

# CRIME RATE ANALYSIS

Crime Busters

# AGENDA

Introduction & Primary Goals

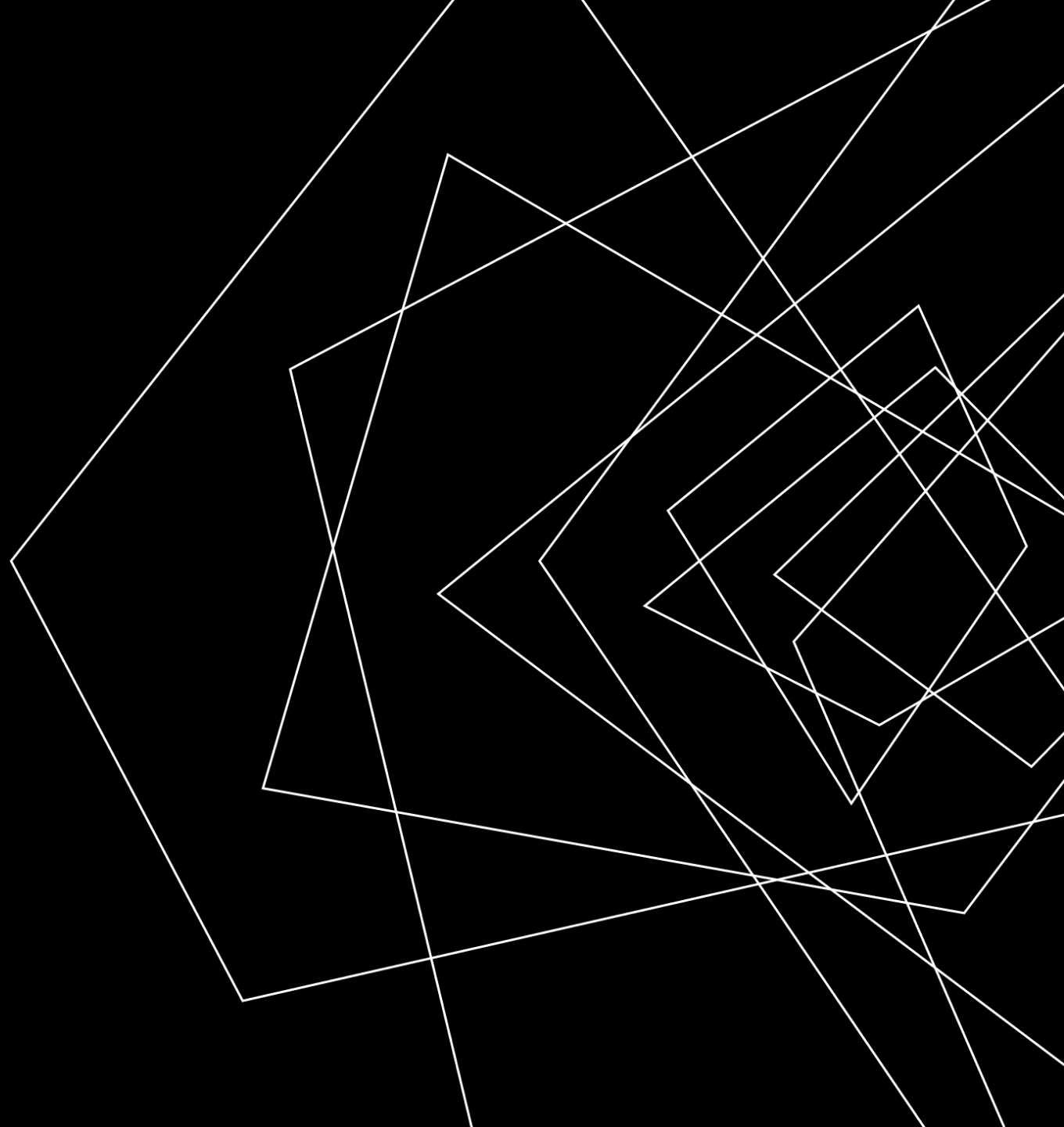
Data quality & Data Cleaning

Data Visualization

Feature vectorization

Random Forrest Classifier using Spark

Summary



# INTRODUCTION

This study aims to analyze crime in Los Angeles ,CA identifying types and assessing their economic impact. By uncovering trends, we seek actionable insights for more effective crime prevention and enhanced public safety.



# PRIMARY GOALS

- ▶ Factors affecting the crimes like Gender ,Age, Ethnicity, Area Density, Area Type
- ▶ Feature vectorisation and Random Forrest classifier to predict crimes
- ▶ This study can be further enhanced by Geospatial analysis, focusing on crime hotspots and patterns within a specific state.

# PLAN FOR PRODUCT LAUNCH

Data setup	Data Schema Creation	Data Quality	Data Cleaning
Finding and understanding the dataset	Converting data types of the columns	→ Formatting date-based columns using spark	→ Cleaning out Anomalies in Gender using spark → Cleaning out incorrect age using spark → Clearing out blank's columns using spark

# USE CASES

- 1 — Type of crime vs Number of Victims  
Result: Bar chart
- 2 — Gender vs Number of victims  
Result: Bar chart
- 3 — Region vs Average age of Victims  
Result: Bar Chart
- 4 & 5 — Ranking criminal incidents by frequency.  
Identifying the most prevalent crime based on occurrence frequency.  
Result: Bar chart

# USE CASES

- 6 — Hypothesis: Burglaries occurred between 10PM and 5Am and Victim's age >60  
Result: False. More burglaries are happening on Senior's citizen age in morning time
- 7 — Geo spatial analysis  
Result: LA Map overview
- 8 — Feature vectorization  
Result: Long vector with the resulting data
- 9 — Random Forrest classifier  
Result: Accuracy – 11.66%



# FUTURE EXTENSIONS

Fine tune the Spark ML code to scale the model

Resizing the current model to accommodate the complete historical data of US crimes, while ensuring seamless data processing without encountering any performance constraints

Identify patterns of similar crimes across the country which would help in improving the Patrolling during the crime interval observed

Sentiment Analysis for the perpetrators



# SUMMARY

## Success

- Achieved design goals as per project proposals
- Classified the crime based on people, property
- Crime rate was higher for Males
- The average age of victim in the central region was around 40-43 years
- Rank the Area with the highest recorded crimes.
  - Central region has the highest number of crimes with crime rate at 42711 reported cases stands number for crimes in LA
  - Hollenbeck has the lowest number of crimes with 28149 cases standing at 10th rank
- 10AM to 5 AM identify the burglaries happened where the age of victim >60
  - Crime rate was higher for Senior citizens above 60 during day time with 9357 crimes reported and the number of cases reported at night was 1984 as compared to night time throughout the LA areas

# RESOURCES

1. <https://catalog.data.gov/dataset/crime-data-from-2020-to-present>

(Data Source for the project)

1. <https://data.gov/>

2. [https://www.census.gov/geo/partnerships/pvs/partnership23v2/st06\\_ca.html](https://www.census.gov/geo/partnerships/pvs/partnership23v2/st06_ca.html)

3. <https://data.gov.in/high-value-datasets>



# THANK YOU

Ajay Raj

Nithin Poovana

Gayathri Ramasubramanian

Sandya K M