EDA Patrol_ DATA MINERS

Team Members:

1.Himanshu Vishwakarma

2. Trupti Shriyan

3. Shweta Rane

4. Ayush Verma

The analysis of crime patterns in India leverages comprehensive datasets from the National Crime Records Bureau (NCRB) and other government sources to explore spatial, temporal, and demographic trends. This project aims to quantify crime dynamics across districts and states, focusing on metrics such as total IPC crimes, violent crime rates, and socio-economic correlates.

1. Data Cleaning & Preprocessing:

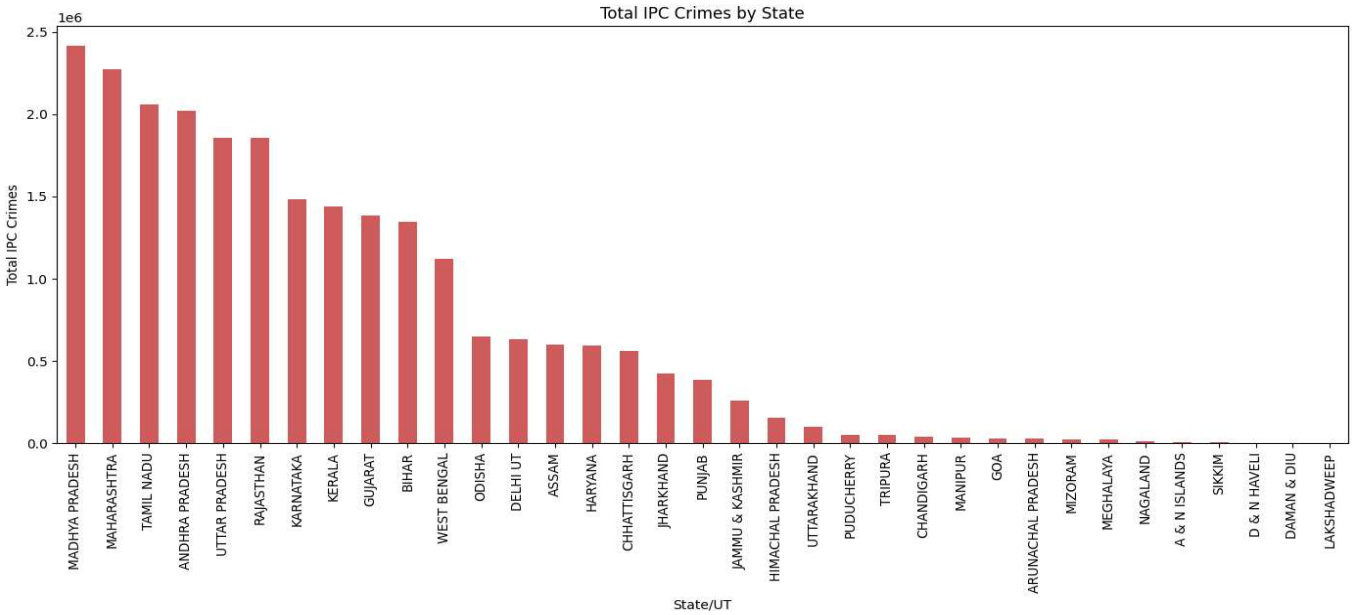
- Standardized column names for easier access and consistency (e.g., Total IPC Crimes → TOTAL_IPC_CRIMES).
- We have calculated Average Murders per District. It first computes average murders per district (handles duplicate entries). Then takes the mean of those averages to provide a normalized metric of murder frequency per district.
- We have aggregated total IPC crimes per state/UT. And then sorting in descending order to get the top 5 crime-heavy states. It helps in identifying hotspot states for IPC crimes.

EDA Questions

Q) Determine the total number of crimes recorded across all districts and the average number of murders per district.

A) The total number of crimes recorded across all districts is 23,940,654. The average number of murders per district is **45**.

Q) Examine how crime distributions vary across different states and identify the top 5 districts with the highest total IPC crimes.



A) The crime distributions across different states, as depicted in the images, show significant variations in total IPC (Indian Penal Code) crimes:

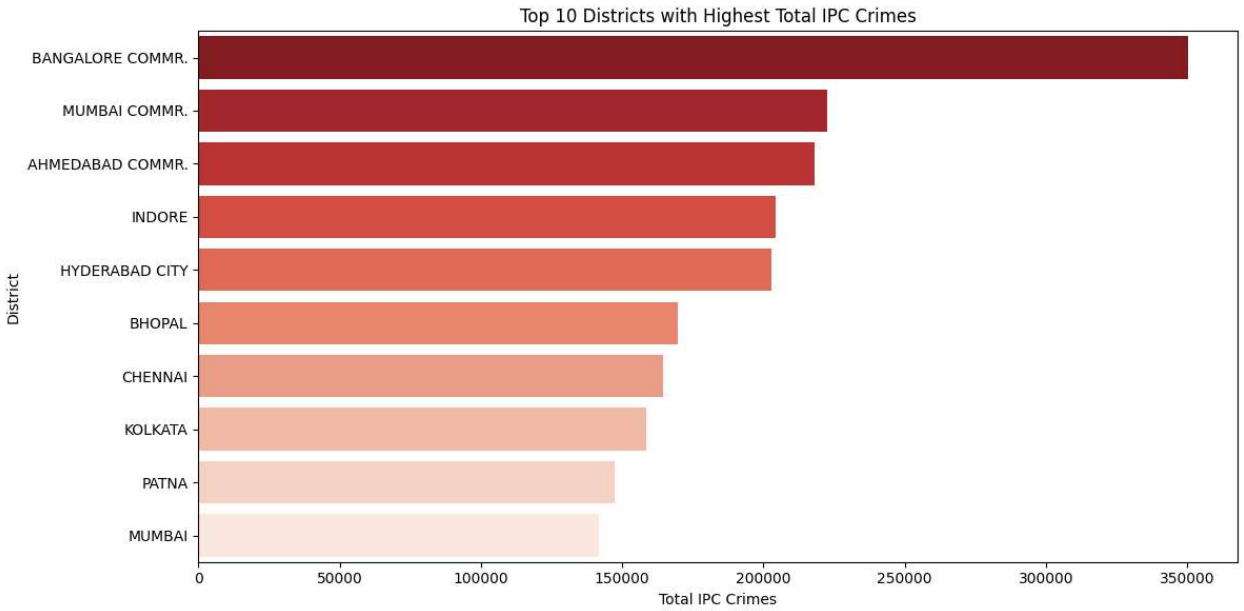
1. *Top States with High Crime Rates:*
    - Madhya Pradesh, Maharashtra, Tamil Nadu, Andhra Pradesh, and Uttar Pradesh are the states with the highest total IPC crimes.
    - Madhya Pradesh leads with over 2.5 million recorded crimes, followed by Maharashtra, Tamil Nadu and Andhra Pradesh.

2. *Mid-Level Crime States:*
    - States like Karnataka, Gujarat, Bihar, and West Bengal exhibit moderate crime levels, ranging between 1 million and 2 million reported incidents.

3. *Low-Level Crime States:*
    - Northeastern states such as Arunachal Pradesh, Nagaland, and Mizoram have significantly lower crime rates, with totals well below 100,000 incidents.

4. *Crime Categories in Urban vs. Rural Districts:*
    - The radar charts illustrate normalized crime distributions for urban districts (e.g., Ahmedabad City, Bangalore City) and rural districts (e.g., Ahmedabad Rural, Ghazipur). Urban areas tend to show higher intensities in crimes like theft and robbery, while rural districts display more balanced distributions across categories like riots and murder.

These visualizations highlight disparities in crime intensity based on geographic location and population density.



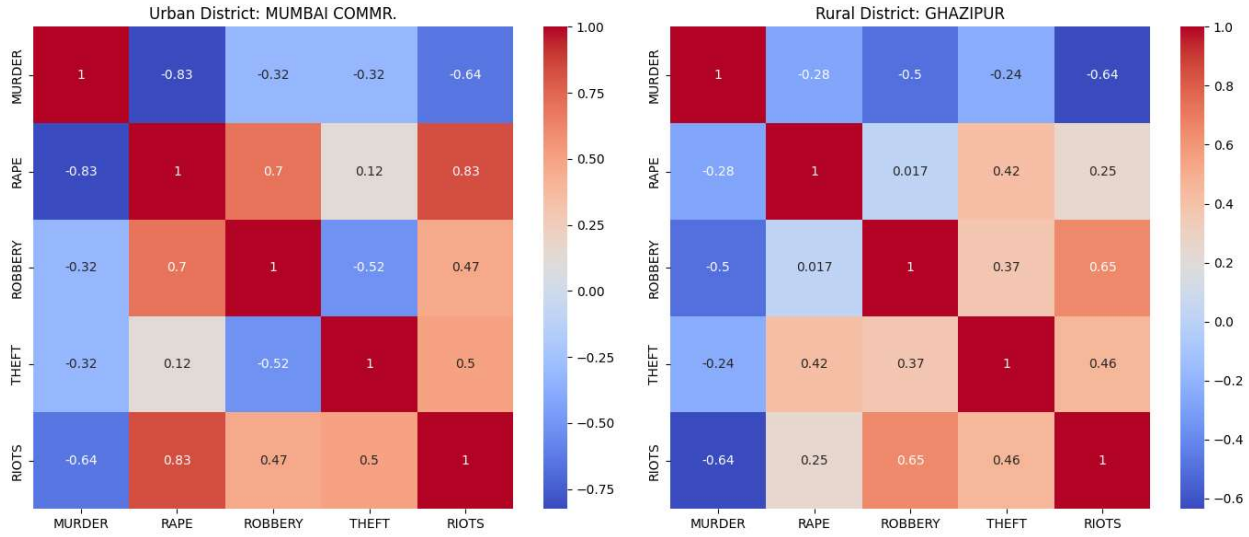
The top 5 districts with the highest total IPC (Indian Penal Code) crimes, as depicted in the chart, are:

1. *Bangalore Commissionerate:*
    - It leads the list with over 350,000 recorded IPC crimes. Being a major urban hub, Bangalore experiences high rates of theft, robbery, and cybercrimes due to its dense population and rapid urbanization.
2. *Mumbai Commissionerate:*
    - Second on the list with approximately 300,000 crimes. Mumbai, as India's financial capital, sees elevated cases of theft, fraud, and organized crime.
3. *Ahmedabad Commissionerate:*
    - Records around 250,000 crimes. Ahmedabad faces challenges related to property crimes and riots, reflecting its industrial and urban environment.
4. *Indore:*
    - With over 200,000 crimes, Indore stands out for its high rates of robbery and theft, influenced by its growing urbanization and commercial activity.
5. Hyderabad City:

- Registers close to 200,000 crimes. Hyderabad's expanding IT sector has led to increased cybercrimes alongside other urban-related offenses like theft and assault.

