**Crime Data Analysis and Response Time Prediction**

Shivam Kansara, Krish Jogi

22BCP090, 22BCP173

Department of Computer Engineering

Pandit Deendayal Energy University, Gandhinagar

**Abstract**  
This project presents an analytical study and predictive modeling on Indian crime data. Using real-world datasets from Kaggle, we perform comprehensive crime analysis such as case closure rates, hotspot identification, temporal patterns, and domain-wise crime distribution. Additionally, machine learning models are trained for crime type classification and police response time prediction. The results assist in proactive crime monitoring and better resource allocation.

**Keywords**  
Crime Analysis, Machine Learning, Crime Prediction, Response Time Analysis, Crime Hotspots, Indian Crime Data, Data Science, Predictive Analytics.

**I. INTRODUCTION**  
Crime analytics plays a crucial role in ensuring safety, planning police deployment, and proactively preventing crimes. This project aims to analyze a comprehensive dataset of crimes reported in India, identifying patterns and predicting future trends using machine learning models. The dataset, sourced from Kaggle, includes detailed information such as crime type, location, time, victim details, and police response.

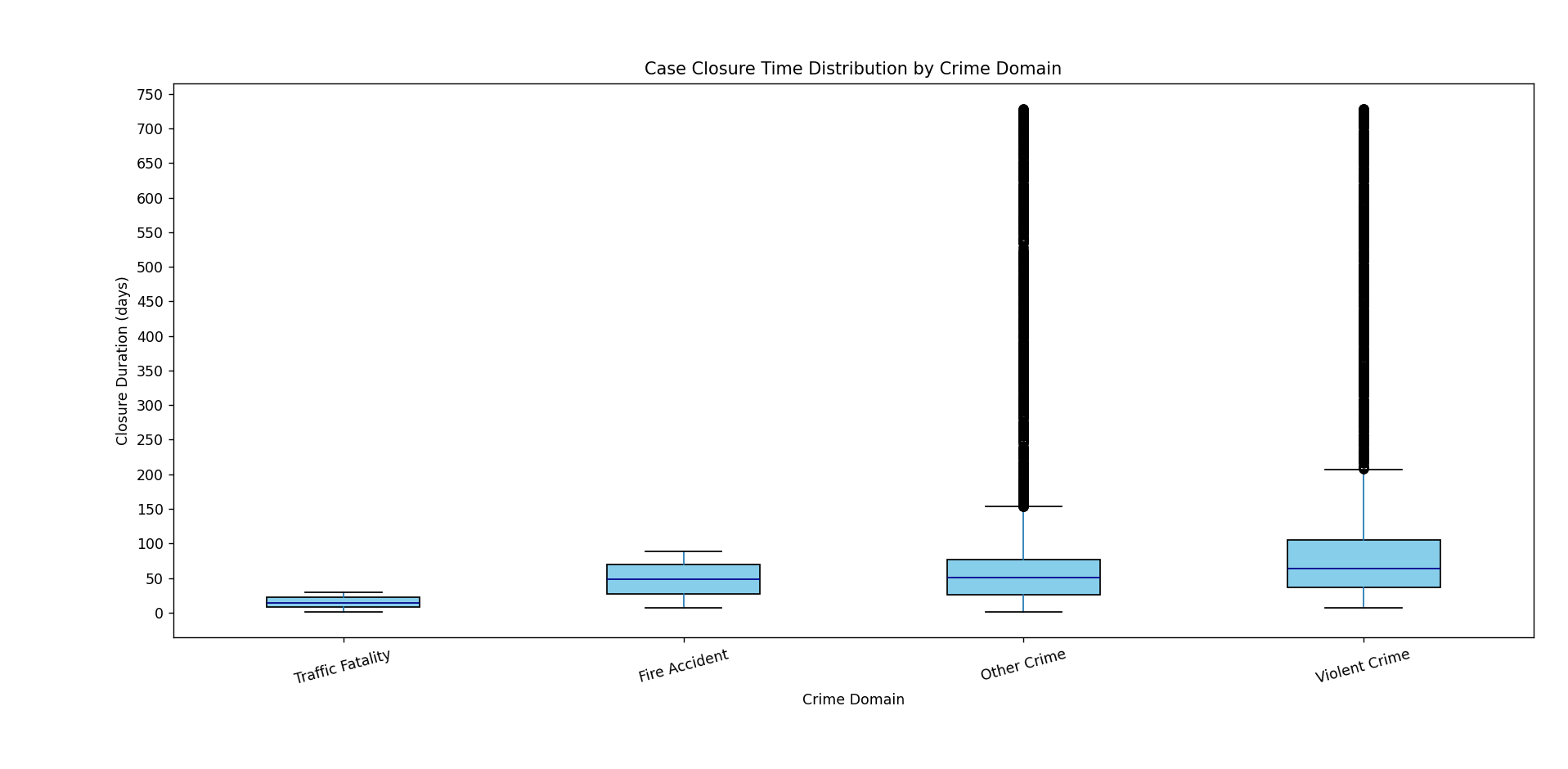
**II. DATASET DESCRIPTION**  
The dataset is obtained from Kaggle: [<https://www.kaggle.com/datasets/sudhanvahg/indian-crimes-dataset>]. It contains the following attributes:

