

Crimes Against Women

Exploratory Data Analysis

Members:

Rakshita

Suraj Kumar

Aakash Kumar Kundan

Mehak

Khushi

Manish

Vinay Gautam

Content

1. Vision & Purpose
2. Python Libraries
3. Work flow
4. Visualization
- 5.

Vision and Purpose of the Project: Crimes Against Women in India

Crimes against women represent a pervasive challenge, impacting not only individuals but also families, communities, and societies. To understand this issue better, data science offers an effective approach for analyzing crime patterns and drawing actionable insights. This report presents an exploratory analysis of a dataset documenting crimes against women across different states and union territories of India. The focus of the analysis is to uncover patterns, trends, and relationships between various types of crimes, and propose preventive and mitigation strategies.

Objectives:

- 1. Identify patterns and trends in crimes against women over time and across different regions.**
- 2. Examine correlations between different crime rates.**
- 3. Highlight geographical hotspots where crimes against women are most prevalent.**
- 4. Identify anomalies and outliers, such as sudden spikes in crime rates.**
- 5. Provide data-driven insights to inform policies aimed at reducing crimes against women.**

Key Python Libraries for Exploratory Data Analysis:

- 1. pandas:** For data manipulation and analysis.
- 2. numpy:** For numerical computations.
- 3. matplotlib:** For basic visualizations.

4. **seaborn**: For advanced visualizations.
5. **plotly**: For interactive visualizations.
6. **scikit-learn**: For data preprocessing and machine learning tasks.
7. **statsmodels**: For statistical modeling.
8. **geopandas**: For handling geographical data.

Work Flow

1. Data Understanding and Loading

Goal: To gain a clear understanding of the data structure, content, and relevance to the problem at hand.

- **Load the Dataset:**
 - Load the dataset using Python (Pandas, NumPy, etc.).
 - Check for correct column names and ensure that data types (categorical, numerical) align with expectations.
- **Feature Familiarization:**
 - Review each feature and understand its role (already done).
 - Check the dataset's size, shape, and granularity (e.g., state-level, year-level).
- **Initial Check for Missing Data and Inconsistencies:**
 - Identify missing values using `isnull()`.
 - Check for inconsistencies like outliers, negative values where not appropriate, or unusual patterns.
- **Key Questions:**
 - What does each feature represent?
 - Are there missing or inconsistent values that need handling?
 - Are the data types correct?

2. Data Cleaning

Goal: To prepare the dataset for analysis by fixing errors, handling missing values, and ensuring uniformity.

- **Handling missing values:**

- Fill, interpolate, or drop missing data, depending on the importance of the feature.
 - Consider imputation techniques like filling missing values with the median, mean, or mode.
 - **Correcting Data Inconsistencies:**
 - Ensure all categorized data is properly labelled.
 - Handle anomalies such as incorrect or unreasonable entries (e.g., negative crime counts).
 - **Removing duplicates and irrelevant Data:**
 - Drop duplicate rows and columns that are irrelevant to the analysis (e.g., identifiers).
 - Standardize formats (e.g., consistent state names).
 - **Standardize state names:** Inconsistencies in state names due to capitalization and formatting were corrected to ensure uniformity (e.g., 'andhra pradesh' -> 'ANDHRA PRADESH').
 - **Key Questions:**
 - How are missing values handled?
 - Are there duplicates or irrelevant data points to remove?
 - Are the data entries standardized and consistent?
-

3. Descriptive Statistics

Goal: To summarize the central tendencies, dispersion, and shape of the dataset's distribution.

- **Summary Statistics:**
 - Use `describe()` to calculate key statistics (mean, median, standard deviation).
 - Explore the distribution of numerical columns to identify data spread and variability.
 - **Group-wise Descriptive Statistics:**
 - Group data by key variables such as state, year, and crime type to get aggregate insights.
 - Use pivot tables or group-by functions to compute grouped statistics.
 - **Key Questions:**
 - What are the basic statistics of the data?
 - How do crime rates differ across states and years?
 - What is the overall distribution of various crime types?
-

4. Exploratory Visualizations

Goal: To identify trends, outliers, and relationships visually.