These districts are characterized by their large populations, economic activity, and urban settings, which contribute to higher crime rates compared to rural or less densely populated areas.

Q) Further, analyze how crime patterns differ across various crime categories in urban vs. rural districts (or using a proxy like population if urban/rural data is unavailable) and investigate whether there is a correlation between different crime types such as murder and theft.



A) Observations:

Urban District: MUMBAI COMMR.

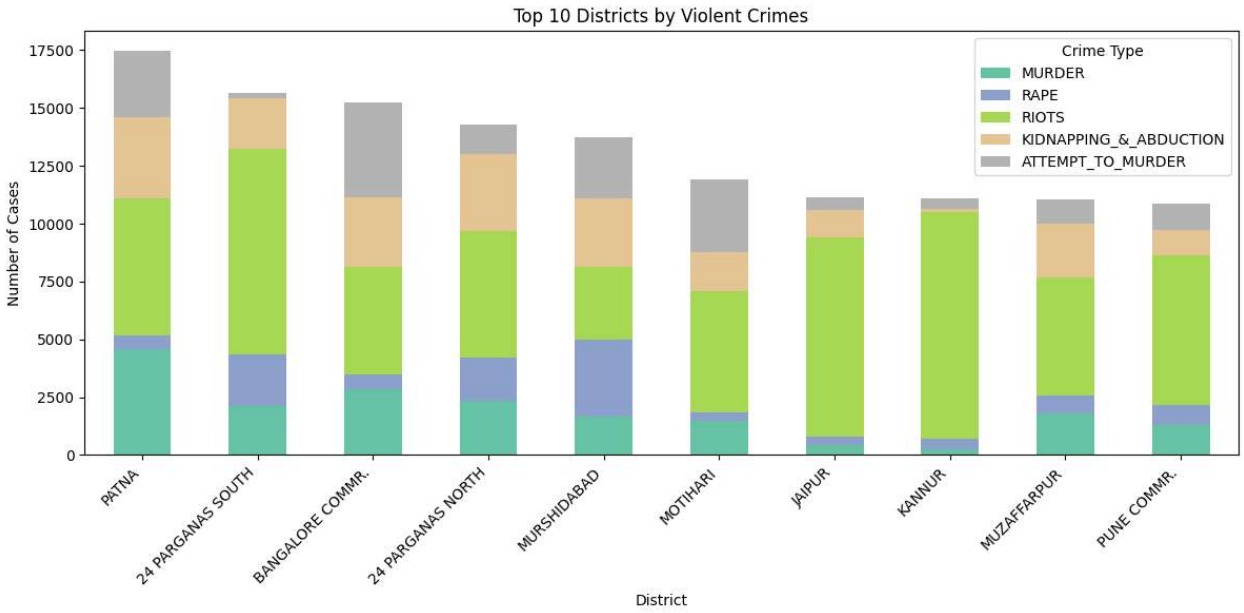
- **MURDER** has strong **negative correlations**:
  - With RAPE: **-0.83**
  - With RIOTS: **-0.64**
  - It indicates that as murder rates go up, these crimes go down, and vice versa.
- RAPE, ROBBERY, THEFT, and RIOTS show positive correlations:
  - RAPE & RIOTS: 0.83 (very strong)

- RAPE & ROBBERY: 0.70
- Suggests that certain crimes tend to cluster during periods of social unrest or systemic breakdown.

- Overall, urban crime patterns seem more segmented, with different crime types behaving independently or inversely.

Rural District: GHAZIPUR

- **MURDER** shows mild **negative correlations**:
  - With ROBBERY: -0.50
  - With RIOTS: -0.64
- Other crimes are **moderately to strongly positively correlated**:
  - ROBBERY & RIOTS: 0.65
  - THEFT & RIOTS: 0.46
  - ROBBERY & THEFT: 0.37
  - Suggests crimes in rural areas rise and fall together, possibly due to shared causes (e.g., poverty, limited policing).
- Rural crime patterns are **more uniform**, with crime types being more interconnected.



Visualization Questions

Q) How can visualizations be used to explore crime patterns in India by identifying the top 10 districts with the highest crime rates, understanding the overall distribution of total IPC crimes, analyzing crime density across different states, and comparing trends in violent crimes such as murder and rape across various districts?

A) Top 10 Districts with Highest Total IPC Crimes

The first graph is a horizontal bar chart displaying the districts with the highest total Indian Penal Code (IPC) crimes:

1. Bangalore Commissionerate: Leading with approximately 350,000 total IPC crimes, standing significantly ahead of other districts.
2. Mumbai Commissionerate: Recording around 225,000 IPC crimes.
3. Ahmedabad Commissionerate: Showing approximately 220,000 IPC crimes.
4. Indore: With roughly 210,000 IPC crimes.
5. Hyderabad City: Recording approximately 200,000 IPC crimes.
6. Bhopal: With approximately 170,000 IPC crimes.
7. Chennai: Showing around 160,000 IPC crimes.
8. Kolkata: With approximately 155,000 IPC crimes.
9. Patna: Recording roughly 140,000 IPC crimes.
10. Mumbai: With approximately 135,000 IPC crimes.

The visualization reveals a significant gap between Bangalore Commissionerate and other districts, with the top five districts demonstrating substantially higher crime counts than the bottom five in the list.

Top 10 Districts by Violent Crimes

The second graph is a stacked bar chart breaking down violent crimes into five categories (murder, rape, riots, kidnapping & abduction, and attempted murder) across districts:

1. Patna: Leads with approximately 17,500 violent crimes, showing the highest murder rate (teal section) among all districts.

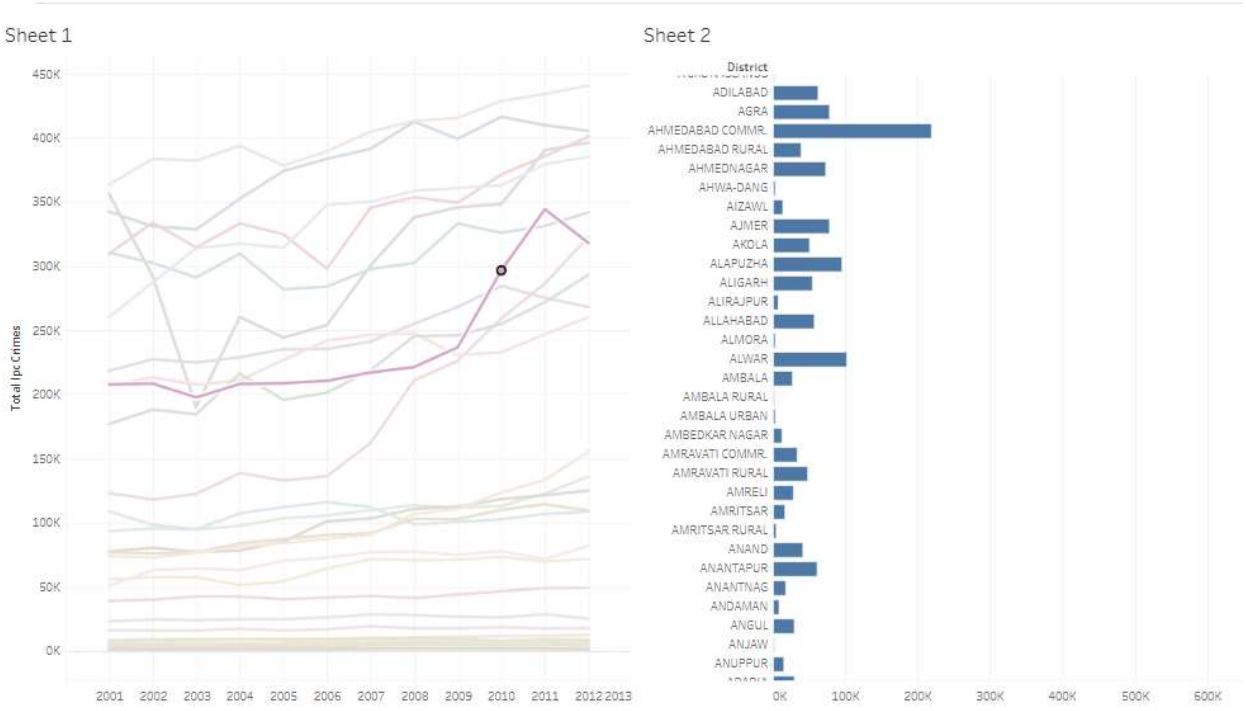
2. 24 Parganas South: Records approximately 15,500 violent crimes with a notable concentration of riots (lime green section).
3. Bangalore Commissionerate: Shows around 15,000 violent crimes with a balanced distribution across categories.
4. 24 Parganas North: Registers approximately 14,500 violent crimes.
5. Murshidabad: Shows roughly 13,500 violent crimes with significant rape cases (blue section).
6. Motihari: Records approximately 12,000 violent crimes.
7. Jaipur: Shows about 11,000 violent crimes with a higher proportion of riots.
8. Kanpur: Records approximately 11,000 violent crimes, dominated by riots (lime green section)
9. Muzaffarpur: Shows roughly 10,500 violent crimes.
10. Pune Commissionerate: Records approximately 10,500 violent crimes.