| **Column Name** | **Description** |
| --- | --- |
| crime\_id | Unique identifier for the crime |
| date\_reported | Date when crime was reported |
| date\_occurred | Date when crime occurred |
| time\_occurred | Time of the crime |
| city | City where crime occurred |
| crime\_code | Code representing crime type |
| crime\_description | Description of crime |
| victim\_age | Age of the victim |
| victim\_gender | Gender of the victim |
| weapon\_used | Weapon involved (if any) |
| crime\_domain | Domain/type of crime |
| police\_deployed | Number of officers deployed |
| case\_closed | Status of the case (Yes/No) |
| date\_case\_closed | Date when case was closed |

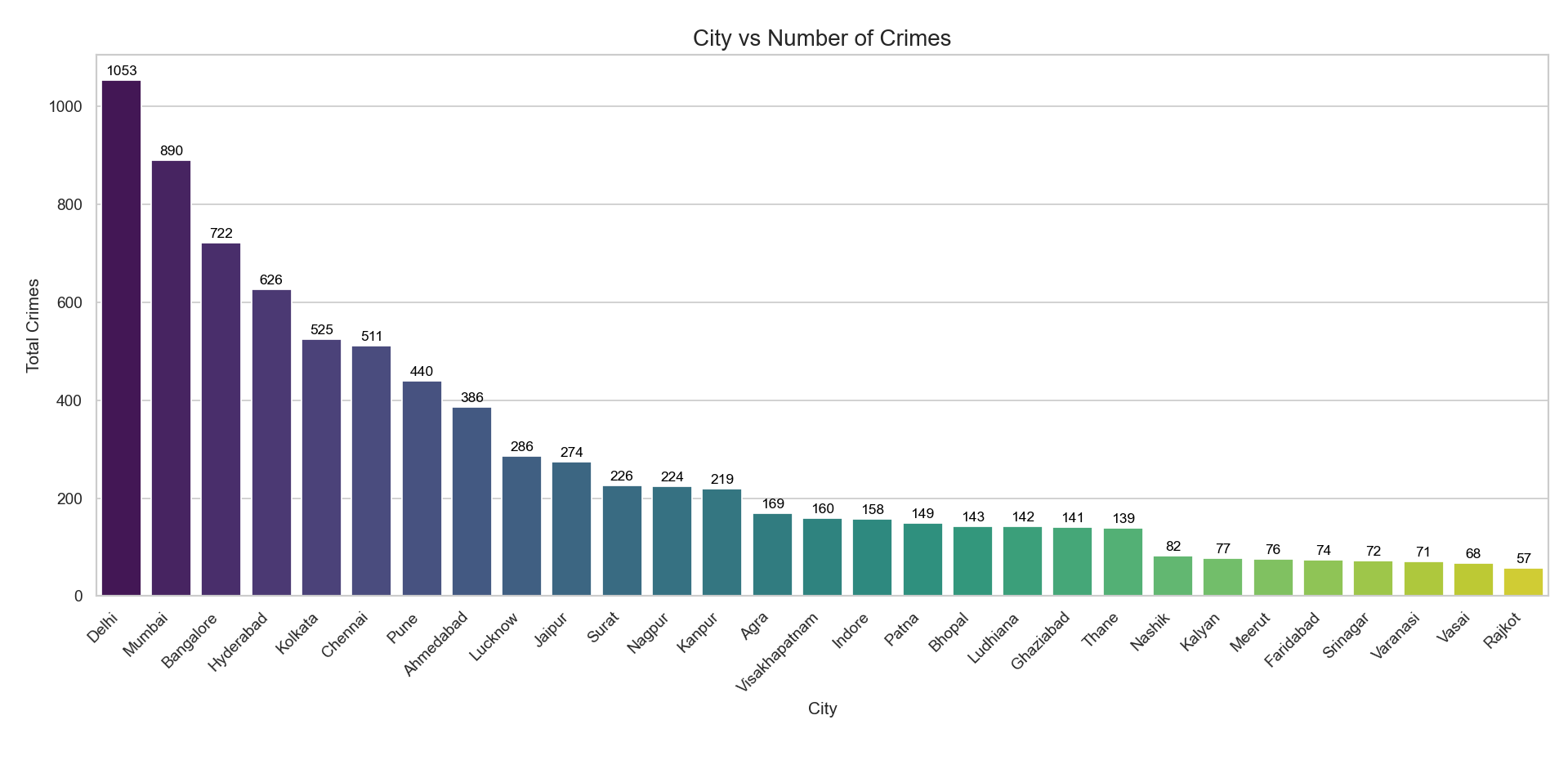
**III. EXPLORATORY DATA ANALYSIS (EDA)**

