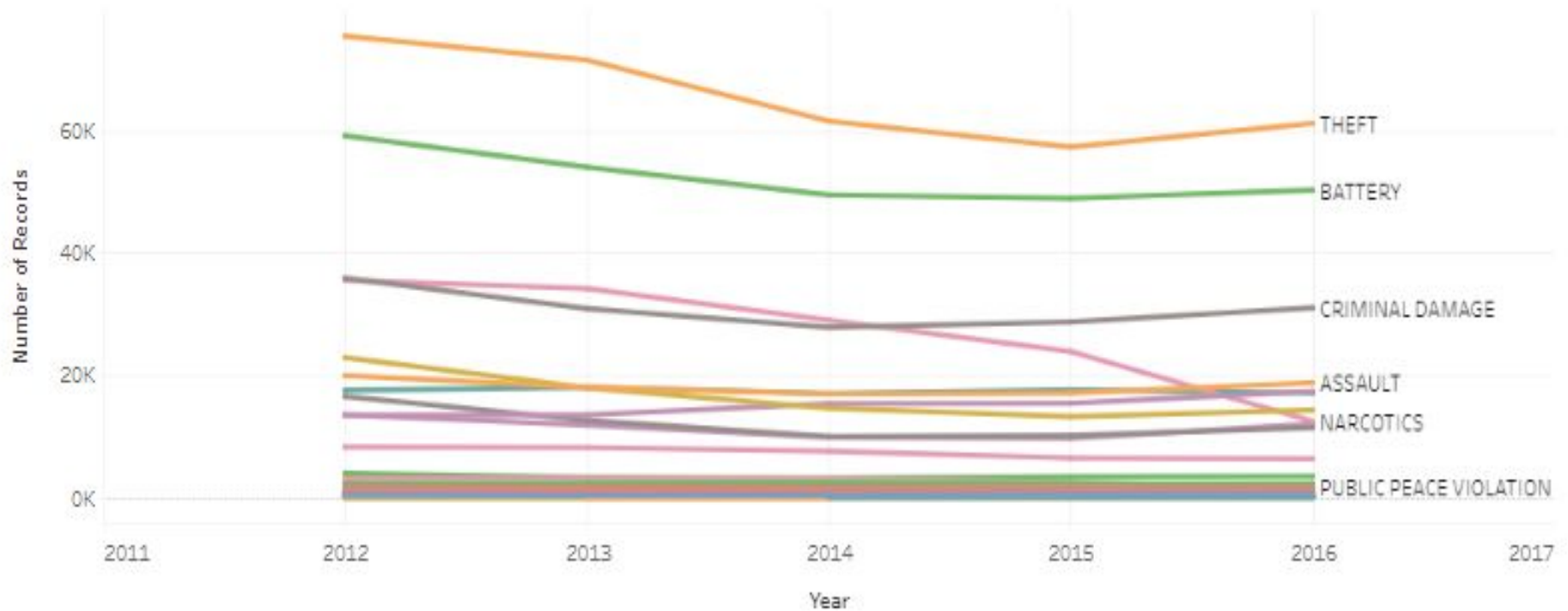
Our journey begins with a plot showcasing primary crime types and their occurrences. Theft emerges as the dominant crime type, while non-criminal offenses are relatively scarce.

Primary Crime Types



A forecasted trend then predicts future crime occurrences, segmented by primary crime type.

