Presented by Team 13,

Los Angeles Crime Data Analysis

Insights into Crime Dynamics and Predictive Factors

One in 31

Los Angeles Crime Rates

Los Angeles has one of the higher crime rates in the United States, with a rate of 32 per one thousand residents. **This statistic places an individual's chance of becoming a victim of either violent or property crime in Los Angeles at one in 31.**

Guiding Research Questions

1 “What are the **primary factors** influencing the **seriousness of crimes** committed?”

2 What is the **predicted number of crimes** for a given area and time period?

Crime patterns in Los Angeles,
using advanced statistical models.

- **Data as a Foundation:**

  LAPD Crime Records (2020 - Present)

- **Focus and Purpose**:

  Understanding Factors Influencing Crime Seriousness and Predicting Crime Number

- **Methodologies**:

  Logistic and Poisson Regression Models

Why **This Study Matters**

Public Safety

Policy Making

Community Awareness

Data Source + Origin

Official

Data provided by the Los Angeles Police Department (LAPD), crime reports from 2020 to present.

Massive

0.8 Million(820,599) observations, 20+ vital variables, comprehensive view of crime patterns in LA

Live

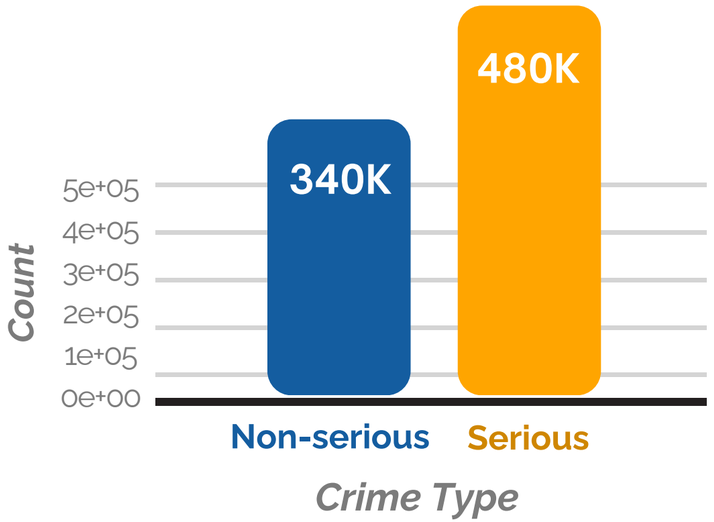
Updated weekly, ensuring current and relevant insights.

Dataset Overview

Each row = crime reported by the Victim

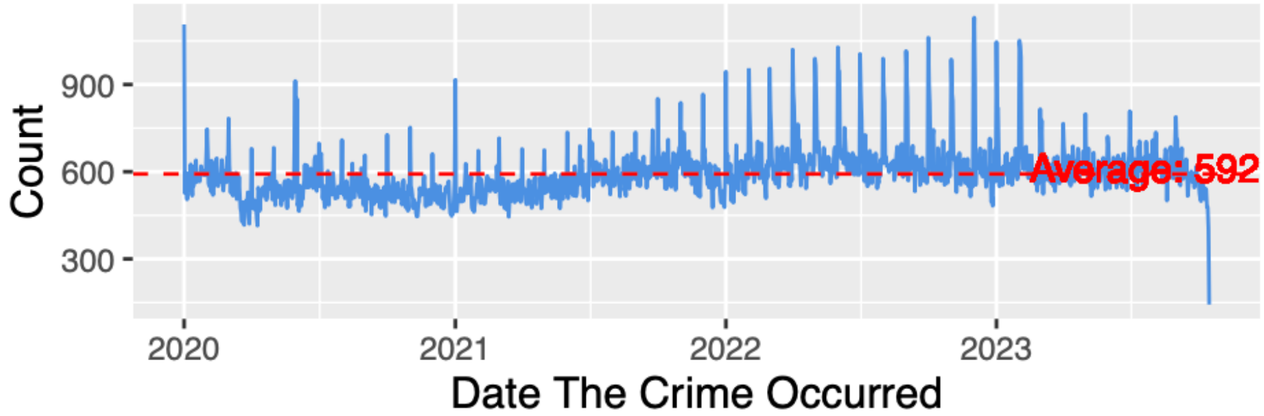
Location	Latitude, Longitude, Area, Street, District.
Victim Demographics	Age, Gender, Ethnicity.
Crime Description	Category, Investigation Outcomes, Weapons.
Date & Time	Date Reported & Occurred, Time Occurred.
Identifier/Classifier	Crime Record Identifier, Mocodes.

