# Crimes Data Analysis Report

## 1. Dataset Description

### 1.1 Source

Data taken from the National Crime Records Bureau (NCRB) dataset on *Crimes Against Children by State and District*, available on data.gov.in.

### 1.2 Columns

- State/UT – Indian State or Union Territory.
- State/UT/District – Corresponding district within each state.
- Murder (Sec.302 IPC) – Reported murders involving children.
- Cyber Crimes (IT Act) – Reported cybercrimes or online exploitation cases against children.
- Other IPC & SLL Crimes – Includes offenses like abetment of suicide, kidnapping, trafficking, and child marriage.
- Total Crimes against Children (IPC + SLL) – Combined total of all reported crimes per district.

### 1.3 Data Quality

- Contains 970 records across 72 columns.
- No missing or corrupted values.
- Columns are numerically consistent, ready for aggregation.
- Data is clean and well-structured for analysis at state and district levels.

## 2. Operations Performed

### 2.1 Data Cleaning & Exploration

- Verified column types and ensured all numeric fields were correctly parsed.
- Confirmed absence of nulls or inconsistencies.
- Checked for duplicate district entries — none found.
- Aggregated district-level data to compute state-wise totals for comparison.

### 2.2 Descriptive Analytics

- Total Crimes Distribution visualized with log-scaled histograms.
- Murder & Cybercrime Distribution studied using separate histograms.

- Average Murders and Cyber Crimes by State represented with bar and line charts.
- Total Crimes by State (Log-Scale) compared via horizontal bar charts.
- District Count by State visualized using pie charts to show geographic coverage.
- Crime Spread by State visualized through boxplots (log-scaled).

### 2.3 Relationship Analysis

- Murders vs. Cyber Crimes: Scatter and hexbin plots revealed no strong linear correlation — states with more murders don't necessarily have more cybercrimes.
- State-wise Total vs. Murder Count: Outliers indicate few states contributing disproportionately high crime numbers.
- Above-Average States: Identified states exceeding the mean total crimes value within the dataset.

## 3. Key Insights

### 3.1 Crime Distribution Overview

- Major disparities across states — some states report over 1,000 total child-related crimes, while others report fewer than 50.
- The distribution is highly right-skewed, meaning a small number of states account for most of the crimes.
- Median total crimes per district is under 100, while a few large states exceed 1,000+ cases.

### 3.2 State-Wise Highlights

- Maharashtra, Uttar Pradesh, Madhya Pradesh, and West Bengal show the highest cumulative totals for crimes against children.
- Murder rates per district are generally low (mostly 0–10 cases), but larger states contribute higher absolute counts.
- Cybercrimes against children remain rare but are increasing in high-population states.

### 3.3 Crime Categories

- Murder (Sec.302 IPC): Concentrated in industrial and densely populated states.
- Cyber Crimes: More prominent in technologically advanced or urbanized regions.
- SLL Crimes (Child Marriage, Labour, Trafficking): Detected mostly in rural-oriented states with weaker enforcement mechanisms.

### 3.4 Trends Identified

- States with high urbanization exhibit increased cyber-related offenses.
- States with larger populations or administrative divisions have higher aggregate totals, suggesting that raw counts must be normalized for accurate comparison.
- Some states show consistently high values across multiple categories, hinting at systemic reporting accuracy or deeper socio-economic issues.

## 4. Recommendations

### 4.1 Strengthen Law Enforcement & Reporting

- Introduce specialized child-protection units in high-crime states.
- Encourage digital literacy and cyber awareness among children and guardians.
- Standardize reporting procedures across all districts to ensure data consistency.

### 4.2 Policy & Governance

- Use state-level insights to prioritize resource allocation in policing, legal aid, and counselling.
- Develop inter-state cooperation frameworks for human trafficking and cyber exploitation cases.
- Launch awareness drives in states showing recurring high SLL crime rates.

### 4.3 Data-Driven Decision Making

- Build predictive analytics models to forecast potential crime hotspots using demographic and socio-economic indicators.
- Implement trend dashboards for continuous monitoring of year-wise changes.
- Normalize crime data by population to derive crime rate per 100,000 children — a fairer comparison metric.

### 4.4 Future Analytical Scope

- Extend this analysis into time-series prediction once multi-year NCRB data is integrated.
- Perform clustering of states by crime profile (murder-dominant, cyber-dominant, SLL-dominant).
- Create geo-visual heatmaps for public awareness and policy formulation.