Trend - Primary Crimes



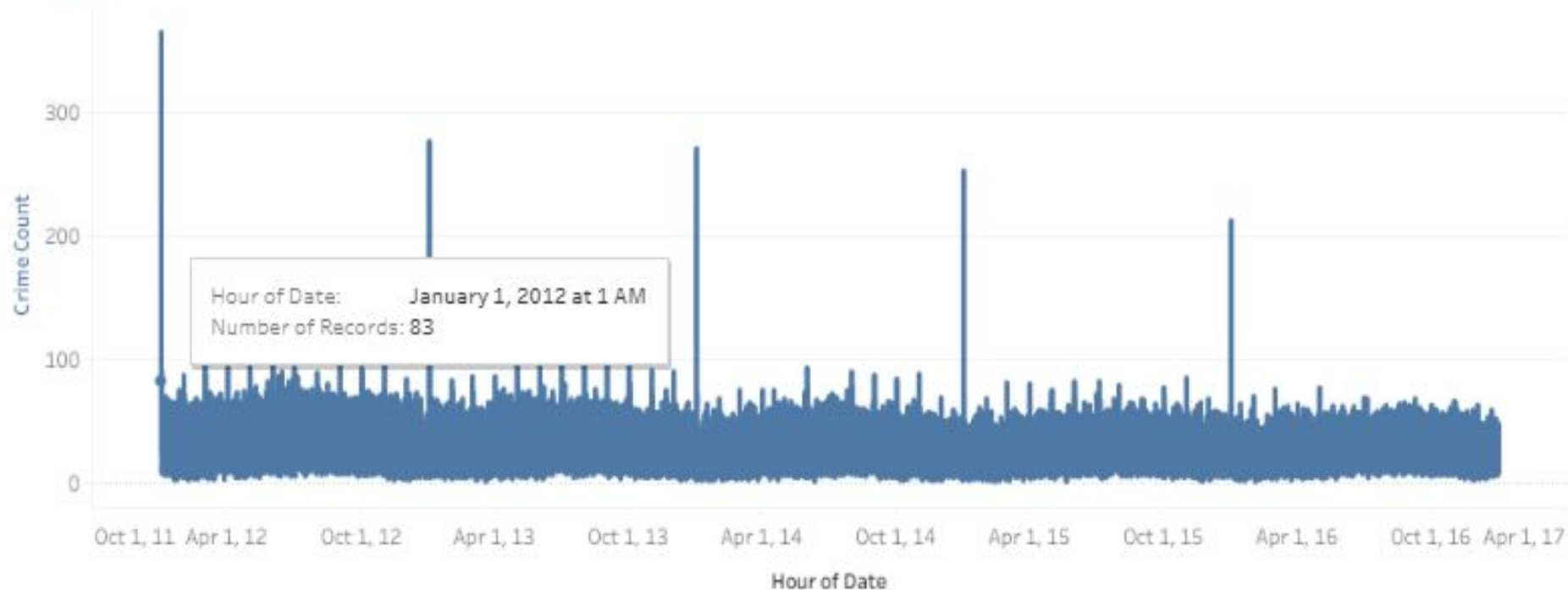
The second worksheet presents a crime timeline across all years, segmented by quarters. Notably, crime records peak in the third quarter and diminish in the fourth quarter consistently over the years.