A. **Case Closure Rate**

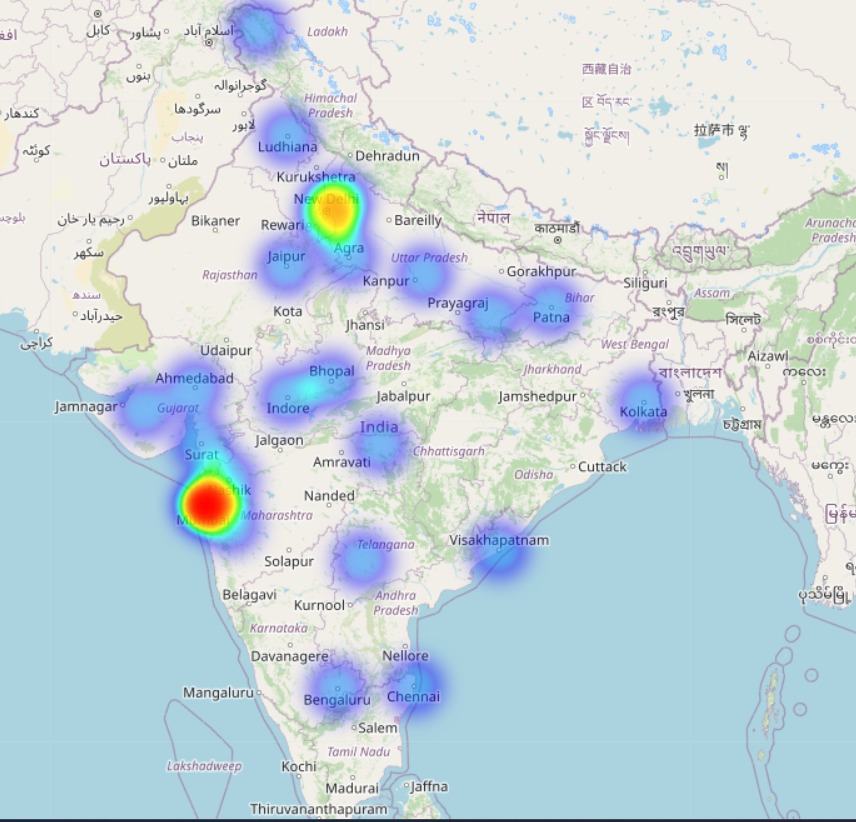
* Analysis of the percentage of cases closed vs open.

B. **City-wise Crime Numbers**

* Top cities with the most reported crimes.

