



# Crime Pattern Analysis in Indian Cities Using Data Visualization Techniques



## Project Details:

**Sector:** Urban Safety and Crime Analytics

**Course:** Data Visualization & Analytics

**Team:** Bhargav Patil (2401020092)  
Ayush Tiwari (2401010121)  
Saumya Soni (2401020058)  
Aparna Singh (2401020089)  
Priyanshu (2401010356)  
Parv (2401010322)

**Institution:** Newton School of Technology

**Date:** 18 February 2026

## 2. Executive Summary

Urban crime in Indian cities varies significantly across locations, crime categories, victim groups, and time periods, posing challenges for effective public safety planning and law-enforcement resource allocation. This project analyzes crime distribution patterns across Indian cities to identify spatial concentration, dominant crime domains, and vulnerable population groups, with the objective of supporting data-driven policing strategies and urban safety planning.

The analysis is based on a structured dataset of over 40,000 reported crime incidents across 29 Indian cities. Following preprocessing, temporal standardization, and categorical normalization, a cleaned analytical subset of 7,891 records was prepared for detailed analysis. Pivot-based aggregation and cross-tabulation in Google Sheets were used to examine crime patterns across cities, crime domains, victim demographics, time of occurrence, and investigation outcomes. An interactive dashboard was developed to enable multi-dimensional visualization of crime distribution.

The results show that crime incidents are strongly concentrated in major metropolitan cities, particularly Delhi, Mumbai, Bangalore, and Hyderabad. Across all cities, *Other Crime* constitutes the largest category (57.2%), followed by *Violent Crime* (28.8%), while *Fire Accidents* (9.35%) and *Traffic Fatalities* (4.65%) represent smaller shares. Gender analysis indicates that females account for 55.6% of victims, with the highest proportion observed within violent crime (56.6%), suggesting elevated gender vulnerability in serious incidents. Temporal patterns reveal peak crime occurrence during night periods. Investigation outcomes show a near-equal distribution between solved (50.3%) and active cases (49.7%), indicating moderate case-closure levels.

Overall, the findings demonstrate that urban crime in India is concentrated in large cities, specific crime domains, and vulnerable victim groups. The study highlights the need for city-specific policing strategies, gender-focused protection measures, and time-based deployment of law-enforcement resources. It illustrates how structured data visualization and pivot-based analysis can convert large administrative crime datasets into actionable insights for urban crime management and policy decision-making.

## 3. Sector & Business Context

### 3.1 Sector Overview

Urban crime and public safety management are major governance challenges in rapidly growing Indian cities. Urbanization, population density, mobility, and socio-economic diversity create complex crime patterns that vary across locations, crime types, and population groups. Law-enforcement agencies must allocate limited resources across multiple urban regions and categories, making it essential to understand where crimes are concentrated and which groups are most vulnerable.

### 3.2 Current Challenges

In India, crime monitoring traditionally relies on aggregated city or state statistics, which provide overall counts but often do not reveal patterns across crime domains, demographics, and time periods. This limits the ability of law-enforcement agencies to identify high-risk locations, dominant crime categories, vulnerable victim groups, and temporal hotspots, reducing the effectiveness of targeted policing and resource allocation.

### 3.3 Why This Problem Was Chosen

This project analyzes crime distribution across Indian cities using structured historical crime data. Understanding variation across cities, domains, victim groups, and time periods helps identify priority areas for law-enforcement attention and urban safety planning. By applying pivot-based aggregation and visualization techniques, the study demonstrates how analytical methods can transform administrative crime data into actionable insights for policing and governance decisions.

## 4. Problem Statement & Objectives

Urban crime presents significant challenges for public safety management and law-enforcement planning in rapidly growing Indian cities. Effective policing requires a data-driven understanding of where crimes are concentrated, which categories dominate, and which population groups are most vulnerable. However, crime data is often reported in

aggregated form, limiting the identification of structural patterns across locations, crime domains, and time periods and reducing the effectiveness of targeted policing and resource allocation.