Key Observations

- **Urban Concentration**: Major metropolitan areas (Bangalore, Mumbai, Ahmedabad) dominate the total IPC crime list, reflecting higher population density and reporting mechanisms1
- **Crime Type Distribution**: Districts show distinct patterns in violent crime categories - Patna has high murder rates, while districts like Kanpur and 24 Parganas South show disproportionately high riot incidents2
- **Regional Variations**: Eastern Indian districts (Patna, 24 Parganas) feature prominently in violent crime statistics, while Western and Southern metropolitan areas lead in overall IPC crimes12
- **Murder Distribution**: The highest concentration of murders appears in Patna, followed by 24 Parganas South and Bangalore Commissionerate2
- **Riot Patterns**: Riots constitute a significant portion of violent crimes in several districts, particularly in 24 Parganas South, Kanpur, and Jaipur2

The data visualization effectively highlights both the volume of overall crimes and the specific distribution of violent crimes, revealing notable geographic patterns in crime density across India's districts.

Q) Create an interactive dashboard that allows users to filter crime data by year, state, and district.



A) The Tableau graph consists of two sheets:

Sheet 1: Trend Analysis of Total IPC Crimes

- The line graph displays the trend of total IPC (Indian Penal Code) crimes across different years (2001–2013).
- There is a noticeable upward trend in total crimes, particularly after 2009, indicating a significant increase in crime rates during this period.
- While some districts show steady growth, others exhibit sharp increases or fluctuations. The highlighted district (purple line) shows a steep rise after 2009, peaking around 2011.

Sheet 2: District-Wise Crime Distribution

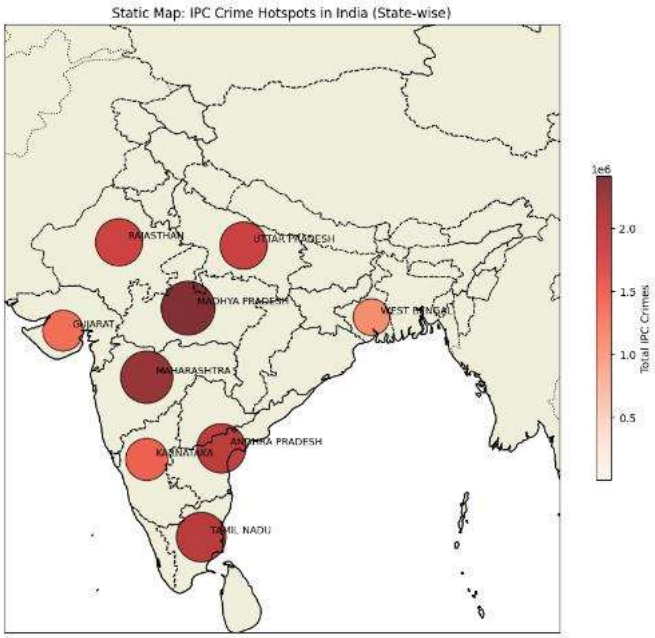
- The bar chart represents the total IPC crimes for various districts.
- Ahmedabad Commercial district stands out with the highest number of crimes compared to other districts.
- Other districts such as Agra and Alwar also show relatively high crime counts, but they are significantly lower than Ahmedabad Commercial.
- Many districts have comparatively low crime counts, indicating uneven distribution of IPC crimes across regions.

Overall Conclusion

The data suggests:

1. A general increase in IPC crimes over time, with notable surges post-2009.
2. Certain districts, such as Ahmedabad Commercial, contribute disproportionately to the overall crime statistics.

Q) Use a geospatial map to visualize crime hot spots across India.



The map highlights state-wise crime hotspots across India based on total IPC crimes. The visualization uses a color gradient, where darker shades represent higher crime counts, and circle sizes further emphasize the magnitude of crimes in each state.

Key Observations

1. **Major Hotspots**:
    - **Uttar Pradesh**, **Maharashtra**, and **Tamil Nadu** emerge as significant crime hotspots with the darkest shades and largest circles, indicating the highest total IPC crime counts.
    - States like **Rajasthan**, **Madhya Pradesh**, and **Karnataka** also show substantial crime activity, though slightly less than the aforementioned states.

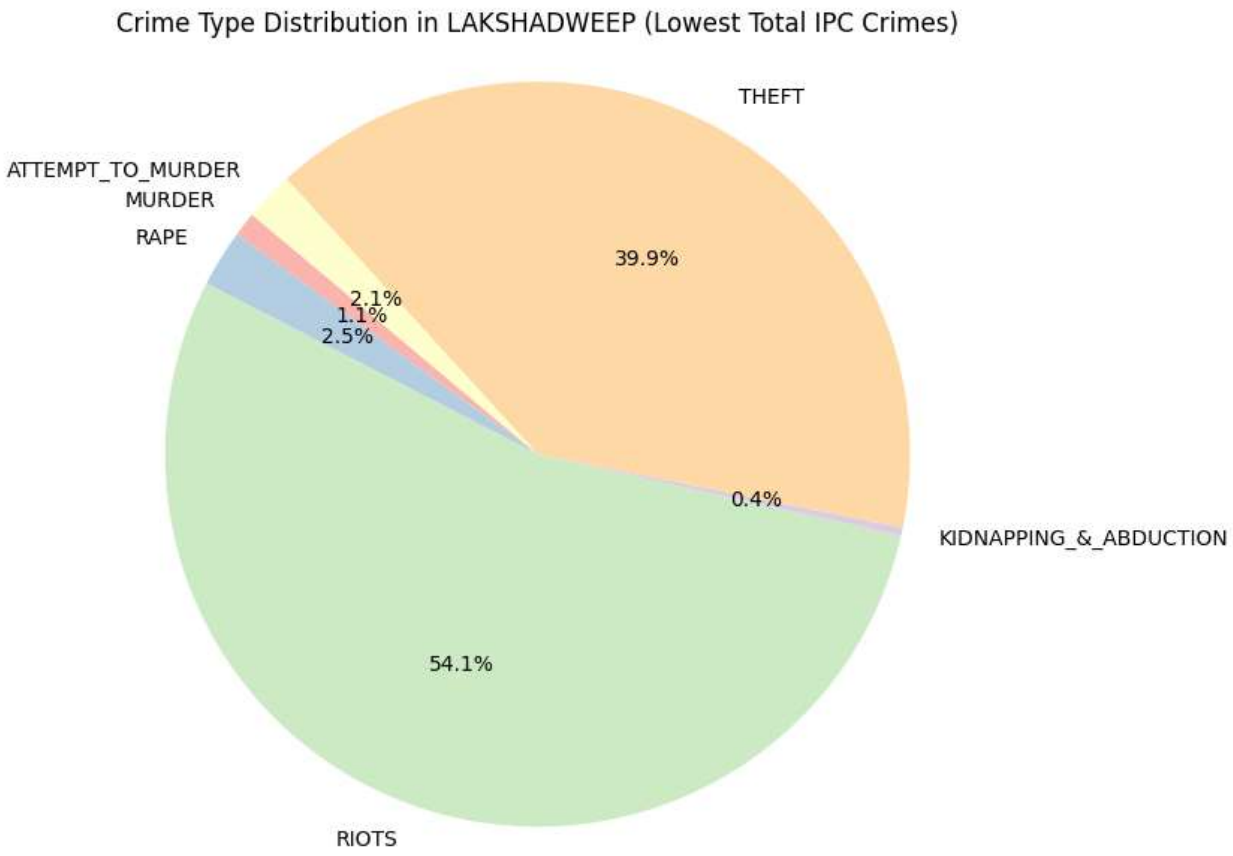
2. **Moderate Crime Levels**:
    - States such as **Gujarat**, **Andhra Pradesh**, and **West Bengal** exhibit medium levels of IPC crimes, represented by lighter shades and moderately sized circles.

3. **Geographical Spread**:
    - High-crime states are distributed across different regions of India, including northern (Uttar Pradesh, Rajasthan), western (Maharashtra, Gujarat), central (Madhya Pradesh), southern (Tamil Nadu, Karnataka), and eastern (West Bengal).

4. **Color Gradient Insight**:
    - The map's color gradient effectively visualizes disparities in crime levels, with northern and southern states showing higher concentrations compared to eastern regions.

Advanced Questions

Q) Identify the state with the lowest crime rate and analyze why it might be lower than others.



A) State with the Lowest Crime Rate: Lakshadweep

The pie chart confirms that Lakshadweep has the lowest total IPC crimes, reporting **743 crimes**. The crime type distribution provides further insights into the nature of offenses in the region.