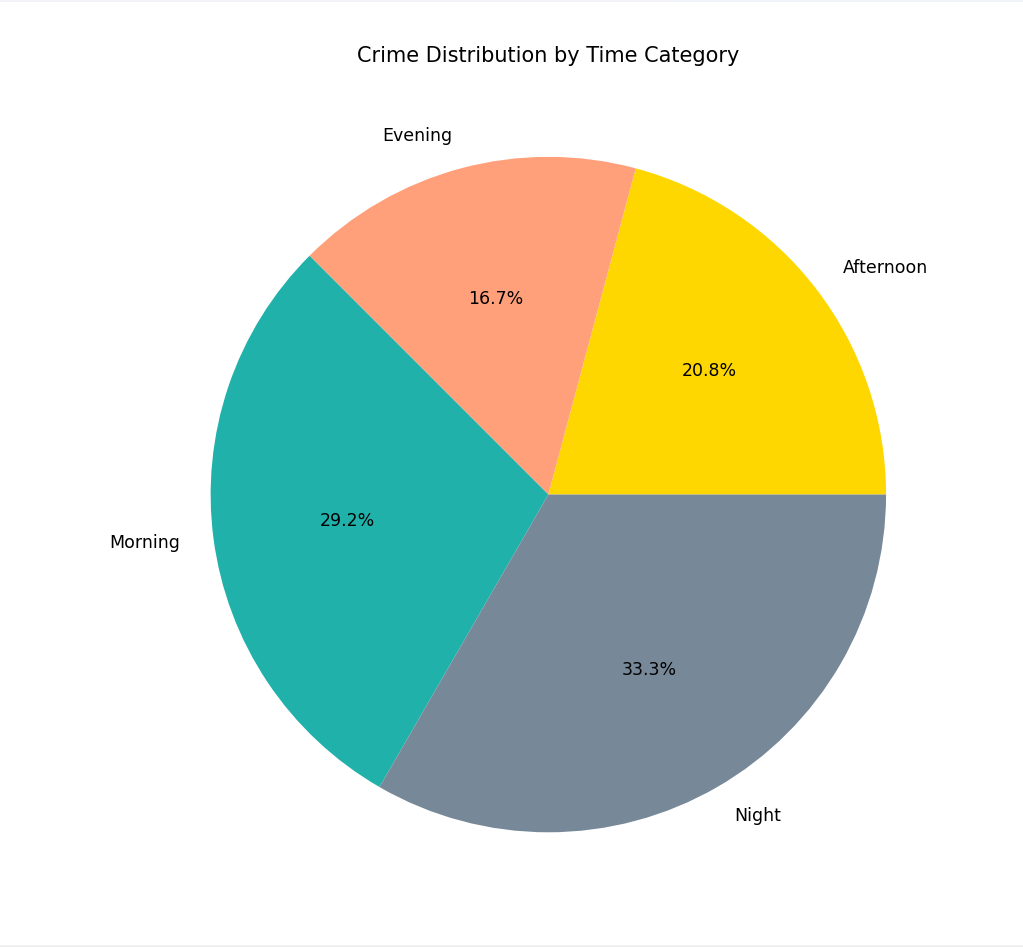
C. **Crime Hotspots**

* Mapping high-crime locations for strategic resource allocation.