This project analyzes historical crime data across Indian cities to understand distribution patterns across crime categories, victim groups, and temporal cycles. By transforming structured crime records into analytical insights, the study aims to support evidence-based policing strategies, resource prioritization, and urban safety planning.

## 4.1 Core Problem Statement

To analyze crime distribution across Indian cities and identify patterns in high-risk locations, crime domains, and vulnerable victim groups to support data-driven law-enforcement and public safety decisions.

## 4.2 Project Scope

The study focuses on reported crime incidents across 29 Indian cities and examines distribution patterns across crime domains, victim demographics, temporal occurrence, and case resolution attributes using structured historical crime data. The analysis is limited to descriptive and comparative pattern identification using pivot-based aggregation and visualization techniques.

## 4.3 Objectives of the Study

- To examine crime distribution across Indian cities and identify high-concentration urban regions
- To analyze crime domain shares and determine dominant categories
- To evaluate victim gender and group patterns to identify vulnerable populations
- To assess temporal variation in crime occurrence across periods and cities
- To examine case resolution patterns across crime types
- To generate insights supporting targeted policing and urban safety planning

## 4.4 Success Criteria

The project is considered successful if clear spatial, demographic, and temporal crime patterns are identified across cities and crime domains, and if actionable insights are generated that support targeted policing strategies and urban safety planning decisions.

# 5. Data Description

The analysis uses a structured historical crime dataset containing 40,160 reported incidents across 29 Indian cities, covering 5 crime domains and 22 crime categories. The dataset was sourced from Kaggle (*Indian Crimes Dataset*, Sudhanva HG, 2023) and is publicly available at:

<https://www.kaggle.com/datasets/sudhanvahg/indian-crimes-dataset>

Each record represents an individual crime case with attributes describing location, time of occurrence, victim characteristics, crime classification, and investigation status, enabling spatial, demographic, and temporal analysis of crime patterns.

For analytical consistency, a cleaned subset of 7,891 records was prepared through preprocessing and standardization. The dataset includes categorical (city, crime domain, victim gender), temporal (date and time attributes), and numerical (victim age, police deployed) variables supporting multi-dimensional aggregation and visualization.

## 5.1 Columns Explanation

The dataset contains fields such as report number, city, crime domain, crime description, victim age and gender, time and date of occurrence, police deployed, and case closure information, capturing geographic, temporal, demographic, and investigative characteristics of each incident.

## 5.2 Data Size

- Original dataset: 40,160 records
- Analytical subset: 7,891 records
- Coverage: 29 cities, 5 crime domains, 22 categories

## 5.3 Data Limitations

The dataset reflects reported crimes and may exclude unreported incidents. Some attributes contained missing values requiring cleaning, and contextual factors such as socio-economic conditions or policing intensity are not included, limiting causal interpretation.

# 6. Data Cleaning & Preparation

The original crime dataset required preprocessing to ensure consistency and analytical suitability for pivot-based aggregation and visualization. All primary data cleaning and transformation steps were performed in Google Sheets, supported by initial structuring in Python. From the initial 40,160 records, a cleaned analytical subset of 7,891 records was prepared by removing incomplete entries and standardizing categorical and temporal variables. Approximately 32,269 records were excluded due to missing or inconsistent key attributes required for analysis.

## 6.1 Data Cleaning Procedures

Categorical fields such as city names, crime domains, and crime descriptions were standardized to ensure consistent grouping across records. Missing values in investigation attributes were treated and formatted during preprocessing. In particular, the *Date Case Closed* field contained approximately 50% missing entries, representing ongoing or unresolved cases; these were recoded as "Not Closed" to enable consistent derivation of case status and resolution metrics.

## 6.2 Date and Time Standardization

Temporal variables were converted into uniform formats to support reliable time-based analysis. The *Date of Occurrence*, *Date Reported*, and *Date Case Closed* fields were standardized into consistent date formats, and *Time of Occurrence* was standardized into a uniform time format. This enabled accurate extraction of month, year, and time-of-day attributes for temporal grouping and comparison.