Quarterly Crime Analysis

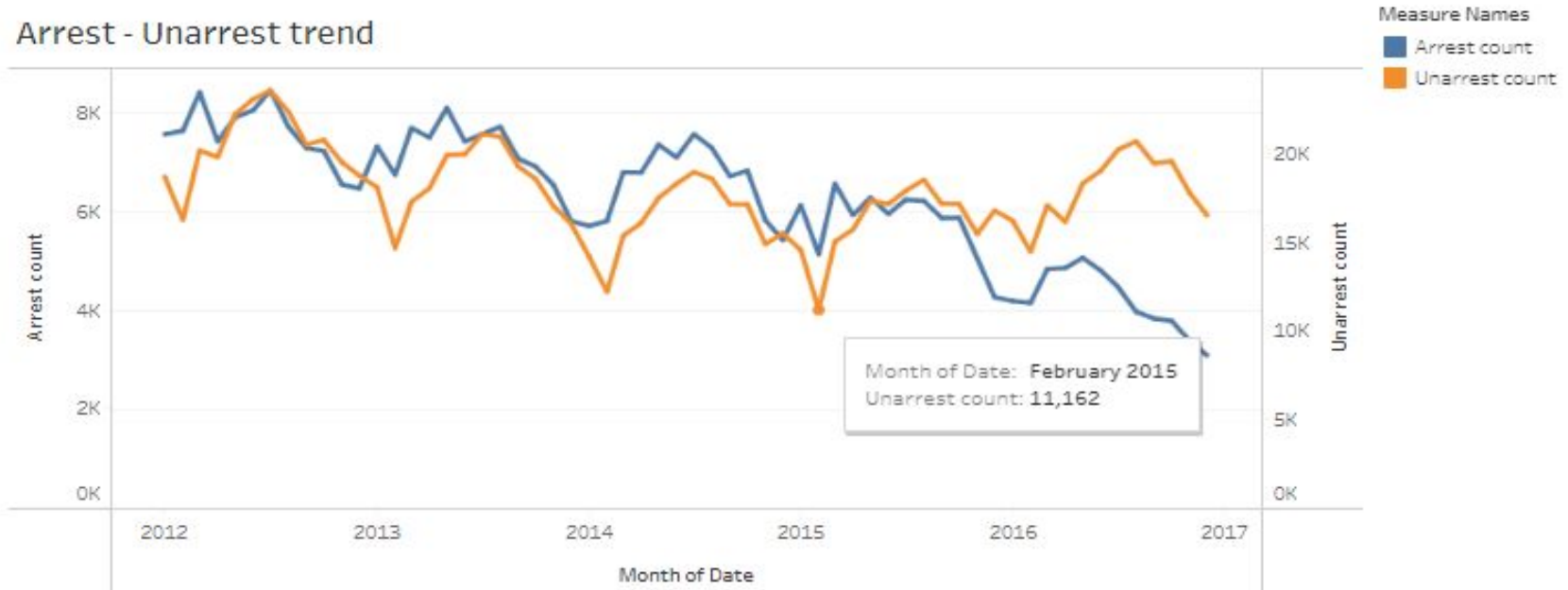


We explore crime occurrences per hour on specific dates. Notably, the visualization suggests a heightened risk of crime on January 1st.