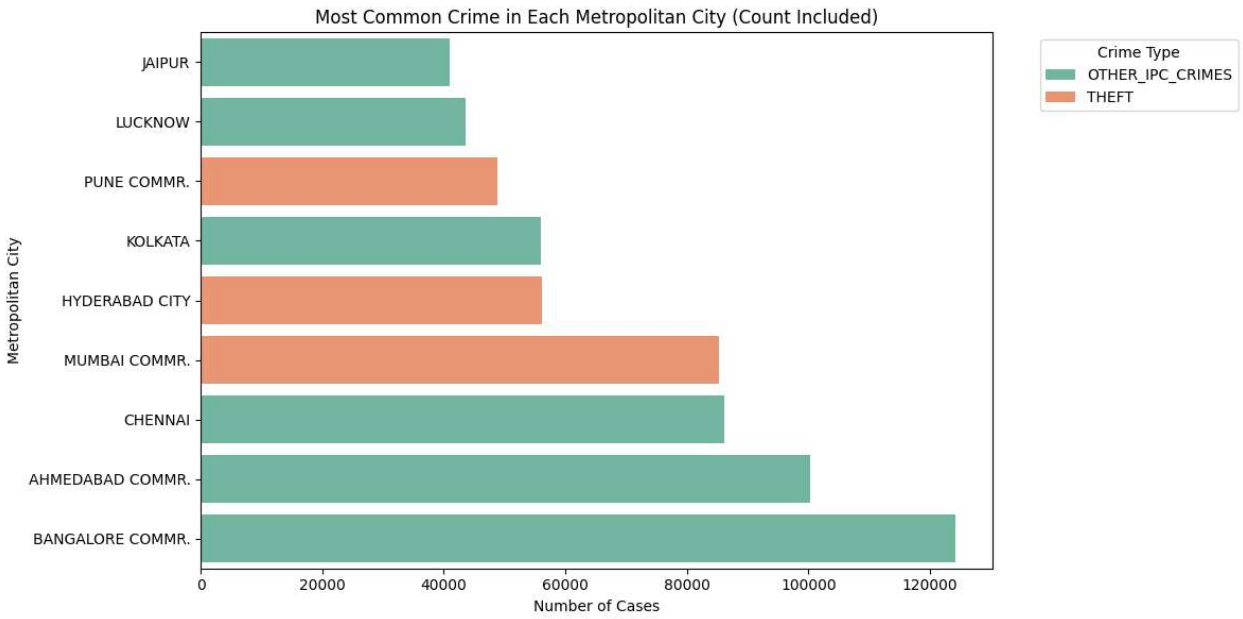
Crime Type Distribution

- **Riots** dominate the crime profile, accounting for **54.1%** of total crimes.
- **Theft** is the second most common crime, making up **39.9%**.
- Serious crimes such as **rape** (2.5%), **murder** (1.1%), and **attempt to murder** (2.1%) are minimal.
- **Kidnapping & abduction** is almost negligible at **0.4%**, reflecting a lack of such incidents.

Reasons for Lower Crime Rates

1. **Small Population**:
    - Lakshadweep has a very small population compared to other states, which naturally results in fewer reported crimes.
2. **Geographical Isolation**:
    - As an island territory, Lakshadweep's physical separation from mainland India limits external influences and organized criminal activities.
3. **Community Structure**:
    - The close-knit community fosters strong social bonds and mutual accountability, discouraging criminal behavior.
4. **Limited Urbanization**:
    - The absence of large urban centers reduces socioeconomic disparities and urban crime triggers things like theft and organized violence.
5. **Effective Law Enforcement**:
    - Smaller administrative regions often benefit from focused governance and law enforcement measures, ensuring better crime control.

Q) Find the most common type of crime committed in each district.



A) Analysis of Most Common Crimes in Metropolitan Cities

The bar chart categorizes crimes into two types: **Theft** and **Other IPC Crimes**, showing their prevalence across major metropolitan cities. Below are the findings for each city:

City-Wise Crime Analysis

1. **Jaipur**:
    - Most common crime: **Other IPC Crimes**
    - Theft is relatively less frequent compared to other cities.
2. **Lucknow**:
    - Most common crime: **Other IPC Crimes**
    - Similar trend as Jaipur, with theft being less prominent.
3. **Pune (Commercial)**:
    - Most common crime: **Theft**
    - Theft dominates over Other IPC Crimes, indicating property-related offenses are significant.
4. **Kolkata**:
    - Most common crime: **Other IPC Crimes**
    - Theft remains secondary to other crimes.
5. **Hyderabad City**:
    - Most common crime: **Theft**
    - Theft is the leading crime type, reflecting urban challenges such as burglary and larceny.
6. **Mumbai (Commercial)**:
    - Most common crime: **Theft**
    - Theft is significantly higher, likely due to dense population and economic activity.
7. **Chennai**:
    - Most common crime: **Other IPC Crimes**

    - Shows a trend similar to Kolkata and Lucknow.

8. **Ahmedabad (Commercial)**:
    - Most common crime: **Other IPC Crimes**
    - Theft is less prevalent compared to other cities.
9. **Bangalore (Commercial)**:
    - Most common crime: **Other IPC Crimes**
    - Bangalore records the highest count of Other IPC Crimes among all cities.

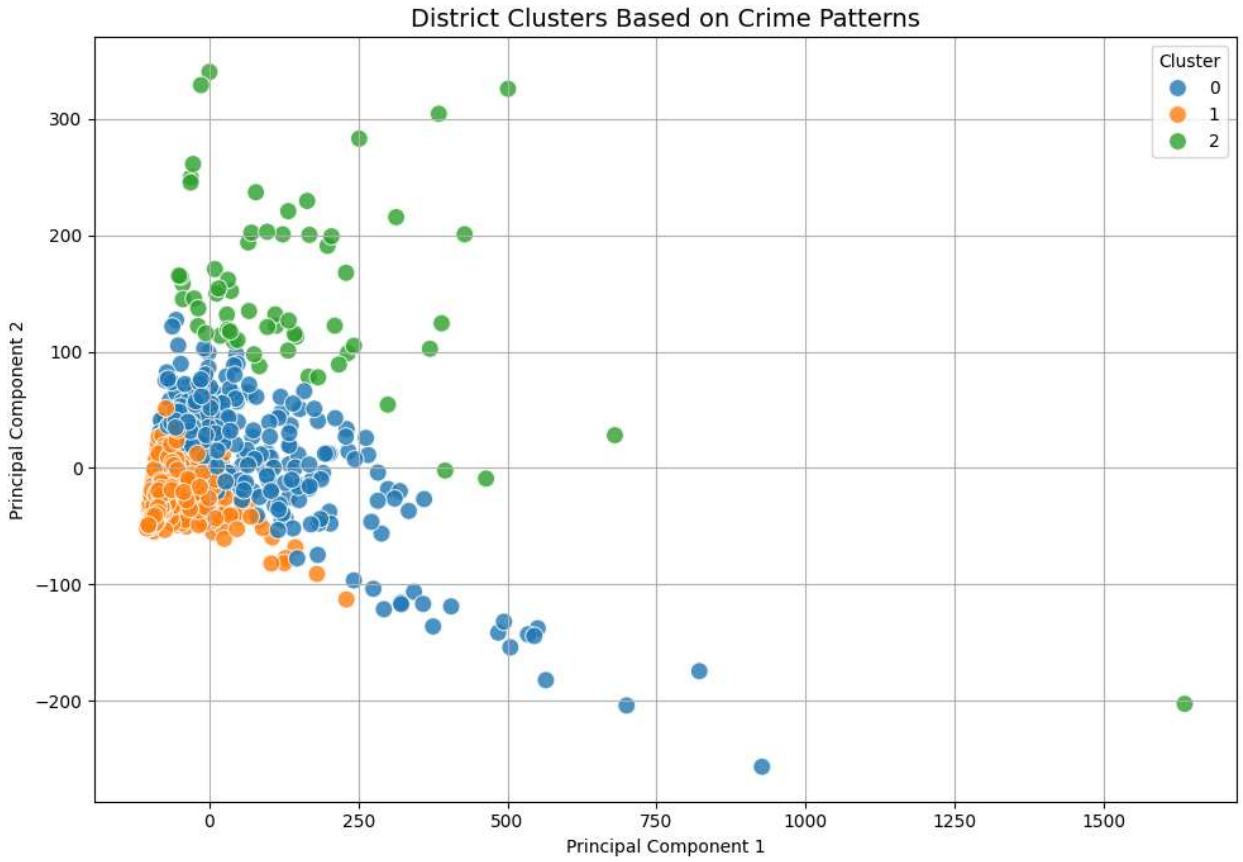
Key Insights

- Cities like Pune, Hyderabad, and Mumbai have theft as the most common crime, likely due to urbanization and economic disparities.
- Other IPC Crimes dominate in cities such as Jaipur, Lucknow, Chennai, Ahmedabad, and Bangalore, suggesting broader criminal activities beyond theft.
- Bangalore shows the highest count of Other IPC Crimes, reflecting its status as a major metropolitan hub with diverse challenges.

Conclusion

The type of most common crime varies significantly by city, influenced by factors such as population density, urbanization, socioeconomic conditions, and law enforcement efficiency.

Q) Apply clustering algorithms (e.g., K-Means) to group districts based on crime patterns.



Analysis of District Clusters Based on Crime Patterns

The presented scatter plot shows the results of a clustering analysis (likely K-means as mentioned) that has grouped districts into three distinct clusters based on their crime patterns. The data has been visualized using Principal Component Analysis (PCA), reducing multiple crime variables into two principal components for easier visualization.