## 6.3 Derived and Engineered Variables

Additional analytical variables were created to support demographic, temporal, and resolution-based analysis. These included victim group (age categories), resolution days, reporting delay, resolution category (fast < 30 days, moderate 30–90 days, slow > 90 days, unresolved), time-of-crime period, and month and year of crime. These derived fields enabled structured analysis of crime patterns across cities, victim groups, time periods, and investigation outcomes.

## 6.4 Preparation of Analytical Subset

The filtered analytical subset of 7,891 records formed the final dataset used for exploratory analysis, visualization, and dashboard development. Through these preprocessing steps, the raw administrative dataset was transformed into a structured analytical dataset suitable for multi-dimensional crime pattern analysis.

# 7. KPI & Metric Framework

To evaluate investigation outcomes, crime trends, and policing resource allocation, a set of key performance indicators (KPIs) was defined using aggregated crime records from the analytical dataset. These KPIs quantify crime volume, investigation status, resolution efficiency, and policing effort, supporting assessment of crime patterns and law-enforcement performance in line with the study objectives.

## 7.1 Total Registered Cases

**Definition:** Total number of crime incidents in the analytical dataset.

**Formula:** Count of reported crime records.

**Why it matters:** Establishes overall crime scale for all comparative metrics.

**Objective linkage:** Crime distribution across cities and domains.

## 7.2 Year-on-Year Crime Growth Rate (%)

**Definition:** Percentage change in total crime cases between consecutive years.

**Formula:**  $(\text{Crimes\_current} - \text{Crimes\_previous}) / \text{Crimes\_previous} \times 100$

**Why it matters:** Indicates temporal increase or decline in crime levels.

**Objective linkage:** Temporal variation in crime occurrence.

## 7.3 Ongoing Investigations

**Definition:** Number of cases that remain active or unresolved.

**Formula:** Count of cases with status = Not Closed.

**Why it matters:** Reflects investigative workload and unresolved crime volume.

**Objective linkage:** Case resolution patterns.

## 7.4 Solved Cases

**Definition:** Number of cases marked as closed.

**Formula:** Count of cases with status = Closed.

**Why it matters:** Indicates completed investigations.

**Objective linkage:** Case resolution patterns.

## 7.5 Solve Rate (%)

**Definition:** Proportion of cases successfully closed.

**Formula:**  $\text{Solved Cases} / \text{Total Registered Cases} \times 100$

**Why it matters:** Measures investigation effectiveness.

**Objective linkage:** Investigation efficiency assessment.

## 7.6 Average Resolution Time (Days)

**Definition:** Average duration from occurrence to case closure.

**Formula:** Mean (Resolution Days).

**Why it matters:** Indicates investigation speed and efficiency.

**Objective linkage:** Case resolution performance.

## 7.7 Average Police Allocated per Crime

**Definition:** Average number of police personnel deployed per case.

**Formula:**  $\text{Total Police Deployed} / \text{Total Registered Cases}$

**Why it matters:** Reflects policing effort and resource allocation intensity.

**Objective linkage:** Resource allocation analysis.

Together, these KPIs summarize crime volume, temporal change, investigation status, resolution efficiency, and policing effort, directly supporting the analytical objectives of spatial, temporal, and investigative crime pattern assessment.

## 8. Exploratory Data Analysis

Exploratory data analysis was conducted to examine crime distribution across domains, cities, victim groups, time periods, and investigation outcomes using pivot-based aggregation and percentage analysis. Cross-tabulated analysis was also used to examine relationships between variables such as gender and crime domain and city and crime volume. The analysis reveals clear patterns in crime structure, geographic concentration, demographic vulnerability, temporal occurrence, and case resolution.