- **Basic Visualizations:**
 - **Histograms:** For understanding the distribution of crime categories (e.g., how frequent is each crime type?).
 - **Bar Charts:** For comparing crime occurrences by state or year.
 - **Temporal Analysis:**
 - **Line Charts:** To visualize crime trends over time for different categories and states.
 - **Stacked Area Plots:** To show total crimes over time, broken down by type.
 - **Geospatial Visualizations:**
 - **Heatmaps/Choropleths:** For visualizing regional distributions of crime (e.g., mapping crime intensity by state).
 - **Key Questions:**
 - What are the visual patterns in crime distribution across time and space?
 - Are there visible outliers or unexpected trends?
 - What states show the highest or lowest crime rates?
-

5. Trend and Correlation Analysis

Goal: To uncover relationships and potential causal patterns between different features.

- **Trend Analysis:**
 - Analyze how crime rates evolve over time for specific categories.
 - Identify years or periods of significant increase/decrease in crimes.
 - **Correlation Matrix:**
 - Use a correlation matrix to assess relationships between different crime categories (e.g., is domestic violence strongly correlated with dowry deaths?).
 - Generate a **heatmap** to visualize correlations.
 - **Advanced Correlation Analysis:**
 - **Pairwise Scatter Plots:** For deeper visual understanding of relationships.
 - **Key Questions:**
 - What are the strongest correlations between crime types?
 - Are there any patterns suggesting co-occurrence of crimes?
 - Are there year-over-year trends for specific states?
-

6. Regional Analysis

Goal: To analyze crime patterns across different geographical areas, understanding regional differences.

- **Geographical Distribution:**
 - Create choropleths or regional heatmaps to visualize crime intensity by state or region.
 - Divide the dataset by regions (e.g., North, South, East, West) to explore regional crime trends.
 - **Key Questions:**
 - What regions or states have consistently high or low crime rates?
 - How do rural and urban areas compare in terms of crime?
-

7. Key Insights and Problem Areas

Goal: To summarize the most important findings from the analysis.

- **Identifying High Crime States and Categories:**
 - Identify states or regions with persistently high crime rates.
 - Flag crime types that show concerning trends (e.g., dowry deaths increasing in certain states).
 - **Key Questions:**
 - Which areas require immediate intervention?
 - What types of crimes have shown consistent increases?
 - Are there patterns that explain why certain crimes are more prevalent in specific regions?
-

8. Proposing Solutions and Mitigation Strategies

Goal: To translate the findings into actionable recommendations for policy intervention and societal change.

- **Strengthening Legal Frameworks:**
 - Based on regions with high crime rates, propose strengthening of laws, enforcement, or judicial oversight.
- **Targeted awareness campaigns:**
 - Identify areas or crime types where public awareness or intervention programs are needed (e.g., domestic violence hotlines in regions with high DV rates).
- **Resource Allocation:**

- Suggest the allocation of resources (e.g., policing, social support services) to regions or states with the most severe crime rates.
 - **Collaboration with Local Stakeholders:**
 - Recommend collaboration with local governments, law enforcement agencies, and NGOs to implement targeted interventions.
-

9. Reporting and Visualization

Goal: To present the analysis in a clear, impactful way that drives decision-making.

- **Create Dashboards:**
 - Develop interactive dashboards using tools like Power BI or Tableau to present dynamic views of the data.
 - Include filterable options for crime type, state, and year.
 - **Data Storytelling:**
 - Use storytelling techniques to highlight key insights, trends, and calls to action.
 - Make the reports accessible to non-technical stakeholders (e.g., policymakers, NGOs).
 - **Key Questions:**
 - How do you present your findings to effectively communicate the urgency of addressing specific crime types?
 - What visualization formats are most effective for your audience?
-

1.CrimesOnWomen1.csv (Data):

- **State:** Name of the state.
- **Year:** The year the data was recorded.
- **Rape:** Number of rape cases.
- **K&A:** Kidnap and Assault cases.
- **DD:** Dowry deaths.
- **AoW:** Assault on Women.
- **AoM:** Assault on Minors.