Dataset Overview

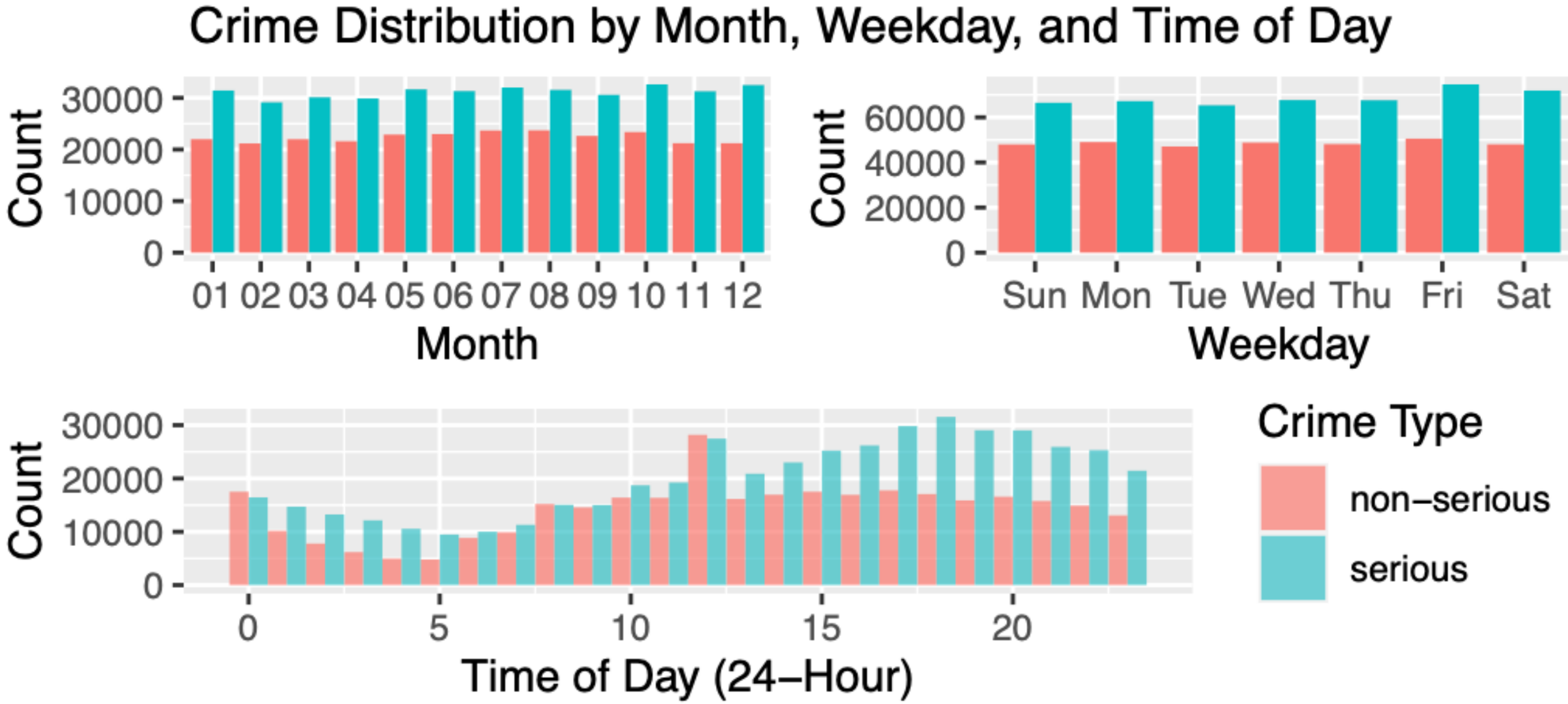


Distribution of Crime Types

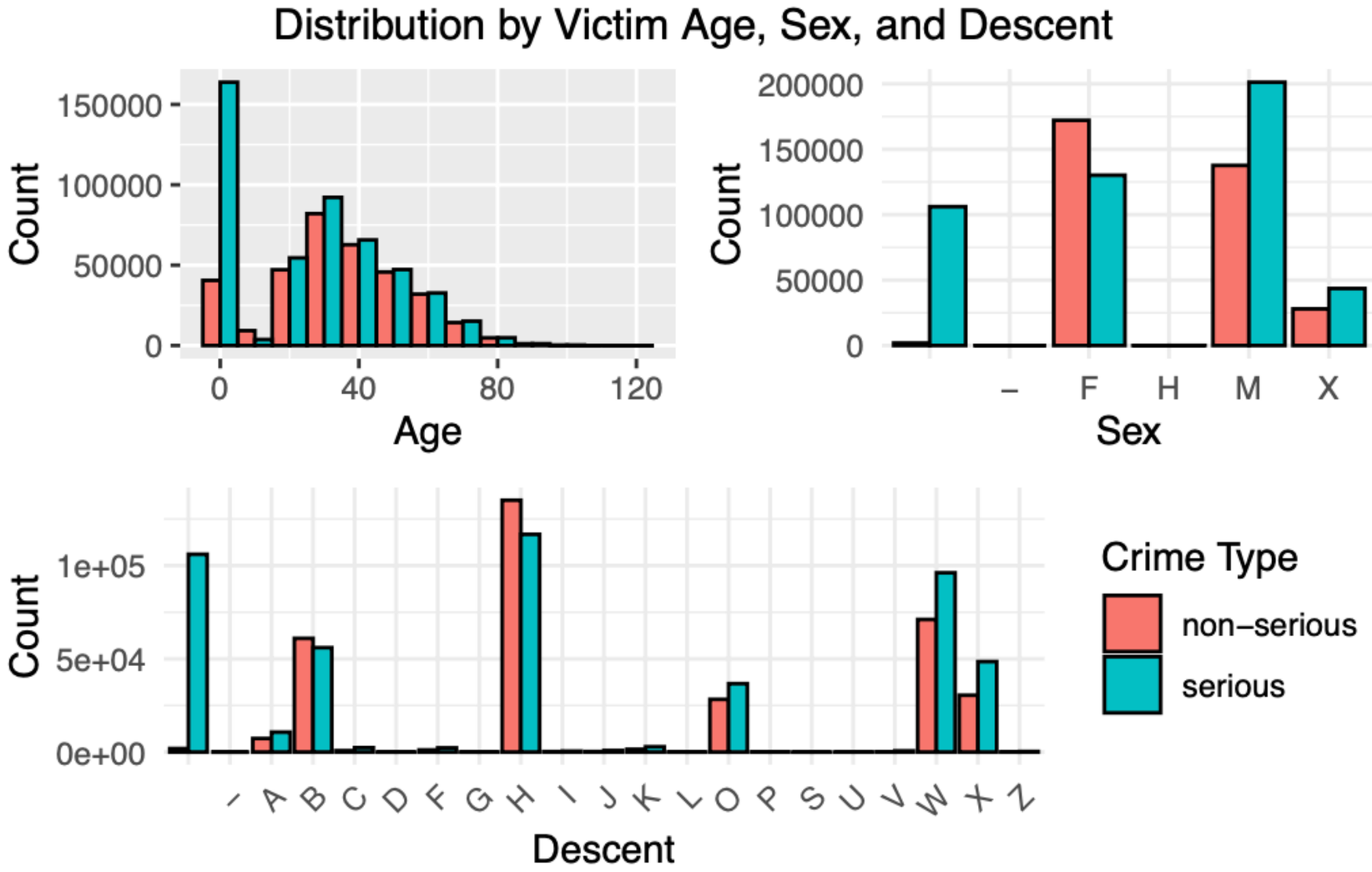
Daily Crime Counts



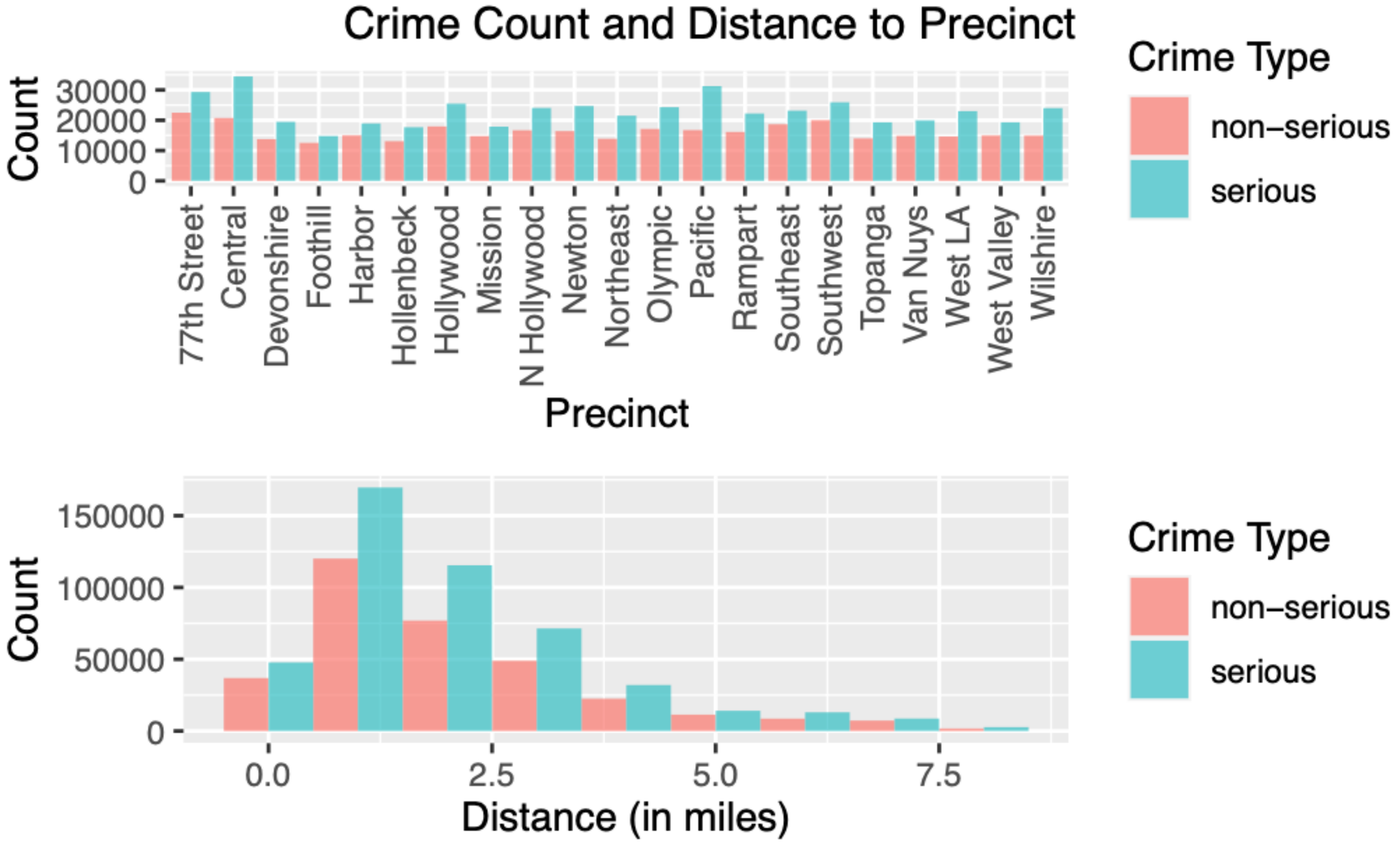
Dataset Overview



Dataset Overview



Dataset Overview



Research Question:

1 “What are the **primary factors** influencing the **seriousness of crimes** committed?”

Logistic Regression:

Understanding Crime Seriousness

01. Appropriateness for a binary outcome (serious vs. non-serious crimes)

02. Easily interpretable to demonstrate
how various factors influence crime seriousness

Variables and **Data Prep**

Outcome Variable

Crime Seriousness

(categorical)

Independent Variables

Times Occurred	(categorical)
Distance to Precinct	(continuous)
Demographic Information	(age, sex, descent)
Weapon Used or Not	(categorical)

Key Insights from Logistic Regression

01. Times Occurred

Time of crime occurrence, particularly at night or early morning, showed a higher likelihood of serious crimes.

02. Distance to Precinct

Proximity to certain precincts was linked to the frequency of serious crimes.

03. Demographic Info

Variations in serious crime occurrences were noted based on victim race, gender, and descent.

04. Weapon Used