### 8.1 Crime Distribution by Domain

Crime incidents are unevenly distributed across categories. *Other Crime* constitutes the largest share (57.2%), followed by *Violent Crime* (28.8%), while *Fire Accidents* (9.35%) and *Traffic Fatalities* (4.65%) contribute smaller proportions. This indicates that non-violent offenses dominate overall crime patterns, although violent crime remains a substantial component of urban incidents.

### 8.2 Crime Distribution by City

Crime is strongly concentrated in major metropolitan regions. Delhi records the highest number of incidents, followed by other large cities such as Mumbai, Bangalore, and Hyderabad, while smaller cities contribute comparatively lower volumes. This pattern reflects higher crime concentration in large urban centers with greater population density and activity levels.

### 8.3 Gender Distribution of Victims

Gender analysis shows that females account for 55.6% of victims, exceeding male victims (33.6%). This indicates significant gender vulnerability across reported crime incidents in the dataset.

### 8.4 Gender within Crime Domains

Female victimization is highest within violent crime (56.6%), exceeding the overall female share. This suggests elevated gender exposure in serious interpersonal crime categories and indicates that gender vulnerability is particularly pronounced in violent incidents.

### 8.5 Temporal Distribution of Crime

Crime occurrence varies by time of day, with the night period showing the highest incidence across cities. This indicates temporal concentration during lower-visibility hours and highlights the importance of time-based policing strategies.

### 8.6 Investigation Status and Resolution

Investigation analysis shows a near-equal distribution between solved cases (3,971; 50.33%) and ongoing investigations (3,919; 49.7%). The average resolution time of 90.68 days indicates moderate investigation duration across incidents.

### 8.7 Police Deployment Patterns

Police allocation varies across incidents, reflecting differences in case complexity and investigative requirements. The average police deployment per crime provides an indicator of operational response intensity across crime categories and cities.

## 9. Advanced Analysis

Additional analysis examined segmentation and risk concentration across cities, crime domains, victim groups, and time periods. Cross-tabulated patterns show that major metropolitan cities account for the highest crime volumes across most domains, indicating spatial concentration in large urban centers.

### 9.1 Spatial Risk Concentration

Cross-tabulated analysis shows that major metropolitan cities account for the highest crime volumes across most crime domains, indicating strong spatial concentration in large urban centers.

## 9.2 Demographic Risk Segmentation

Demographic segmentation reveals that female victims constitute the majority of incidents and show elevated exposure within violent crime categories, indicating demographic risk concentration.

## 9.3 Temporal Risk Patterns

Temporal analysis shows consistent crime peaks during night periods across cities, suggesting systematic time-based risk conditions.

## 9.4 Investigation Outcome Patterns

Investigation patterns indicate a near-equal distribution between solved and active cases with moderate resolution duration, reflecting balanced but constrained investigative outcomes.

Overall, combined spatial, demographic, and temporal patterns indicate that crime risk is highest in large cities, during night periods, and within violent incidents affecting female victims.

# 10. Dashboard Design

The crime analytics dashboard was implemented in Google Sheets using pivot tables, calculated KPIs, formulas, and interactive filters to enable multi-dimensional exploration of crime distribution patterns across Indian cities. The dashboard integrates aggregated metrics and visualizations derived from the cleaned analytical dataset and supports interactive analysis of crime patterns across spatial, demographic, temporal, and investigative dimensions.

## 10.1 Dashboard Objective

The dashboard provides a consolidated view of key crime indicators to support rapid interpretation of spatial, demographic, temporal, and investigative patterns. It enables identification of high-crime cities, dominant crime domains, vulnerable victim groups, peak crime periods, and case resolution status, allowing users to quickly detect concentration patterns and operational priorities.

## 10.2 View Structure

The dashboard is organized into a KPI summary section and supporting visualizations. The top section displays core indicators including total registered cases, solved cases, ongoing investigations, solve rate, and average resolution time, providing an overall performance snapshot. The visualization section presents charts showing crime distribution by city, crime domain, victim gender, time of occurrence, and investigation status, enabling comparative pattern analysis across key variables.