- **DV:** Domestic violence.
- **WT:** Witchcraft-related crimes.

2. Understanding the Dataset Structure

The first step in exploratory analysis was understanding the dataset's structure and gaining insights into the distribution of variables. Summary statistics such as the mean, median, and standard deviation provided a quantitative sense of the dataset's central tendencies and variability.

3. Trend Analysis

Crime trends over time were analyzed to identify patterns across the years for different states. Line plots for various crime categories (e.g., Rape, Dowry Deaths, Domestic Violence) were created to visualize how crime rates have evolved.

- **Rape Cases:** States like Uttar Pradesh, Rajasthan, and Madhya Pradesh show consistently rising trends in rape cases. This suggests either an increase in reported crimes or enhanced law enforcement and reporting mechanisms.
- **Domestic Violence:** The number of domestic violence cases has generally increased across all states, pointing to either improved reporting systems or a genuine rise in incidents.

Prevention Insight: States with an increasing trend should implement targeted programs to reduce these crimes, such as public awareness campaigns, educational programs, and strengthening of reporting mechanisms.

4. Crime Distribution

We utilized **heatmaps** to explore crime distribution across states and years. This helped in identifying high-crime regions and the intensity of different crime types.

- **High-Crime States:** Uttar Pradesh, Maharashtra, and Madhya Pradesh consistently exhibit the highest number of crimes across multiple categories. Smaller states like Goa and Sikkim have significantly fewer reported cases.

5. Correlation Analysis

A **correlation matrix** was created to check the relationships between different crime types. For instance, dowry deaths are strongly correlated with domestic violence, suggesting that these crimes often occur together in the same regions.

Key Findings:

- **Strong Correlation:** Dowry Deaths and Domestic Violence show a strong positive correlation, indicating a possible cultural and systemic issue in certain regions where women face multiple forms of violence.

Prevention Insight: Regions showing high correlations between crimes like dowry deaths and domestic violence require comprehensive interventions, including stricter law enforcement, community education, and rehabilitation programs for victims.

6. State-wise Crime Rankings

Bar charts were used to **rank states** based on total crimes. This visual representation helps in identifying which states are the most crime-prone.

- **Top Crime States:** Uttar Pradesh, Maharashtra, and Madhya Pradesh rank the highest in terms of total crime rates. Conversely, states like Goa and Sikkim have lower crime burdens.

7. Prevention and Mitigation Strategies

- **Public Awareness Campaigns**

Public campaigns to educate citizens about women's rights, legal protections, and reporting mechanisms are crucial. These campaigns should focus on high-crime regions and aim to reduce social stigma around reporting crimes.

- **Legal Reforms and Enforcement**

Stronger legal frameworks and stricter penalties should be enforced, especially for crimes like domestic violence and dowry deaths. Fast-tracking court cases related to women's safety will ensure timely justice and act as a deterrent to future crimes.