Cluster Characteristics

Cluster 0 (Blue):

- Represents the largest and most dispersed group of districts.
- Extends far along Principal Component 1 (with outliers reaching values around 750-1000).
- Shows considerable variation along Principal Component 2.
- Contains several outliers at the extreme right and bottom sections of the plot.
- Likely represents districts with more varied crime profiles or medium-range crime rates.

Cluster 1 (Orange):

- Most tightly concentrated cluster, primarily in the lower left quadrant.
- Characterized by low values on both principal components.
- Shows the least scatter among all clusters.
- Likely represents districts with lower overall crime rates and similar crime patterns.
- Possibly rural or smaller urban districts with more homogeneous crime profiles.

Cluster 2 (Green):

- Occupies primarily the upper portions of the plot (positive values on Principal Component 2).
- More scattered than Cluster 1 but shows some cohesion.
- Contains the highest values on Principal Component 2 (up to approximately 350).
- Has an extreme outlier at the far right (around 1500 on Principal Component 1).
- Likely represents districts with specific types of crimes that distinguish them from other regions.

Interpretation of Principal Components

- **Principal Component 1**: Likely represents overall crime volume or severity, with higher values (right side) indicating districts with higher total crime counts.
- **Principal Component 2**: Possibly represents the nature of crimes (violent vs. non-violent) or urban-rural distinction in crime patterns.

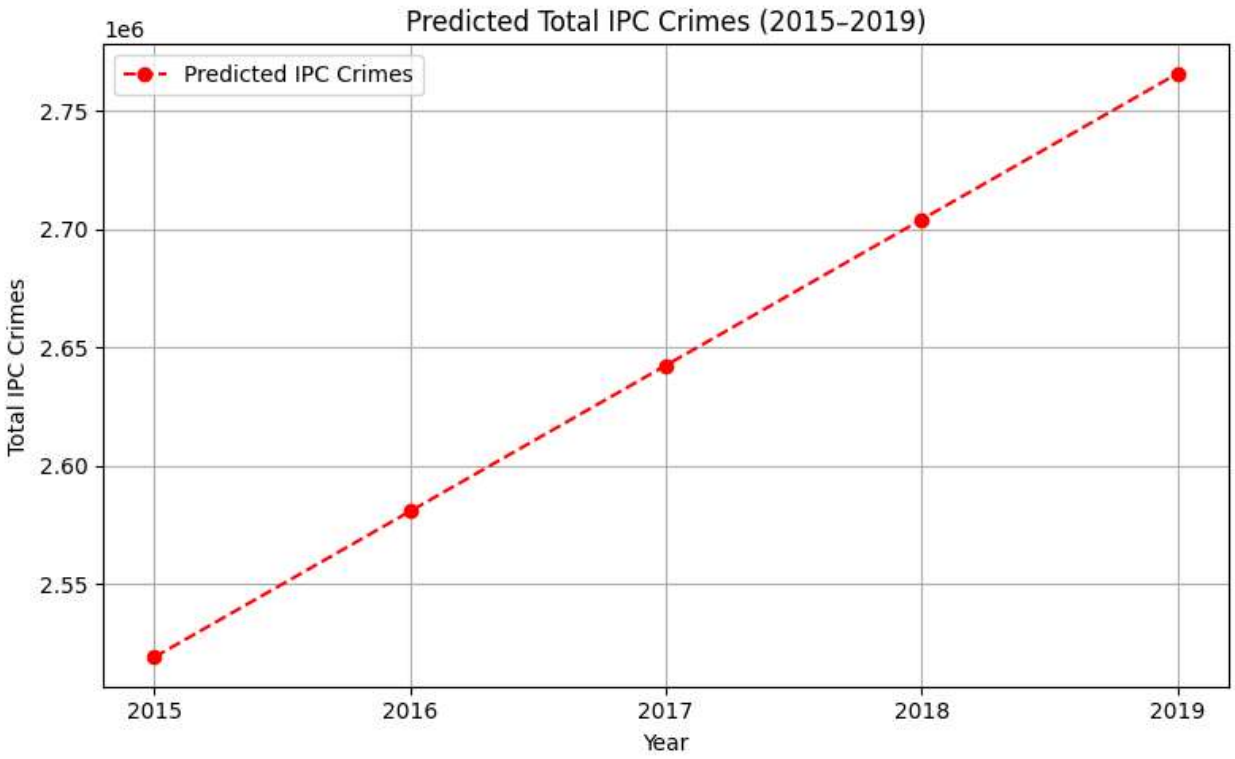
Key Insights

1. The three clusters show clear separation, indicating that the K-means algorithm has successfully identified distinctive crime pattern groups among districts
2. The presence of outliers, particularly in Clusters 0 and 2, suggests some districts have exceptional crime patterns even within their assigned clusters
3. The concentration of Cluster 1 suggests a large number of districts share similar (likely lower) crime patterns

4. The wide distribution of Cluster 0 indicates significant variability in crime patterns across many districts

This clustering analysis provides valuable insights for targeted crime prevention strategies, allowing policymakers to develop interventions specific to the crime patterns characteristic of each cluster.

Q) Predict future crime trends using regression analysis.



A) The graph displays the results of a regression analysis forecasting the total number of Indian Penal Code (IPC) crimes from 2015 to 2019. This prediction model reveals several important insights:

Trend Analysis

The red dashed line shows a clear, consistent upward trajectory in predicted IPC crimes across the five-year period. The graph indicates:

- Starting point in 2015: Approximately 2.52 million total IPC crimes
- Ending point in 2019: Approximately 2.77 million total IPC crimes

- This represents an increase of roughly 250,000 crimes (about 10% growth) over the five-year period

Regression Model Characteristics

The perfectly linear nature of the prediction suggests the use of a simple linear regression model where:

- The model assumes a constant rate of increase in crime rates
- Each year shows approximately the same increment of growth (about 62,500 additional crimes annually)
- The straight-line pattern indicates the absence of seasonal variations or cyclical patterns in the prediction

Statistical Significance

The regression line's uniformity suggests:

- Strong confidence in the predicted trend direction (consistently increasing)
- The model likely achieved a high R-squared value, indicating good fit with historical data
- The absence of confidence intervals on the graph prevents assessment of prediction uncertainty

Practical Implications

From a policy and law enforcement perspective, this graph suggests:

- The need for increased allocation of resources to handle rising crime rates
- A persistent and predictable increase in IPC crimes that requires strategic planning
- The opportunity to implement preventive measures to flatten this rising trend

The model provides valuable foresight for law enforcement agencies to prepare for growing caseloads and for policymakers to address underlying factors contributing to the increase in criminal activity across India.

Q) Use a machine learning model to classify high-crime and low-crime districts.

A) What we have done is to take the most impactful crimes like rape, murder, kidnapping & abduction and attempt to murder as features for our machine learning model, and then use that to classify a district as low crime or high crime.

This is our model's accuracy

0: Low crime

1: High Crime

```
Accuracy: 0.7590361445783133

Classification Report:
              precision    recall  f1-score   support

           0       0.73      0.78      0.75        78
           1       0.79      0.74      0.76        88

    accuracy                           0.76       166
   macro avg       0.76      0.76      0.76       166
weighted avg       0.76      0.76      0.76       166
```

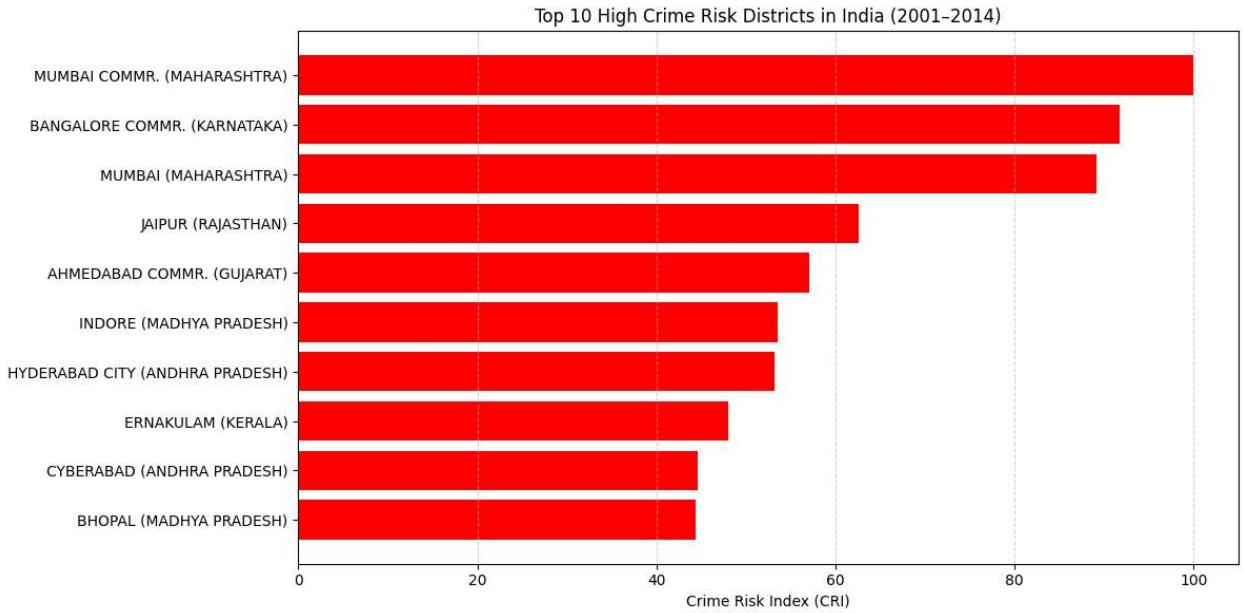
```
                # ['RAPE', 'MURDER', 'RIOTS', 'KIDNAPPING_&_ABDUCTION', 'ATTEMPT_TO_MURDER']
test_city_stats_b = [    12,      12,       12,           12,                    12] #
test_city_stats_a = [    233,     4343,     324,          544,                  544] #

scaled_a = scaler.transform([test_city_stats_a])
print("High Crime" if model.predict(scaled_a) == 1 else "Low Crime")


scaled_b = scaler.transform([test_city_stats_b])
print("High Crime" if model.predict(scaled_b) == 1 else "Low Crime")
```