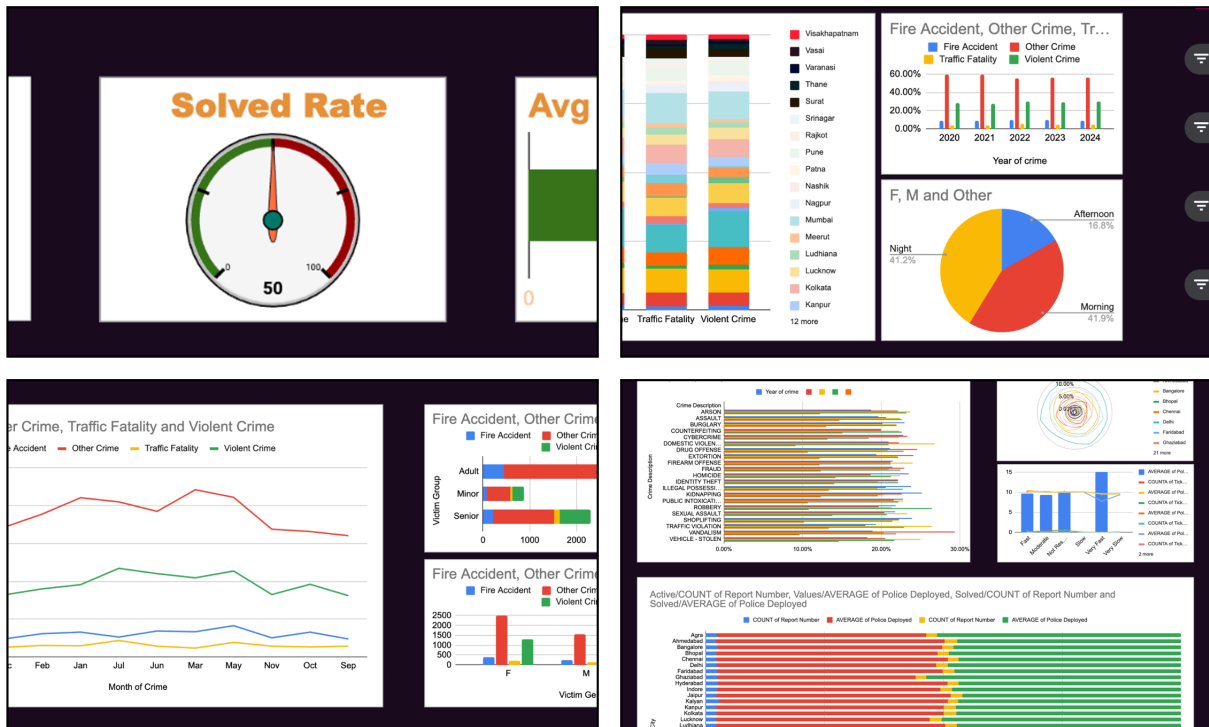
## 10.3 Filters and Drilldowns

Interactive filters are provided for city, crime domain, time of crime, victim gender, and resolution category, allowing users to dynamically examine patterns across data subsets. Pivot-based drilldowns enable comparison of crime distribution across locations, categories, and time periods, while KPI values automatically update with filter selection to reflect context-specific insights.

## 10.4 Screenshots and Interpretation

The dashboard view is shown in the figure below. The layout enables intuitive identification of crime hotspots, dominant categories, demographic vulnerability, and investigation outcomes, supporting data-driven interpretation and decision-oriented analysis of urban crime patterns.

# Crime Analytics Dashboard Overview



## 11. Insights Summary

The analysis of crime patterns across Indian cities reveals key spatial, structural, demographic, temporal, and investigative insights relevant for urban safety planning and policing strategy.

### 11.1 Urban Crime Concentration

Crime incidents are concentrated in major metropolitan regions, particularly Delhi, Mumbai, and Bangalore, indicating that policing resources and preventive interventions should be prioritized in large urban centers.