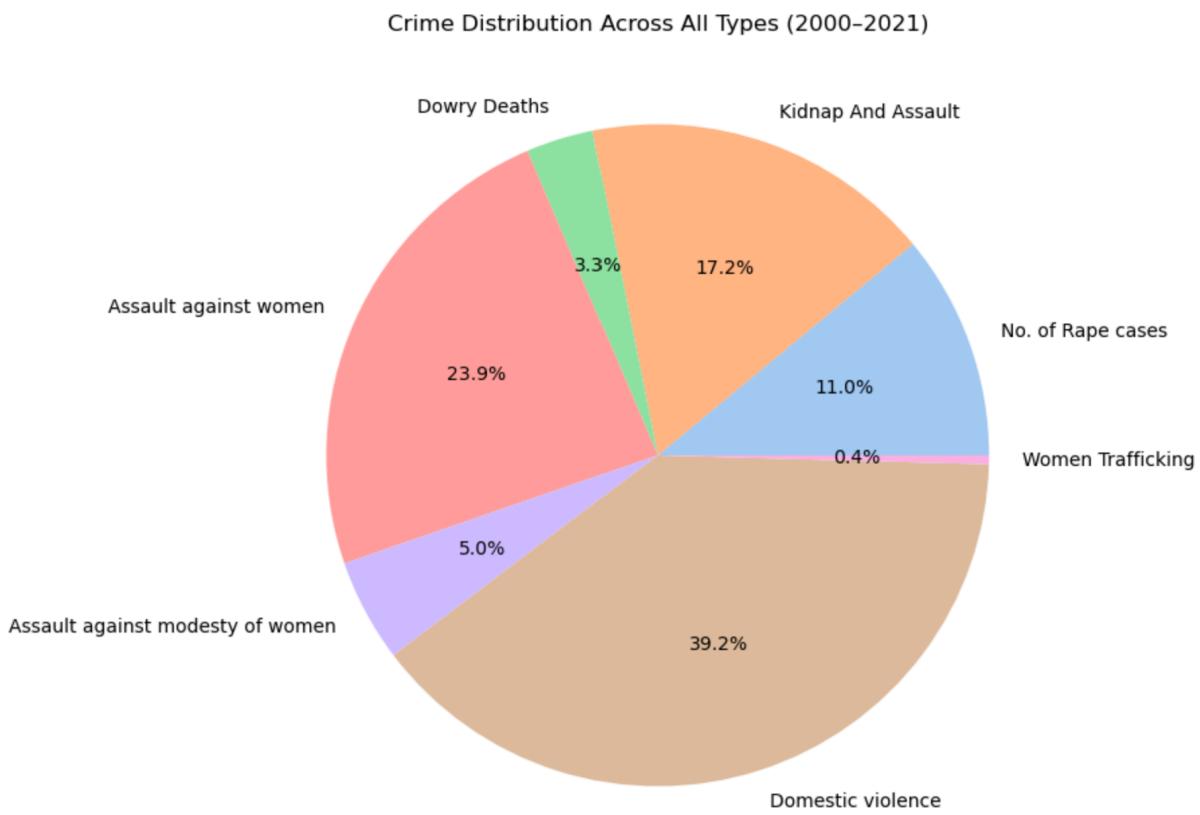
- **Community Policing**

Community-based policing initiatives can help build trust between law enforcement and the local population. By involving community leaders and social workers, local crime rates can be better monitored and addressed.

- **Victim Support and Rehabilitation**

Expanding victim support networks, including shelters, hotlines, and counseling services, is essential for helping victims recover and reintegrate into society. Special attention should be given to rural areas where such services are often lacking.