Crimes By Hour

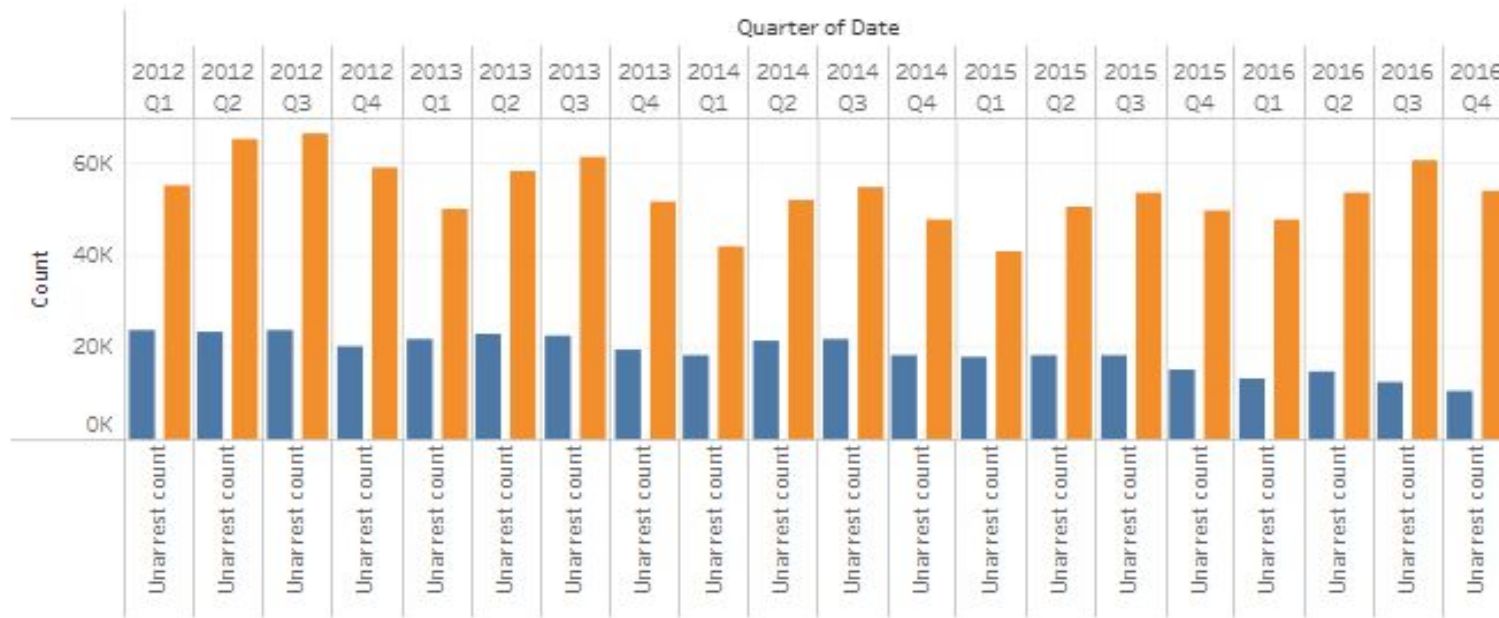


Our attention then shifts to trend analysis, focusing on arrest and unarrest counts over the years. Through categorized measures, we witness a gradual decrease in arrest rates.