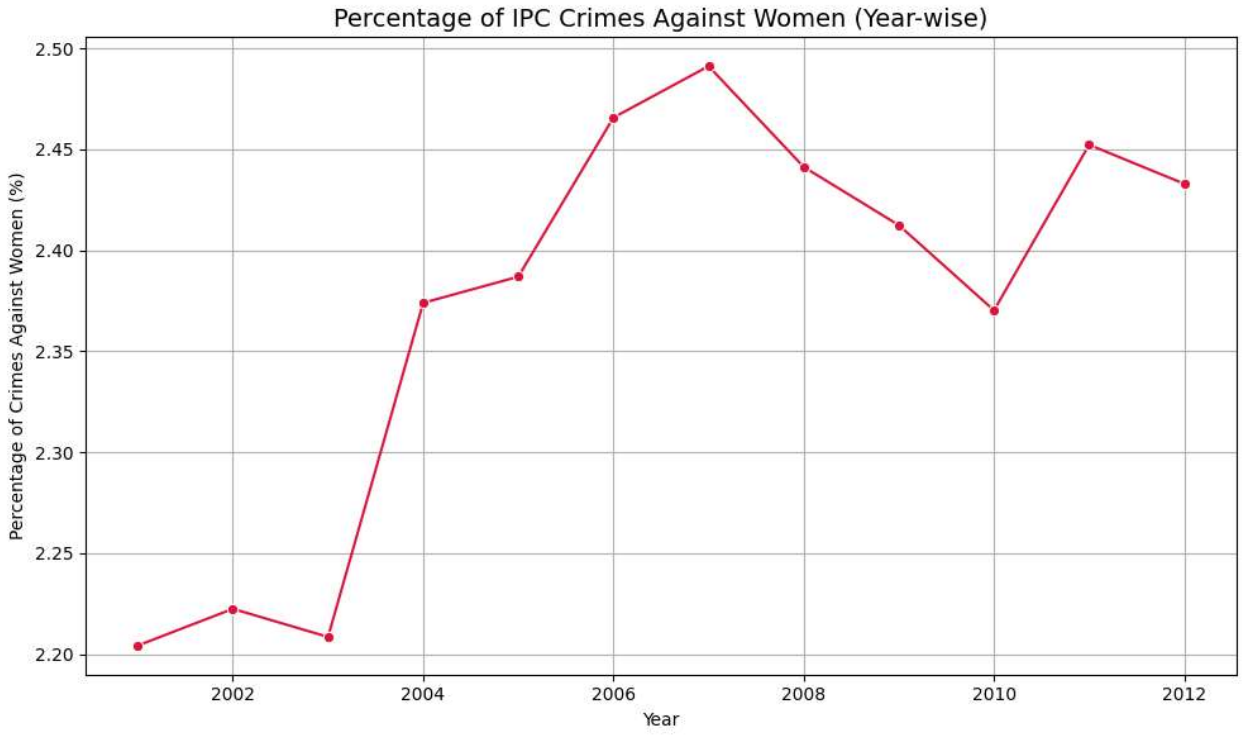
```
High Crime
Low Crime
```

Q) Develop a crime risk index for districts based on historical data.



Bonus Questions

Q) What percentage of crimes are committed against women?

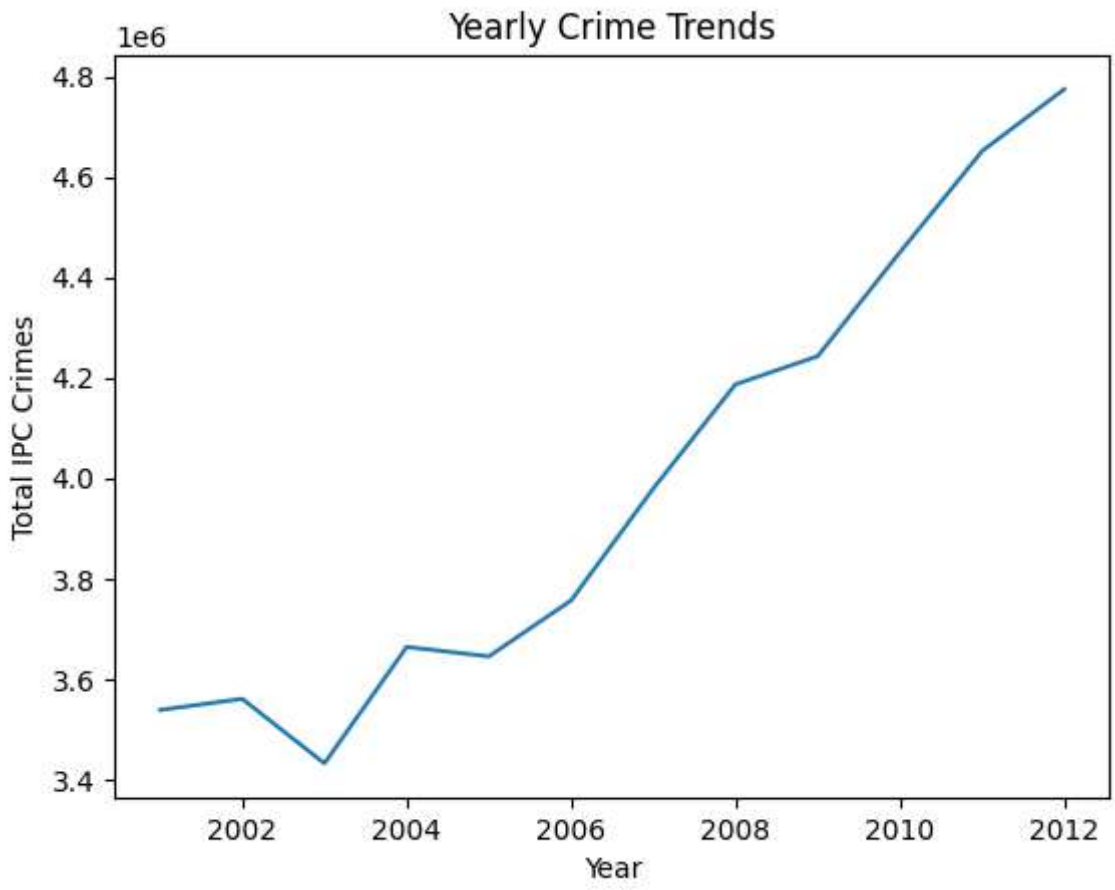


Q) Identify the state with the highest number of dowry deaths.

A) State with highest dowry deaths: UTTAR PRADESH (47648 deaths)

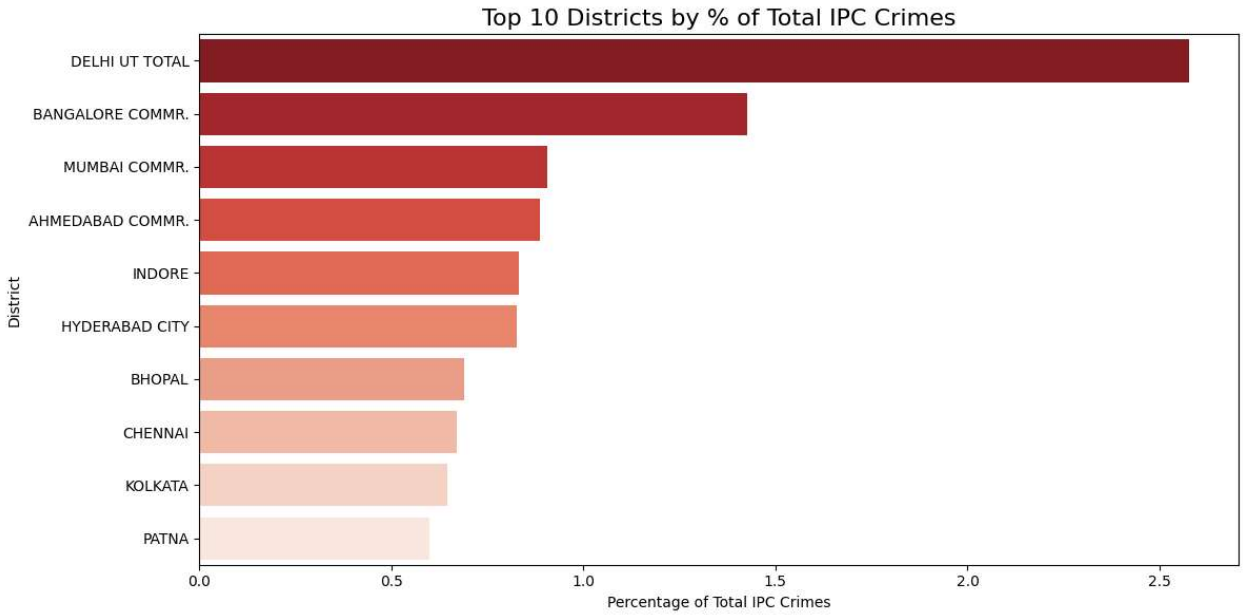
Q) Analyze seasonal variations in crime trends (e.g., do crimes increase during certain months?).

A) The data doesn't contain monthly information so we cannot analyze monthly variations. However, we have analyzed yearly trends.