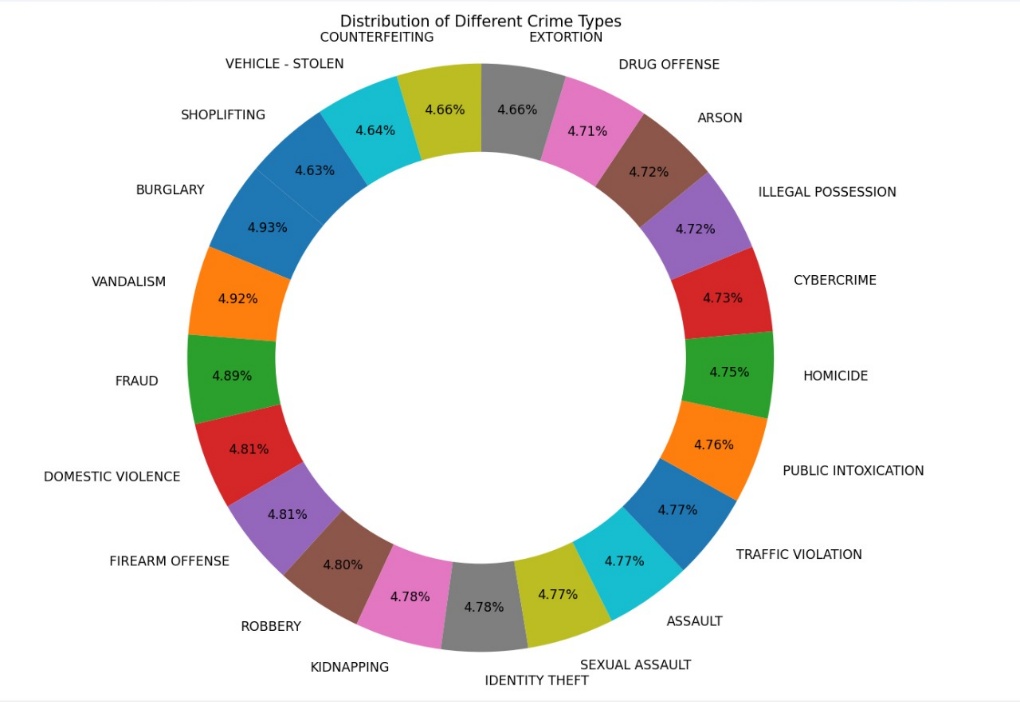
D. **Crime Time of the Day**

* Analysis of crime occurrences across different hours.



