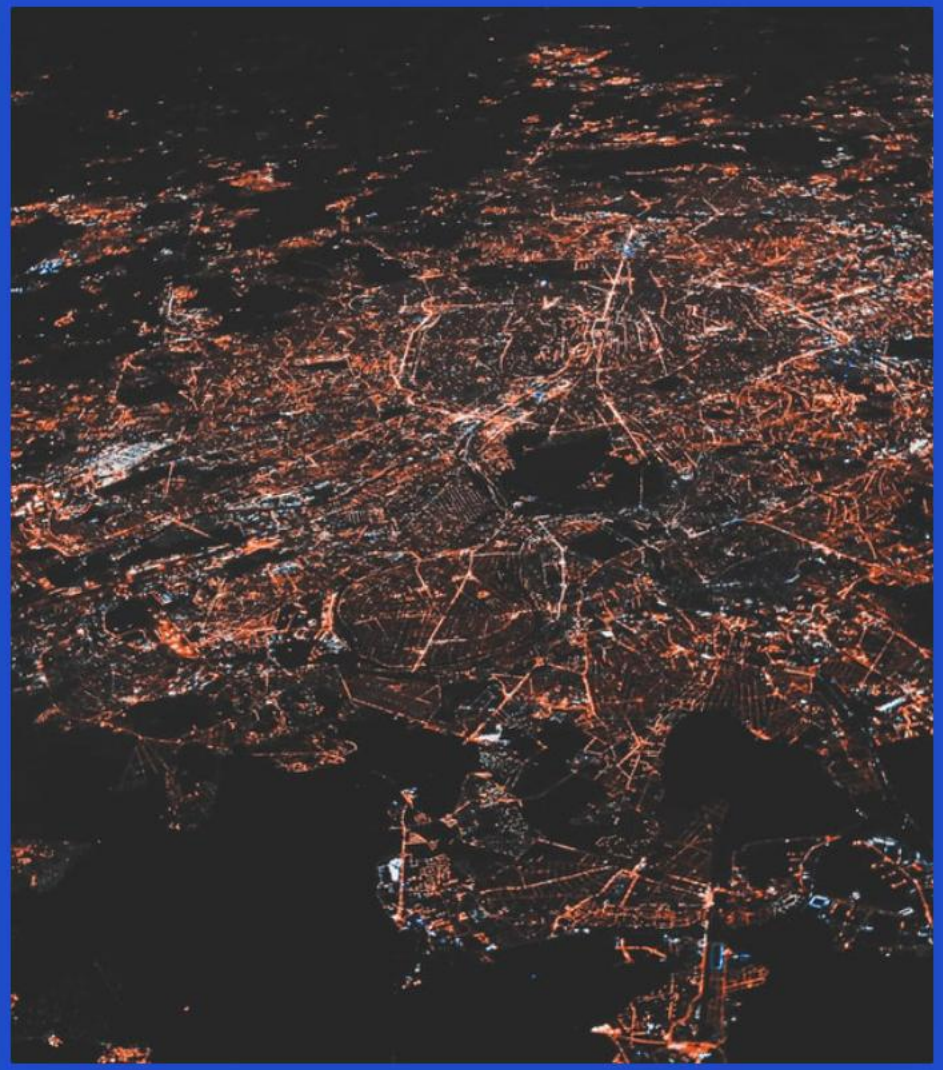
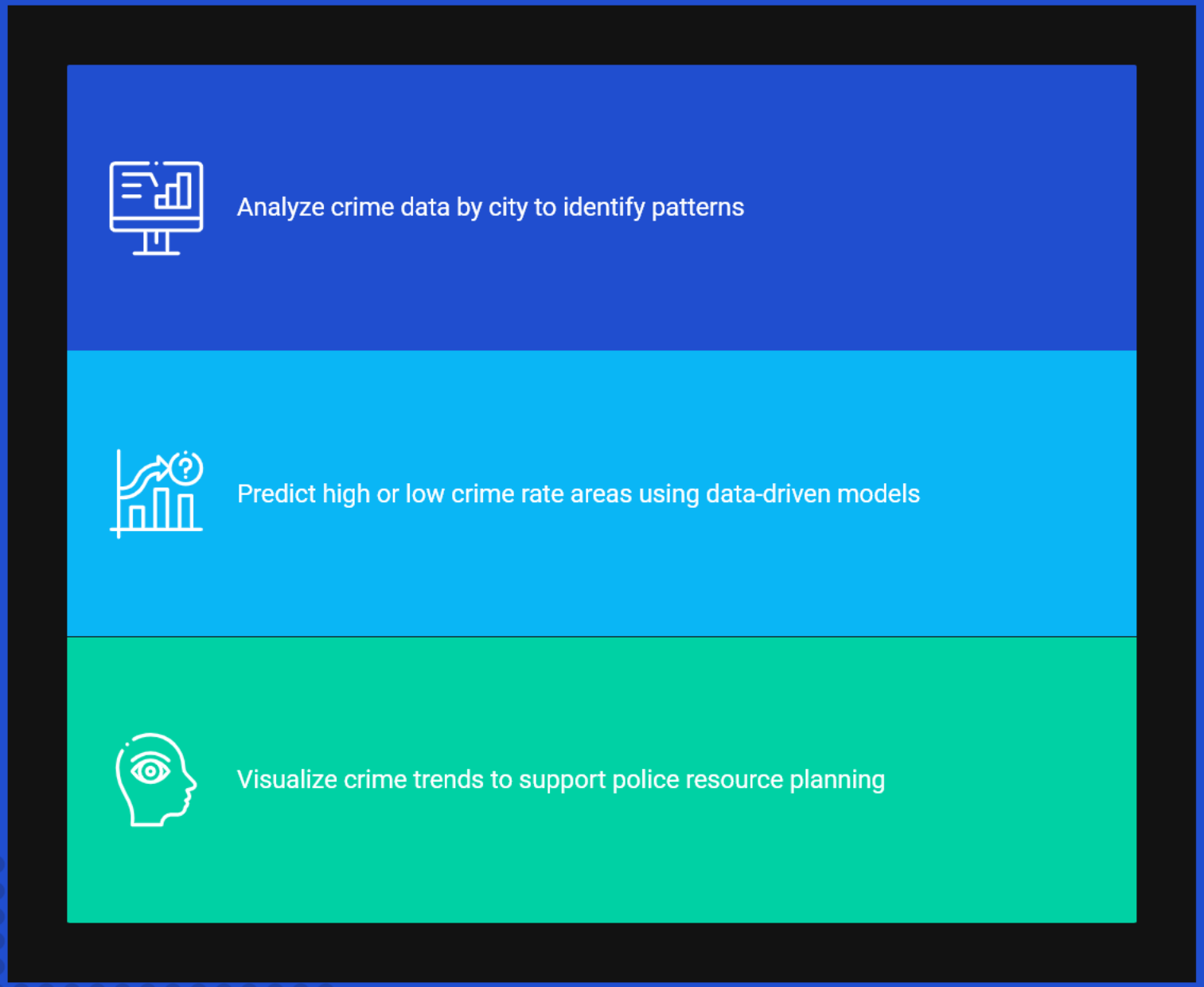
Predicting Crime to Enhance Community Safety