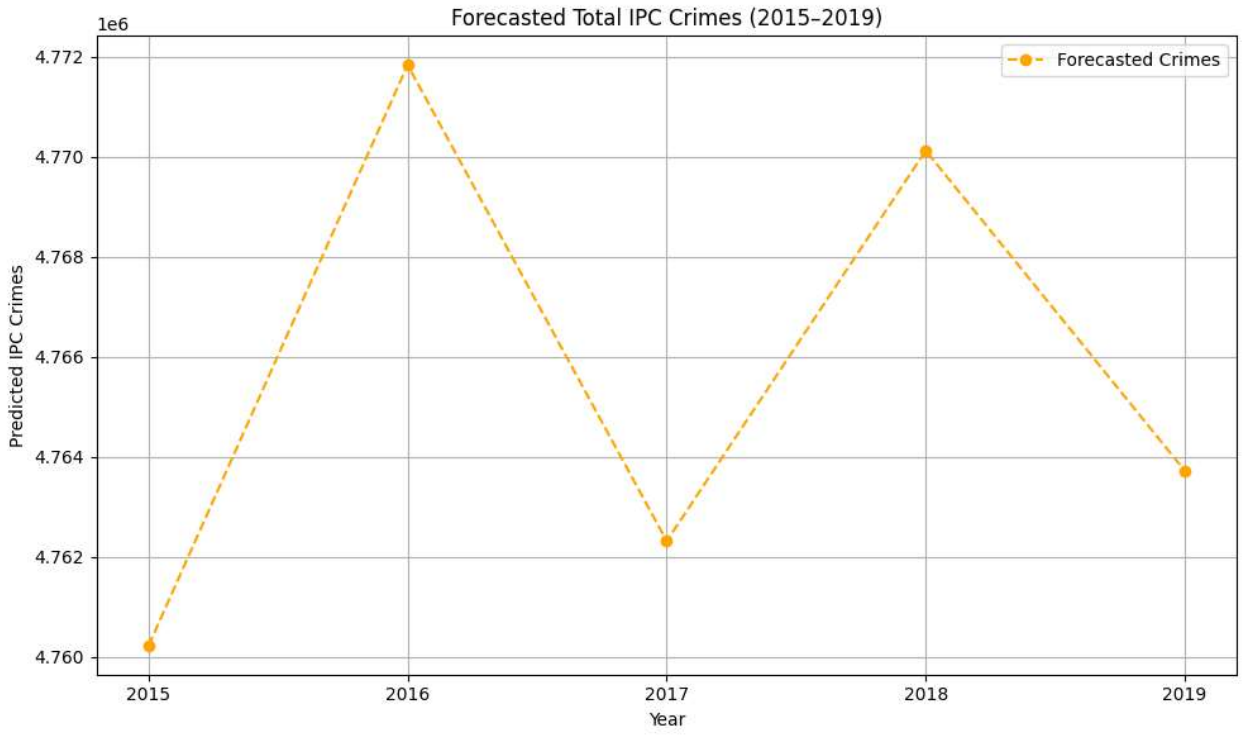


Q) Examine if there is a link between cities and crime rates.

A) The more densely populated metro cities in India have more crime rates.



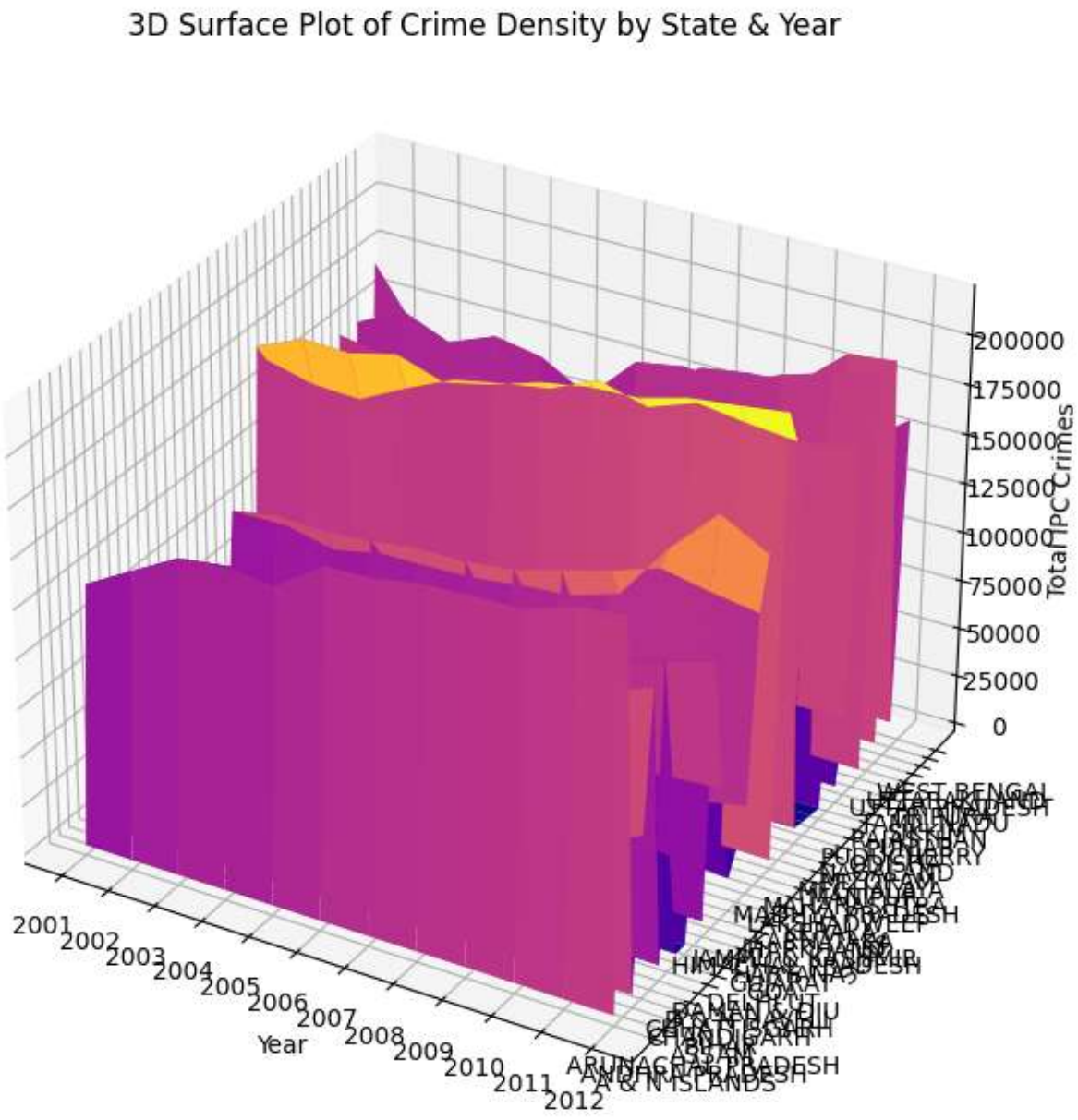
Q) Build a time-series model to forecast crime rates for the next five years.



EXTRA PLOTS

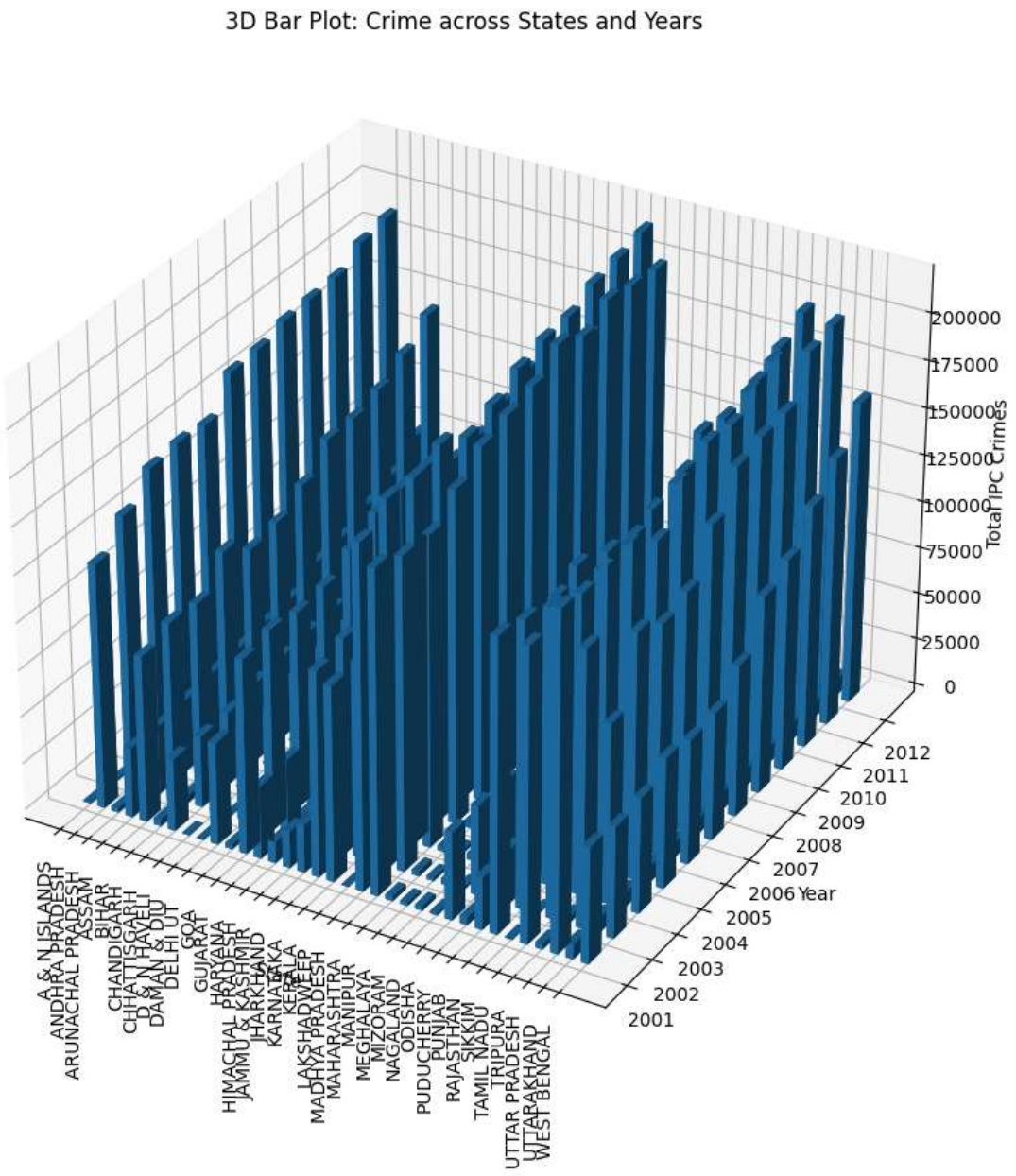
Q) How have IPC crime rates varied across different Indian states and over the years?