Visualization



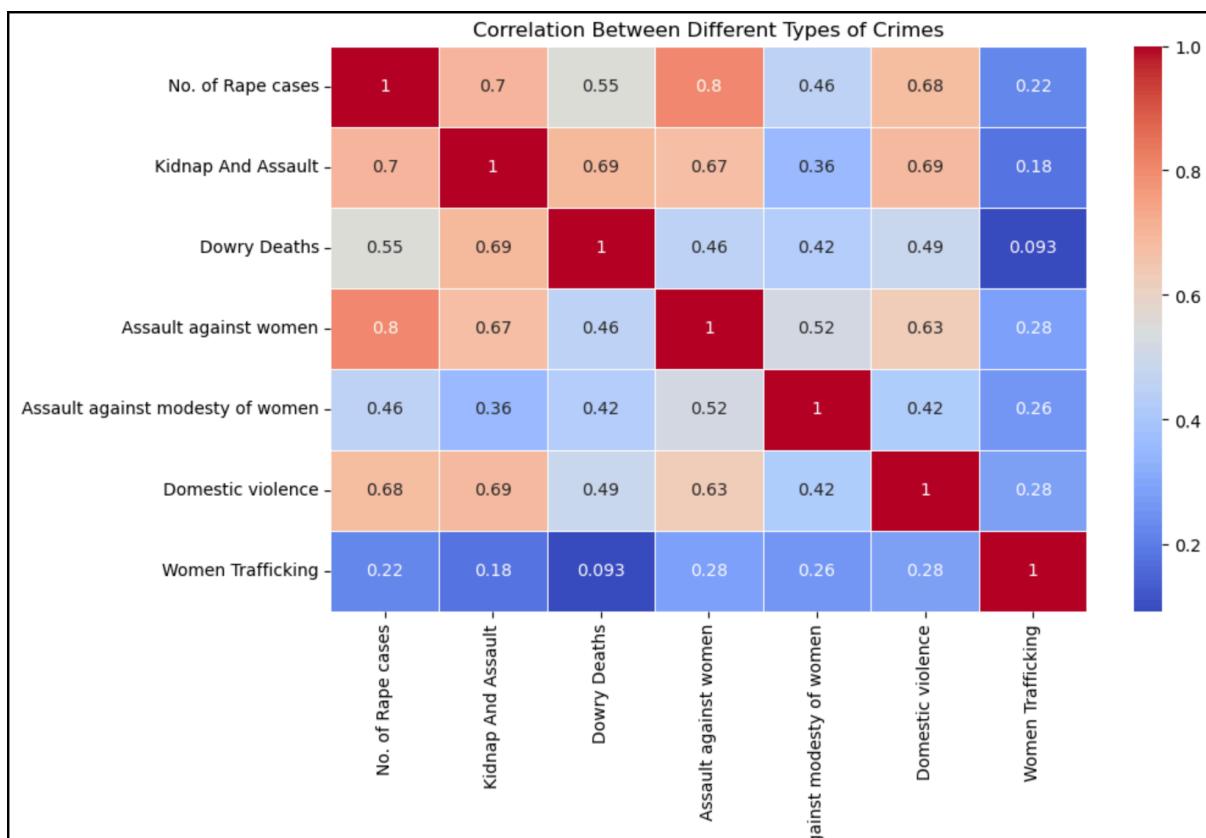
1. Crime Distribution Across all types (2000-2021)

- **Domestic Violence** accounts for the largest share of crimes, making up **39.2%** of the total. This suggests a pressing issue of abuse within households, indicating the need for stronger legal and social measures to protect women in domestic settings.
- **Assault against Women** is the second largest category, representing **23.9%**. This category typically includes physical violence and other forms of aggression, highlighting societal issues related to gender-based violence.
- **Kidnapping and Assault** represents **17.2%**, indicating the significant risk of abduction and physical harm faced by women.
- **Rape Cases** account for **11%** of the total, a critical issue that receives significant attention in both public discourse and legal reforms.
- **Assault against Modesty of Women** forms **5.0%** of reported crimes. This category likely includes sexual harassment and other forms of inappropriate behavior toward women.
- **Dowry Deaths**, although making up a smaller percentage (**3.3%**), represent a severe and often deadly form of gender-based violence tied to marriage customs.
- **Women Trafficking** accounts for **0.4%**, though small in percentage, it reflects a deeply rooted and inhumane crime with long-term consequences on victims.

General Observations:

- The chart confirms that **domestic violence** and **physical assault** are the most prevalent forms of crimes against women in India over the years.
- Rape** and **kidnap and assault** are also significant but comparatively lower in frequency.
- The **modesty of women**, **dowry deaths**, and **women trafficking** categories have lower reported incidences, but they remain critical issues that require attention.
-

2. Correlation between different types of crimes over the year



Strong Correlations:

- Rape and Assault against Women:** The highest correlation (0.8) suggests that areas with more rape cases are also likely to experience a higher number of general assaults against women.
- Rape and Kidnap/Assault:** A strong correlation (0.7) exists between rape cases and kidnapping/assault cases, indicating that such crimes often occur together in similar regions or circumstances.