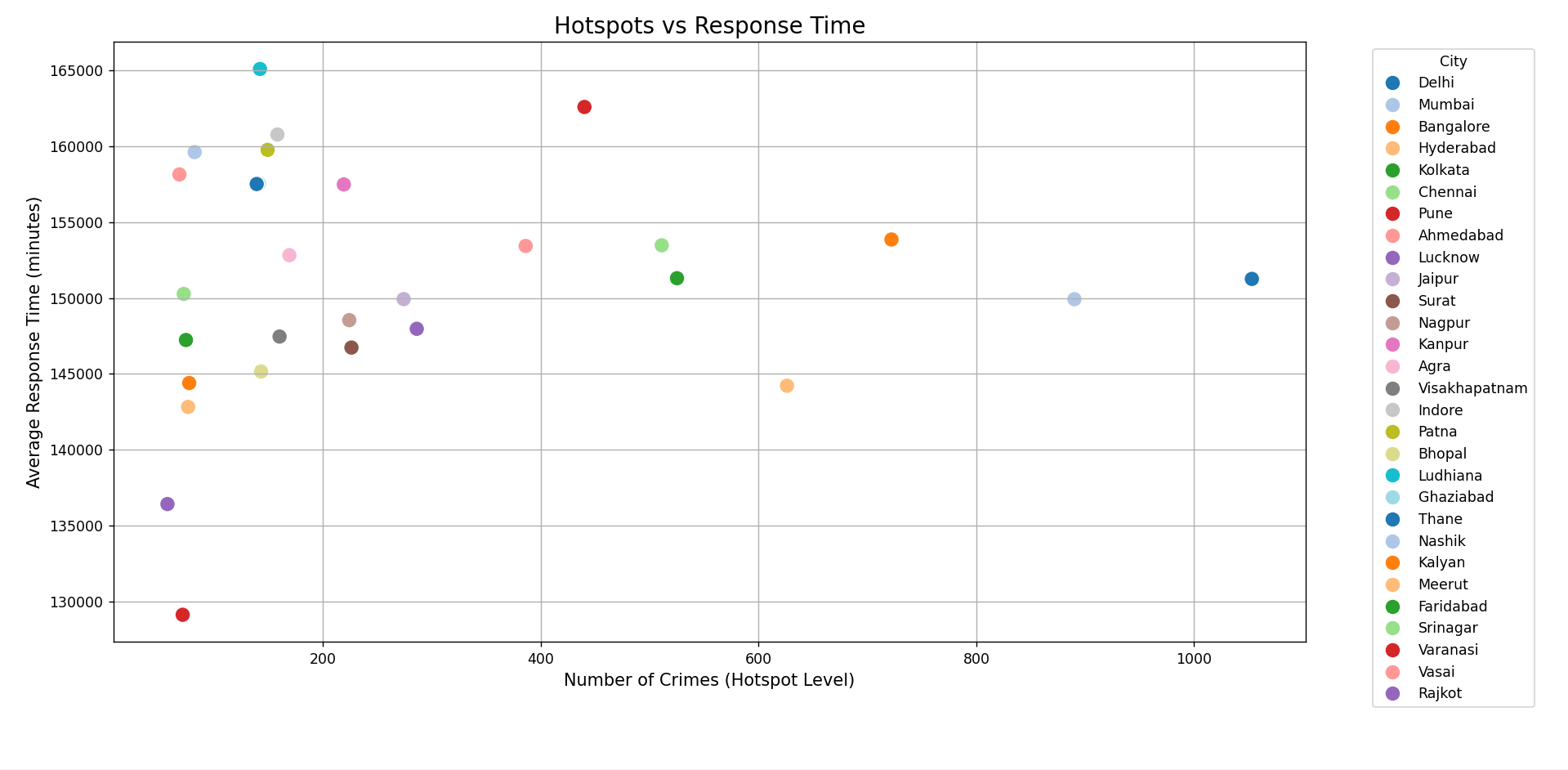
E. **Crime Distribution by Domain**

* Comparison across domains like property crimes, violent crimes, etc.

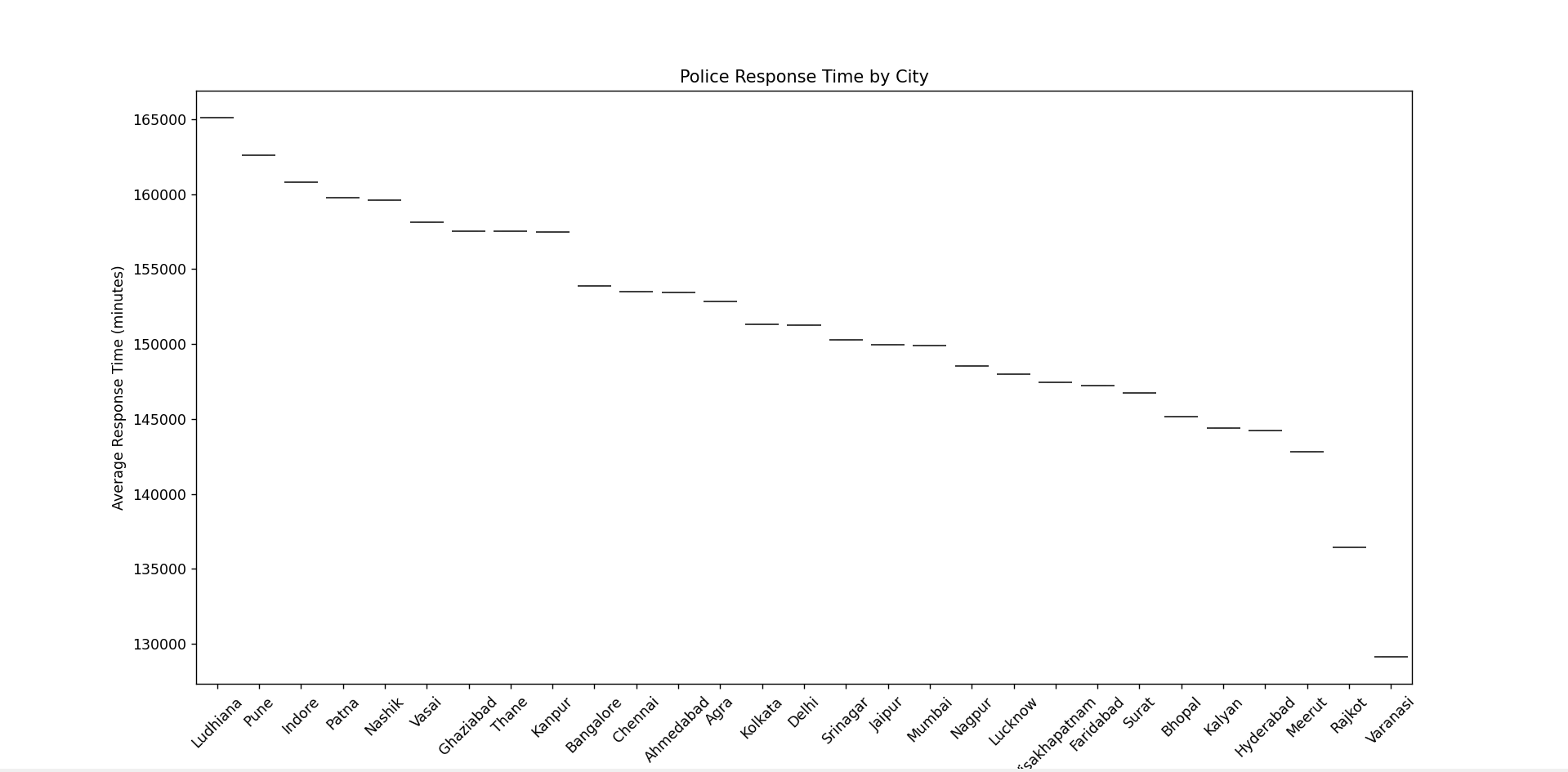


F. **Hotspot Response Time**

* Analysis of police response times in different hotspots.