### 11.2 Dominance of Non-Violent Crime

The crime structure is dominated by *Other Crime* (57.2%), with violent crime (28.8%) forming a substantial secondary share, suggesting that enforcement strategies should address both high-volume non-violent offenses and serious interpersonal crime.

### 11.3 Elevated Female Victimization

Females constitute 55.6% of total victims and 56.6% of violent crime victims, indicating the need for gender-focused safety measures and targeted protection strategies in urban crime prevention.

### 11.4 Night-Time Risk Concentration

Crime occurrence peaks during night hours, indicating that surveillance, patrol deployment, and preventive policing should be strengthened during low-visibility periods.

### 11.5 Balanced but Moderate Investigation Closure

Case resolution shows a near-equal distribution between solved (50.33%) and active cases (49.7%), suggesting moderate investigation effectiveness and highlighting the need to improve closure rates.

### 11.6 Moderate Resolution Duration

The average resolution time of 90.68 days indicates that investigations typically require a moderate duration, suggesting potential opportunities to streamline investigative processes.

### 11.7 Variation in Police Deployment

Differences in police allocation per incident indicate variability in investigative effort and case complexity, suggesting the need for more standardized resource allocation frameworks.

### 11.8 Urban-Centric Crime Risk Profile

Overall patterns indicate that crime risk in India is concentrated in large cities, night periods, and incidents involving female victims, highlighting priority areas for targeted policing and urban safety planning.

## 12. Recommendations

The following recommendations are derived from the analytical insights on spatial, demographic, temporal, and investigative crime patterns. Each recommendation links observed crime patterns with actionable policing or governance interventions and expected business and public safety impact.

Recommendation	Linked Insight	Business Impact	Feasibility
Targeted policing in high-crime cities	Crime concentrated in major metropolitan regions	More effective deployment in high-risk areas; reduction in urban crime hotspots	High – requires resource prioritization rather than new infrastructure
Enhanced protection for female victims	Females form majority of victims, especially in violent crime	Improved public safety perception; reduction in gender-based victimization	Medium-High – requires patrol planning and safety programs
Night-time surveillance intensification	Crime peaks during night hours	Reduction in night-time incidents; improved deterrence	High – patrol timing and lighting adjustments feasible
Focused intervention for violent crime	Violent crime significant with high female exposure	Reduction in serious crime severity and victim harm	Medium – requires specialized units and training
Improved case resolution efficiency	Moderate solve rate and resolution time	Faster case closure; improved justice outcomes	Medium-High – digital tracking and workflow optimization feasible
Data-driven policing and monitoring	Crime varies across cities, domains, and time	Proactive policing; optimized deployment decisions	Medium – requires analytics integration
Community reporting strengthening	Reported crime patterns depend on reporting	Better incident detection and prevention	Medium – requires outreach systems
Integration with urban safety planning	Crime linked to urban environment factors	Long-term reduction in urban crime risk	Medium-Low – cross-agency coordination needed

## 13. Impact Estimation

Implementation of the proposed recommendations is expected to improve crime control, investigation efficiency, and urban safety outcomes across analyzed cities.

### 13.1 Cost Savings

Targeted deployment in high-crime metropolitan regions and data-driven resource allocation can reduce redundant patrol coverage and optimize personnel utilization, lowering operational costs associated with inefficient deployment.

### 13.2 Efficiency Improvement

Digital case tracking, prioritization frameworks, and focused investigation of violent crime are expected to improve closure rates and reduce the current ~90-day average resolution period, enhancing investigative throughput.

### 13.3 Service Improvement

Gender-focused safety measures and night-time surveillance strengthening can improve public safety perception, citizen protection, and responsiveness of policing services in high-risk locations and periods.

### 13.4 Risk Reduction

Intensified patrols in crime hotspots, enhanced protection for female victims, and time-based deployment strategies are expected to reduce incident concentration in major cities, decrease night-time crime occurrence, and mitigate vulnerability in violent crime contexts.