Domestic Violence and Other Crimes:

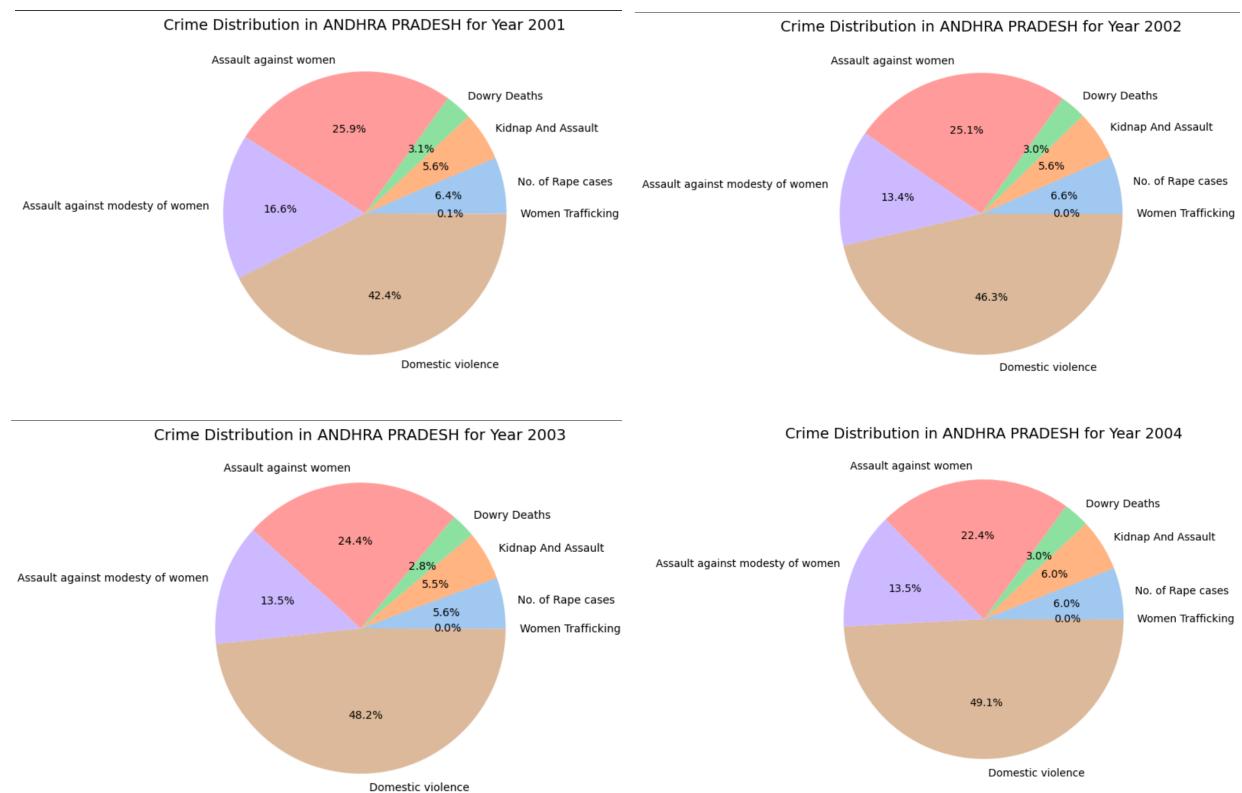
- Domestic violence shows notable correlations with kidnapping (0.69) and general assault (0.63), highlighting its possible overlap with other forms of physical abuse.

Weak Correlations:

- **Women Trafficking:** Correlations with other crime types are relatively low, with the highest being with domestic violence (0.28). This suggests that trafficking might occur under different socio-economic factors compared to more common forms of violence against women.
- **Dowry Deaths:** This type of crime exhibits moderate correlations with kidnapping (0.69) and domestic violence (0.49), indicating its potential linkage with familial or societal pressures.

Assault against Modesty: Exhibits only moderate correlations with other crimes, implying that it may occur under different contexts compared to violent crimes like rape and assault.

3. Crime Distribution of each type for each State & UTs per year



And so on

Insights for Crime distribution in Andhra Pradesh for the year 2001

Domestic Violence Dominates: Domestic violence accounted for 42.4% of crimes against women, making it the most prevalent crime. This highlights the need for intervention in family and home settings, perhaps due to social norms and the underreporting of other forms of violence within households.

Assault Against Women: The second-largest segment is "Assault against women," with 25.9%. This reflects broader physical violence and might include public and private acts of aggression against women.