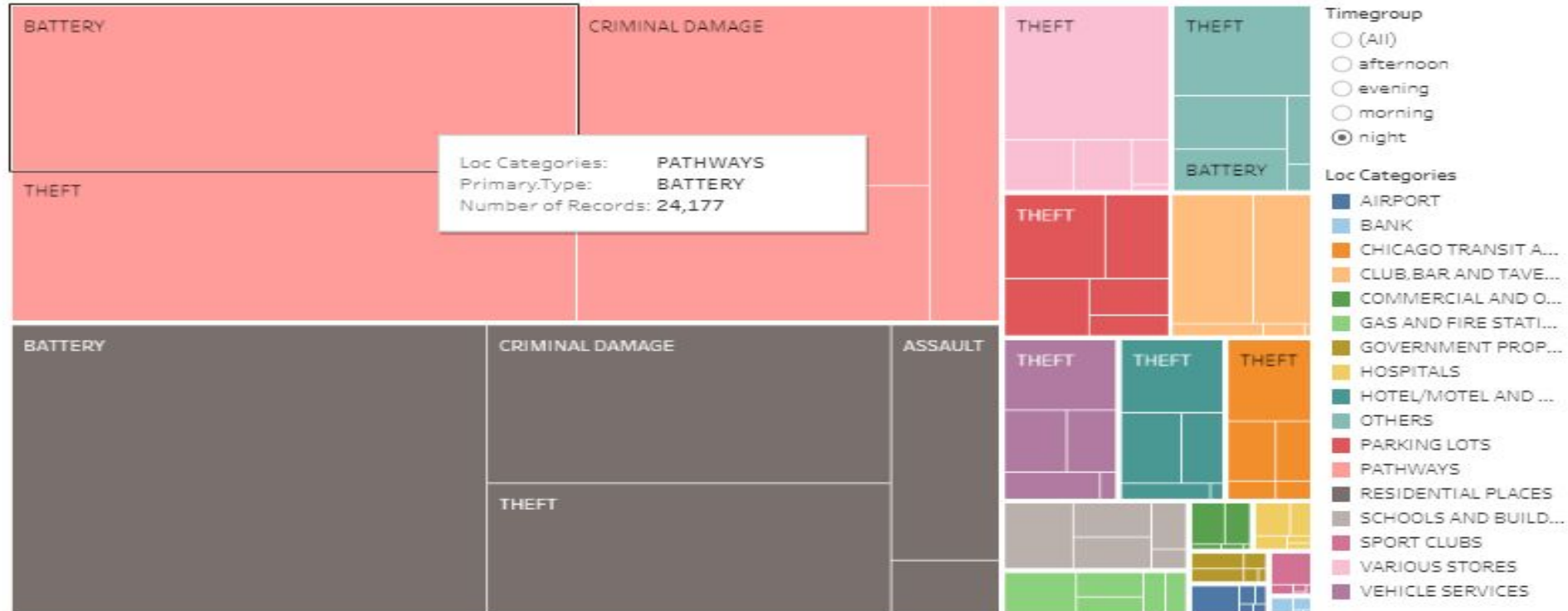


Quarterly comparisons highlight the significant disparity between arrest and unarrest counts. Through this we can clearly say that arrest count very less when compared to crime occurrences