Leveraging machine learning to identify high-risk areas and support proactive law enforcement





Drive Safer Communities with Data Insights

Analyze, predict, and visualize crime patterns to empower police planning and resource allocation

Comprehensive Dataset Overview

Explore key data columns and sample cities shaping crime analysis

Key Dataset Columns	Sample Cities
Report Number, Date Reported, City	Ahmedabad
Crime Code, Description, Victim Age	Chennai
Gender, Weapon Used, Crime Domain	Ludhiana
Police Deployed, Case Closed Status, Date Closed	Pune
Target Variable: High Crime Area Status	

Essential Data Cleaning Steps for Reliable Analysis

Refined the dataset by addressing missing values, standardizing formats, encoding categories, and defining key targets

Removed missing and invalid values to ensure data integrity

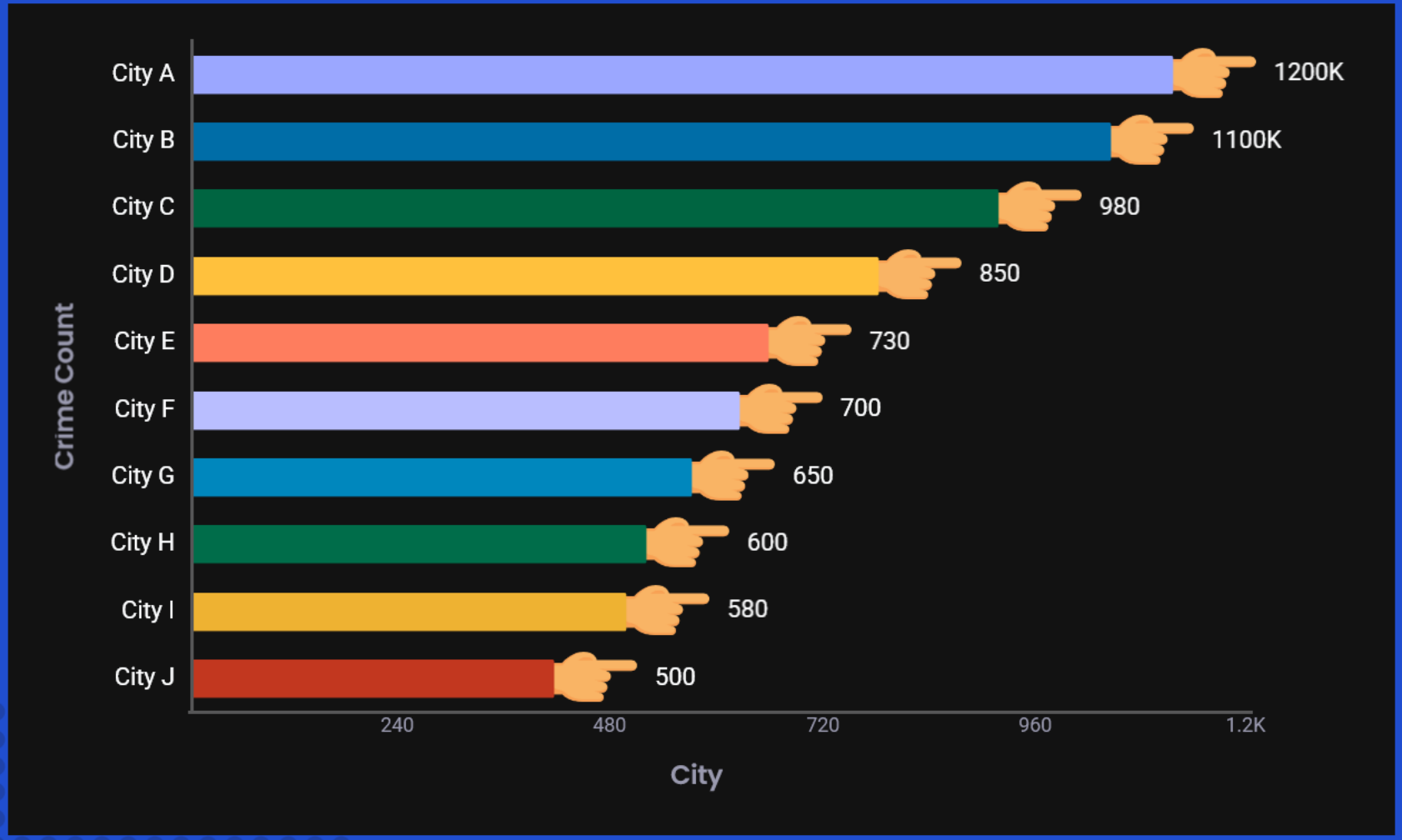
Standardized date columns to consistent, correct formats

Encoded categorical features including City, Crime Domain, and Weapon Used

Created a new target column ‘High Crime Area’ based on crime frequency analysis

Exploratory Data Analysis Reveals High-Crime Cities

Visualizing crime frequency to identify and compare the top 10 cities with the highest crime counts using Seaborn and Matplotlib.



We conducted an exploratory data analysis (EDA) using Seaborn and Matplotlib to examine crime frequency across cities. The analysis identified the top 10 cities with the highest crime counts. This visualization aids in understanding regional crime severity and supports targeted interventions.

Training a Robust Random Forest Model

We trained a **Robust Random Forest Model** to predict high-crime areas using city and crime-related features.

1. It combines multiple decision trees to improve accuracy and stability.
2. The model reduces overfitting and handles complex data patterns well.
3. This helps identify areas with higher crime risk more reliably.

Think of it like this:
Instead of asking one tree (model) to decide if an area has high crime,
you ask 100 trees and take the majority vote.
That makes the prediction more stable and accurate.

Evaluating Model Performance with Precision

Comprehensive analysis using sklearn metrics and confusion matrix visualization to validate crime area classification accuracy

We employed sklearn evaluation tools to measure model performance, focusing on Precision, Recall, and F1-score from the classification report. The confusion matrix provided a detailed comparison of true versus predicted labels, revealing clear distinctions between high and low crime area classifications. Results were visualized with a heatmap to emphasize the model's strong accuracy in identifying crime levels.

Exploring Future Enhancements

Advancing crime prediction accuracy and visualization capabilities

Leverage deep learning models to enhance prediction accuracy and insights

Integrate live crime feeds or APIs from police data sources for real-time updates

Create interactive dashboards using Power BI or Streamlit for advanced visualization

Clear Team Roles Driving Success

Each member's expertise ensured smooth project execution and insightful outcomes.

Task	Members
Data Cleaning	Vanshika
EDA & Visualization	Menika Maurya
Model Training	Deeksha
Evaluation	Aditya Kumar Shukla
Presentation	Divy Agarwal
Overall Assistance	Shoaib Ali

That’s it