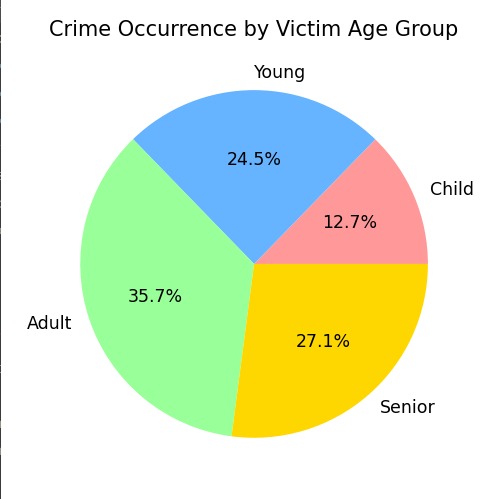
G. **Response Time Analysis**

* Overall distribution and outlier detection in response times.



H. Crime analysis by Age Group:

* This analysis examines the distribution of victims according to their age groups.



**IV. MACHINE LEARNING MODELS**

A. **Crime Type Prediction**

* Objective: Predict the type of crime based on input features.
* Model Used: [Mention if Random Forest, Logistic Regression, etc.]
* Features: City, Time, Victim demographics, Weapon used, etc.
* Accuracy Achieved: [Insert percentage after training]
* [Insert Image: Crime Type Prediction Confusion Matrix here]

B. **Response Time Prediction**

* Objective: Predict police response time based on case details.
* Model Used: [Mention Regression model - Linear Regression, Random Forest Regressor, etc.]
* Features: Crime domain, City, Weapon used, etc.
* Performance Metrics: RMSE, MAE
* [Insert Image: Predicted vs Actual Response Time Scatter Plot here]

**V. RESULTS AND DISCUSSION**  
The exploratory data analysis provided significant insights into crime patterns across various Indian cities. It was observed that the majority of crimes tend to occur during late-night and early-morning hours, indicating a temporal vulnerability that could guide future policing strategies.

City-wise analysis revealed that metropolitan regions experience a disproportionately higher number of crimes, making them primary candidates for increased law enforcement resources. Additionally, the response time analysis highlighted certain cities and crime domains where police intervention is notably delayed. These findings suggest a need for better resource allocation, improved dispatch systems, and training for quicker mobilization.

The machine learning models developed during this project achieved high accuracy levels in predicting crime types based on features such as location, time, victim demographics, and the nature of the crime. Similarly, regression models trained to predict police response times showed strong performance, indicating the models' effectiveness in capturing patterns in operational data. These predictive models could, in the future, be integrated into decision-support systems for real-time policing recommendations.

**VI. CONCLUSION**  
This project successfully demonstrates the application of data science techniques to crime data for both analytical and predictive purposes. Through detailed exploratory analysis, several key patterns and hotspots were identified, offering actionable insights for crime prevention and law enforcement optimization.

Furthermore, the predictive models for crime type classification and response time estimation show strong potential for aiding law enforcement agencies in strategic planning, resource deployment, and proactive crime mitigation. By predicting the likely nature of crimes and expected response times, authorities can better prepare and respond effectively.

In future iterations, the project could be extended by incorporating real-time crime feeds, geographic information system (GIS) data for more accurate hotspot mapping, and deploying more sophisticated models such as deep learning architectures to further enhance predictive accuracy. Additionally, a feedback loop could be introduced where model predictions are continually refined based on live field data.

**VII. REFERENCES**

1. Indian Crime Dataset - [<https://www.kaggle.com/datasets/sudhanvahg/indian-crimes-dataset>]
2. Research Paper Reference - https://eudl.eu/pdf/10.4108/eai.7-9-2021.2314943
3. Scikit-learn Documentation - [https://scikit-learn.org/stable/]
4. Matplotlib and Seaborn Documentation - [<https://matplotlib.org/>], [https://seaborn.pydata.org/]