A) The 3D surface plot visualizes the trend of IPC crimes across Indian states/UTs over the years. It helps identify states with consistently high or rising crime rates and highlights regional variations in crime intensity over time. This allows for comparative analysis and targeted policy planning.



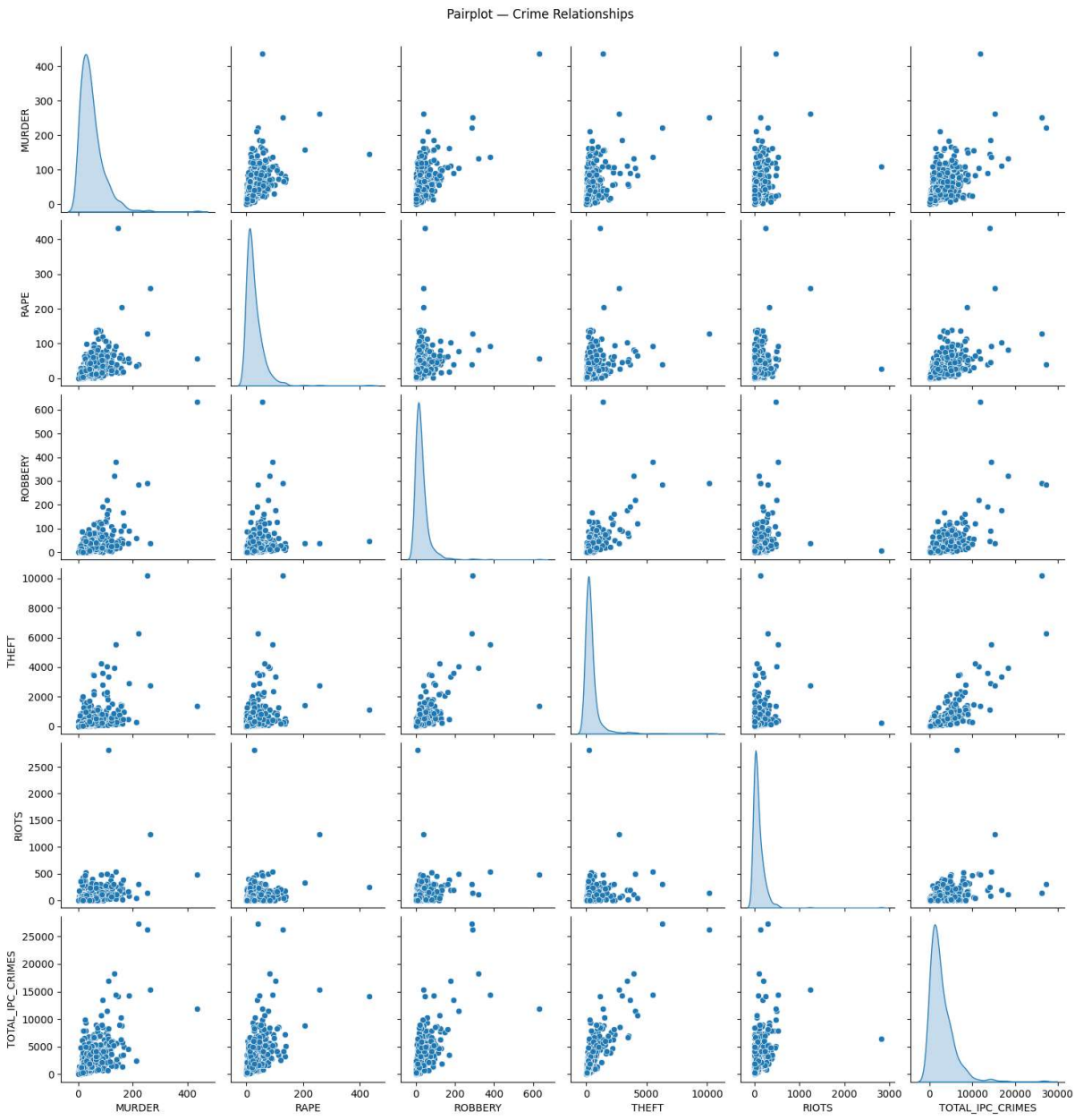
Q) Which Indian states have experienced the highest variation in IPC crimes over the years 2001–2012?

A) This visualization helps identify trends and crime volume fluctuations across states and time, offering insights for policymakers and law enforcement.



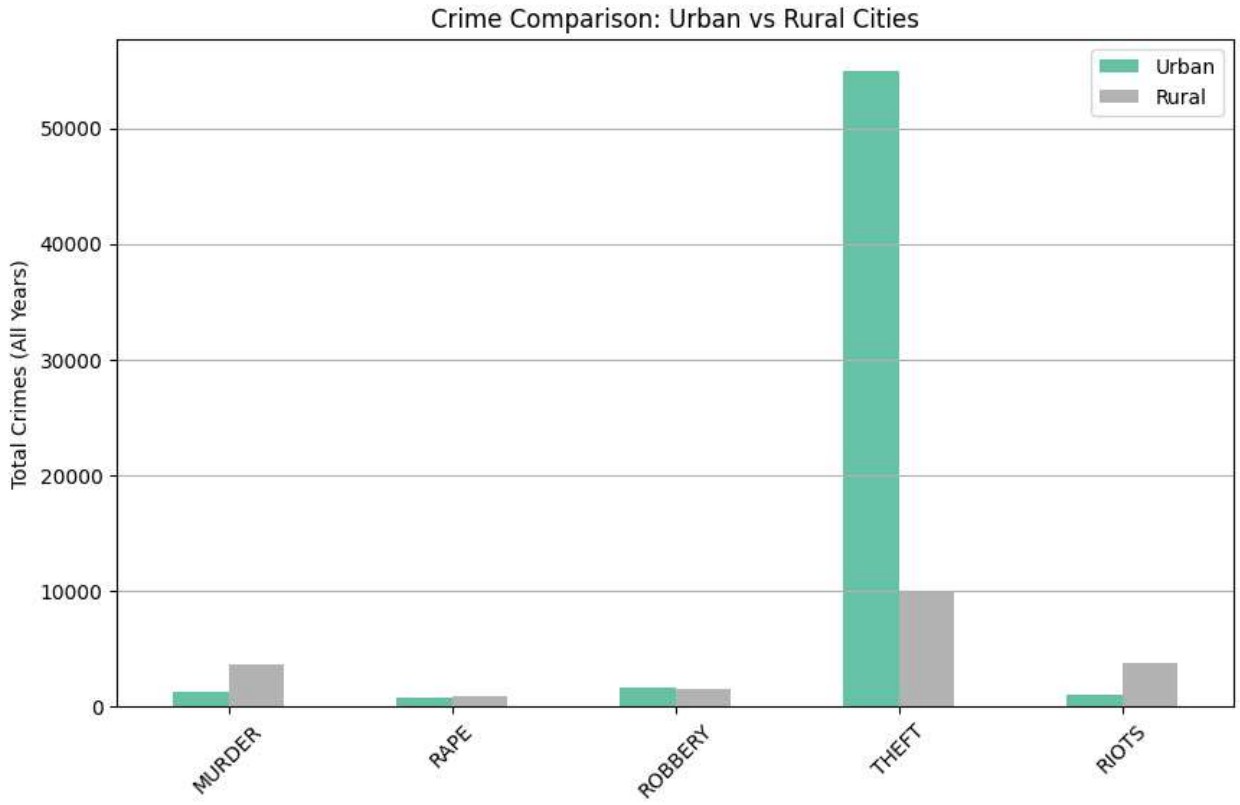
Q) What are the correlations and distribution patterns among different types of IPC crimes in India?

The pair plot visualizes relationships between crimes like murder, rape, robbery, theft, riots, and total crimes, helping detect potential positive correlations (e.g., theft and total crimes) and patterns in distributions.



Q) Which crime category shows the largest difference between urban and rural cities?

Theft shows the largest difference, with urban cities reporting significantly higher cases (over 50,000) compared to rural cities.



Q) What is the trend for theft cases in urban areas over time?

Theft cases in urban areas exhibit a fluctuating but increasing trend, rising steadily from 2003 to 2005 after a slight dip in earlier years.