Weapon usage significantly increased the probability of a crime being classified as serious

Limitations and to-do

01. **Evident that the model does not fit the data particularly well**

02. **Additional measures are required for more accurate inferential result**

To address the remaining concerns:

1. Consider adding more variables or obtain new data by introducing new variables.
2. Consider how to handle values categorized as 'Other' or 'Unknown' in the demographic data.

Analysis:

2 What is the predicted number of crimes for a given area and time period?

Poisson Regression:
Predicting Crime Frequency

- the outcome variable of research question involves counting events, which is a characteristic that aligns well with the Poisson distribution.

- The focus is on the number of events that occur in a fixed interval of time or space, and we try to predict number of crimes may occur based on geographic area, weekday and month.

- With satisfied result, we may support government to improve resource allocation and take proactive approach to prevent crimes.

Variables and Data Preparation

- Key Independent Variables: weekday, month, area generated from Time Occurred, Date Occurred, Geographic Area.
- Outcome Variable: Count of crimes, aggregated daily.
- Data Preparation: Targeted selection of variables, transform time occured and date occurred to weekday and month, addressing missing values.

Spatial and Temporal Patterns in Crime

- Areas such as Devonshire and Foothill show notably lower crime rates, whereas the Central area experiences higher crime frequencies.

- the coefficients for the months indicate seasonal variations, with certain months like September showing a marked decrease in crime occurrences.

- An observation of a slightly higher crime rate on Sundays was confirmed by the model result.

Limitations and to-do

- A dispersion value above 1 suggests over-dispersion, indicating potential issues with the model assumptions.

- Factors contributing to this may include Population Heterogeneity, Model Misspecification, Zero-Inflation and Correlation Among Observations, among which Zero-Inflation and Correlation Among Observations have been ruled out.

- To address the remaining concerns:

1. Consider using a negative binomial model instead of Poisson to account for over-dispersion.
2. Troubleshooting issues related to Population Heterogeneity and Model Misspecification for model improvement.

Thank you