Arrest - Unarrest Count

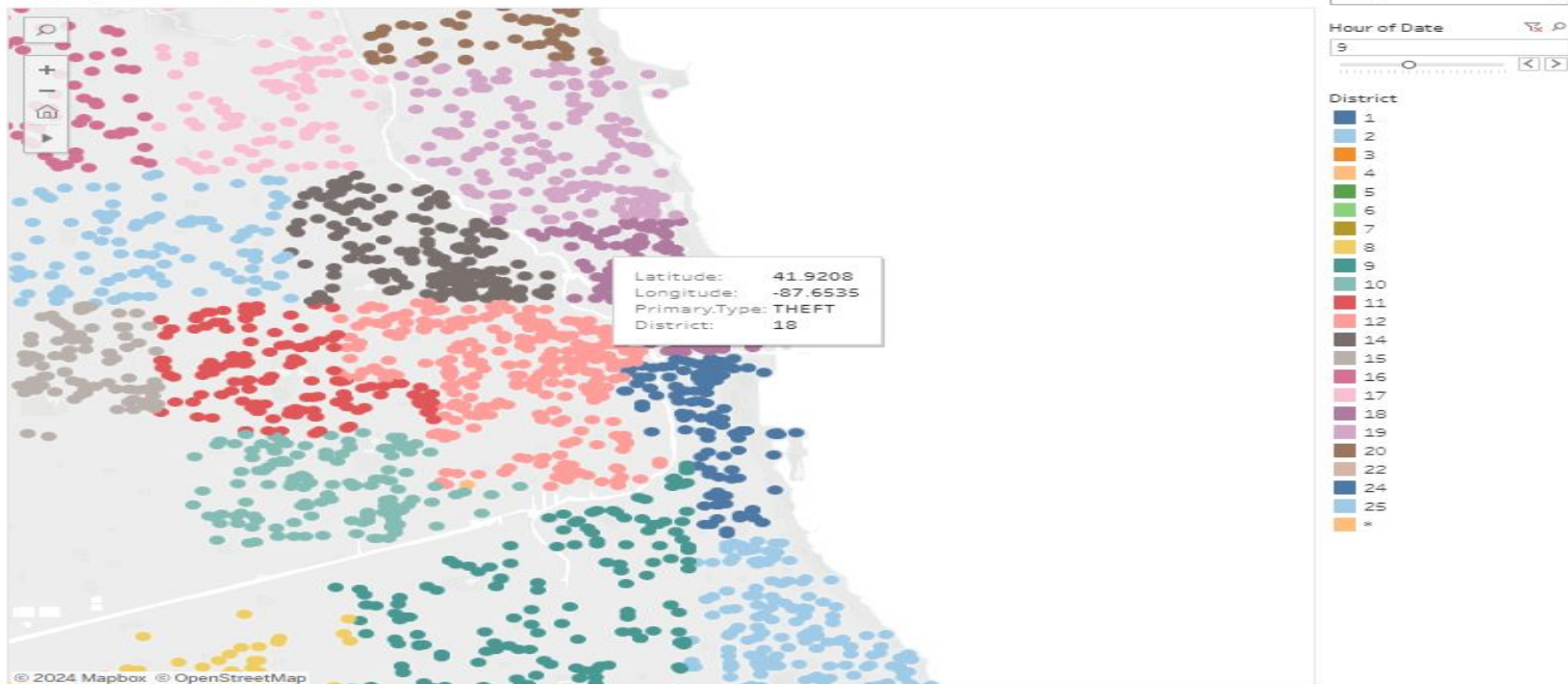


Moving to the next worksheet, we find a detailed enumeration of crimes based on location and time grouping. Color-blocked sections represent crime hotspots, further categorized by crime type. Hovering over these blocks reveals the crime count and types, dynamically adjusting with the filtered time group.



In our final worksheet, we encounter a dynamic display of crime instances across Chicago's districts. Each colored dot represents a specific crime type occurring in a district. As we filter the data by hour, the visualization adapts, showcasing the crime distribution based on the filtered hour.

Hourly Theft in District



Overview

Each visualization within our tableau tells a story, unveiling patterns, trends, and insights within Chicago's crime data landscape. By analyzing Chicago crime data according to months in a year, stakeholders can identify seasonal patterns and allocate resources effectively to address peak periods of criminal activity. Additionally, targeted interventions and prevention strategies can be implemented to mitigate crime during specific months when rates are typically higher.

Models and Results

Models for Prediction

- **Logistic Regression Classifier:**

Logistic Regression is the model that usually works on any data frame and the scores given by the logistic regression are accurate depending upon the data set.

- **Decision-Tree-Classfier:**

We have chosen this because as we have more categorical data, It is useful to derive as it has feature of capturing the descriptive decision-making skills.

- **KNN- Model:**

We have chosen this as it is a simple supervised algorithm where we can assign the class to new data point, and it is non-parametric model.

Scores

Logistic Regression Classifier

- ROC-Accuracy Score: 89.27%
- Logistic Regression score: 88.89%
- F-1 Score: 70.04%