Overall, these measures are expected to enhance policing effectiveness, investigation performance, and urban safety through data-driven resource allocation and targeted intervention strategies.

## 14. Limitations

The analysis is based on reported crime incidents and may not capture unreported or undocumented cases, which can affect representation of actual crime levels. Reporting practices and data recording standards may also vary across cities, influencing comparability.

Some attributes in the original dataset contained missing values, particularly case closure information, which required preprocessing and standardization. The dataset includes incident-level characteristics but lacks contextual factors such as socio-economic conditions or policing intensity; therefore, the analysis identifies distribution patterns rather than causal relationships.

## 15. Future Scope

Future analysis can extend the study by incorporating longitudinal trend analysis across years to examine changes in crime patterns over time and evaluate growth or decline in specific crime domains and cities. More granular spatial analysis at neighborhood or zone level could improve identification of localized crime hotspots.

Integration of additional contextual datasets such as population density, socio-economic indicators, policing resources, and urban infrastructure variables would enable causal analysis and risk modeling of crime occurrence. Predictive modeling approaches could also be applied to forecast crime risk across locations and time periods.

Development of automated dashboards and real-time data integration from policing systems would further enhance continuous monitoring and decision support for urban crime management.

## 17. Appendix

### 17.1 Data Dictionary

Column Name	Data Type	Unit	Description
Report Number	Integer	ID	Unique identifier for each crime case
Date Reported	DateTime	YYYY-MM-DD HH:MM	Date and time when crime was reported
Date of Occurrence	Date	YYYY-MM-DD	Actual date of crime occurrence
Time of Occurrence	Time	HH:MM	Time when crime occurred
City	Text (Categorical)	NA	City where crime occurred
Crime Code	Integer	Code	Numeric code representing crime type
Crime Description	Text (Categorical)	NA	Specific crime category
Victim Age	Integer	Years	Age of victim

Column Name	Data Type	Unit	Description
Victim Gender	Text (Categorical)	NA	Gender of victim
Weapon Used	Text (Categorical)	NA	Weapon or instrument involved
Crime Domain	Text (Categorical)	NA	Broad crime category grouping
Police Deployed	Integer	Count	Number of officers assigned
Ticket Status	Text (Categorical)	NA	Investigation status (Active/Solved)
Date Case Closed	Date	YYYY-MM-DD	Date of case closure
Victim Group	Text (Categorical)	NA	Age group (Minor/Adult/Senior)
Resolution Category	Text (Categorical)	NA	Closure speed classification
Time Of Crime	Text (Categorical)	NA	Time period (Morning/Afternoon/Night)
Month of Crime	Integer	Month	Extracted month of occurrence
Year of Crime	Integer	Year	Extracted year of occurrence
Resolution_Days	Integer	Days	Days between occurrence and closure
Reporting_Delay	Integer	Days	Days between occurrence and reporting

## 17.2 SQL / Python Logic

Initial dataset structuring and extraction from the source file were supported using Python, followed by primary preprocessing and transformation in Google Sheets. Key preprocessing steps included record extraction, missing-value treatment, date standardization, categorical normalization, and creation of derived analytical variables used in pivot analysis and dashboard metrics.

## 18. Contribution Matrix

Team Member	Dataset & Sourcing	Cleaning	KPI & Analysis	Dashboard	Report Writing	PPT	Overall Role
Bhargav Patil	✓		✓				STRATEGY LEAD
Ayush Tiwari	✓					✓	PPT AND QUALITY LEAD
Saumya Soni		✓	✓	✓			ANALYSIS AND DASHBOARD LEAD
Aparna Singh	✓				✓		PROJECT LEAD
Priyanshu		✓					DATA LEAD
Parv						✓	PPT

**Declaration:** We confirm that the above contribution details are accurate and verifiable through version history and submitted artifacts.

**Team Signature Block:** \_\_\_\_\_