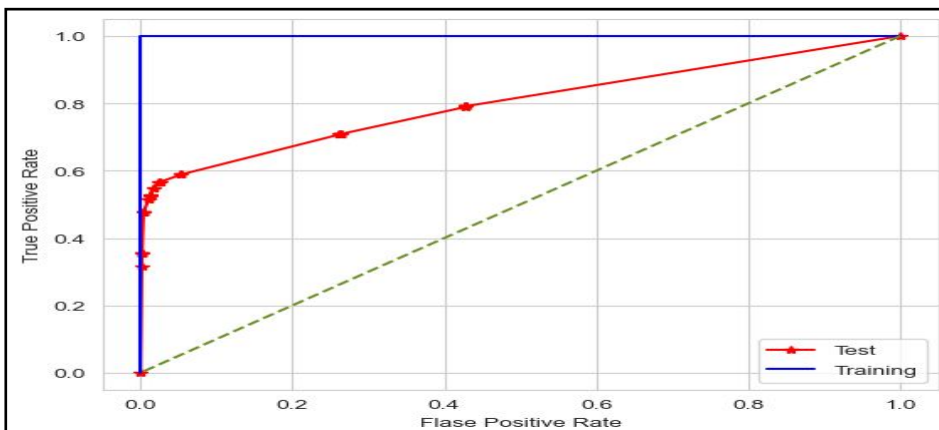
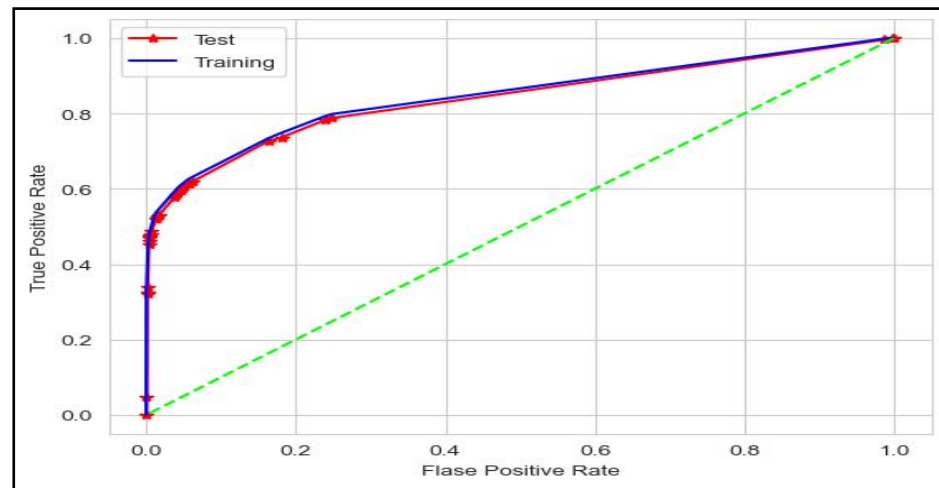
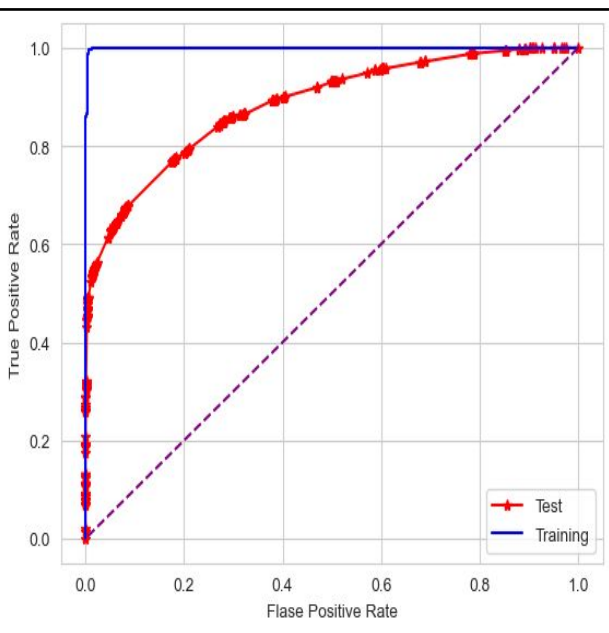
Decision-Tree Classifier

- ROC-Accuracy Score: 85.60%
- Decision Tree score: 88.56%
- F-1 Score: 69.48%

KNN-Model Classifier

- ROC-Accuracy Score: 82.66%
- KNN- model score: 86.56%
- F-1 Score: 61.94%

ROC Curves for 3 Models





Comparing Models

Logistic Regression performs well among the decision tree and KNN model

The ROC accuracy score is higher for the logistic regression as it maintains 89.27%

And comparing the F1 score among the models here also the logistic regression is best with 70.04%

Coming to the crimes most of the crimes are theft based, in 10th district the crimes are high and most of the crimes occurred in 10th month i.e., October.



INSIGHTS:

In summary, analyzing Chicago's crime data from 2001 to the present reveals a complex interplay of factors influencing fluctuations in crime rates.

From the introduction of data-driven policing strategies like COMPSTAT to socioeconomic conditions, gang activity, and shifts in law enforcement priorities post-9/11, various elements have impacted crime trends.

Community policing initiatives, technology adoption, violence reduction programs, and changes in criminal justice policies further shape the landscape of crime in the city. Economic conditions, including both booms and crises, also play a significant role.

Ultimately, understanding Chicago's crime dynamics requires a multifaceted approach encompassing historical data, community dynamics, law enforcement strategies, and socioeconomic factors.

As our tableau of Chicago crime statistics unfolds, it reveals not just numbers, but narratives of resilience, challenges, and the evolving landscape of urban safety. Through these visualizations, we gain a deeper understanding of the dynamics shaping our city, empowering us to enact informed strategies for a safer tomorrow.