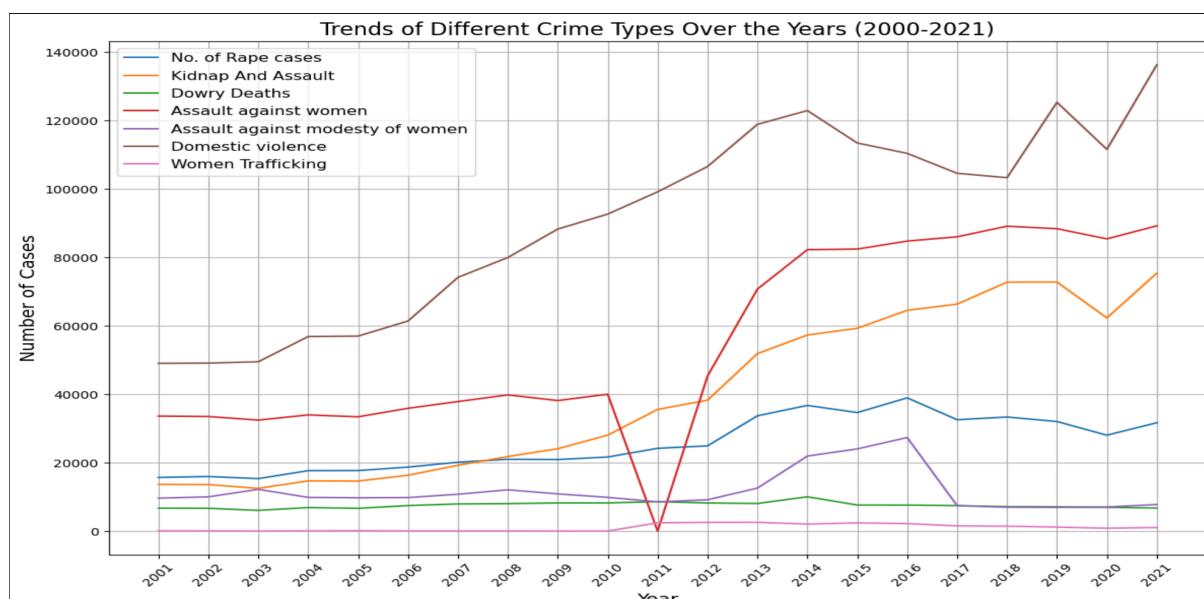
Significant Incidence of Modesty-related Crimes: Crimes "against modesty of women" make up 16.6%. This points to widespread cultural issues around women's personal autonomy and dignity, reinforcing the need for educational reform and awareness.

Low Reporting or Less Prevalent Crimes: Dowry deaths (3.1%), Kidnap and Assault (5.6%), and No. of Rape cases (6.4%) together constitute a smaller share but represent severe forms of violence that have long-lasting psychological and physical impacts.

Women Trafficking: At just 0.1%, trafficking appears minimal, although it is possible that this figure reflects severe underreporting, as trafficking crimes often occur covertly.

Likewise we have collected crime distribution for each type of crimes for each State & UTs per year.

4. Trends of different Crime Types over the years (2001-2021)



Domestic Violence (brown line):

- This type of crime shows a significant and steady increase over the years, peaking around 2020 with over 130,000 cases. This suggests a concerning rise in domestic violence incidents during this period.

Assault against Women (red line):

- There was a sharp drop in 2011, but the numbers surged dramatically in 2012. After 2013, the trend stabilized but remains at a high level, around 85,000 cases. This may indicate societal or policy changes affecting reporting.

Kidnap and Assault (orange line):

- A consistent rise is seen in kidnap and assault cases from around 2010 onwards, peaking in 2020 with approximately 75,000 cases.

No. of Rape Cases (blue line):

- Rape cases have gradually increased over time, with a peak around 2015–2016. However, after 2016, the numbers seem to stabilize between 25,000 to 35,000 cases annually.

Assault against Modesty of Women (purple line):

- After a sharp rise in 2012, the cases dropped after 2014, stabilizing at a much lower level around 2016 and onward.

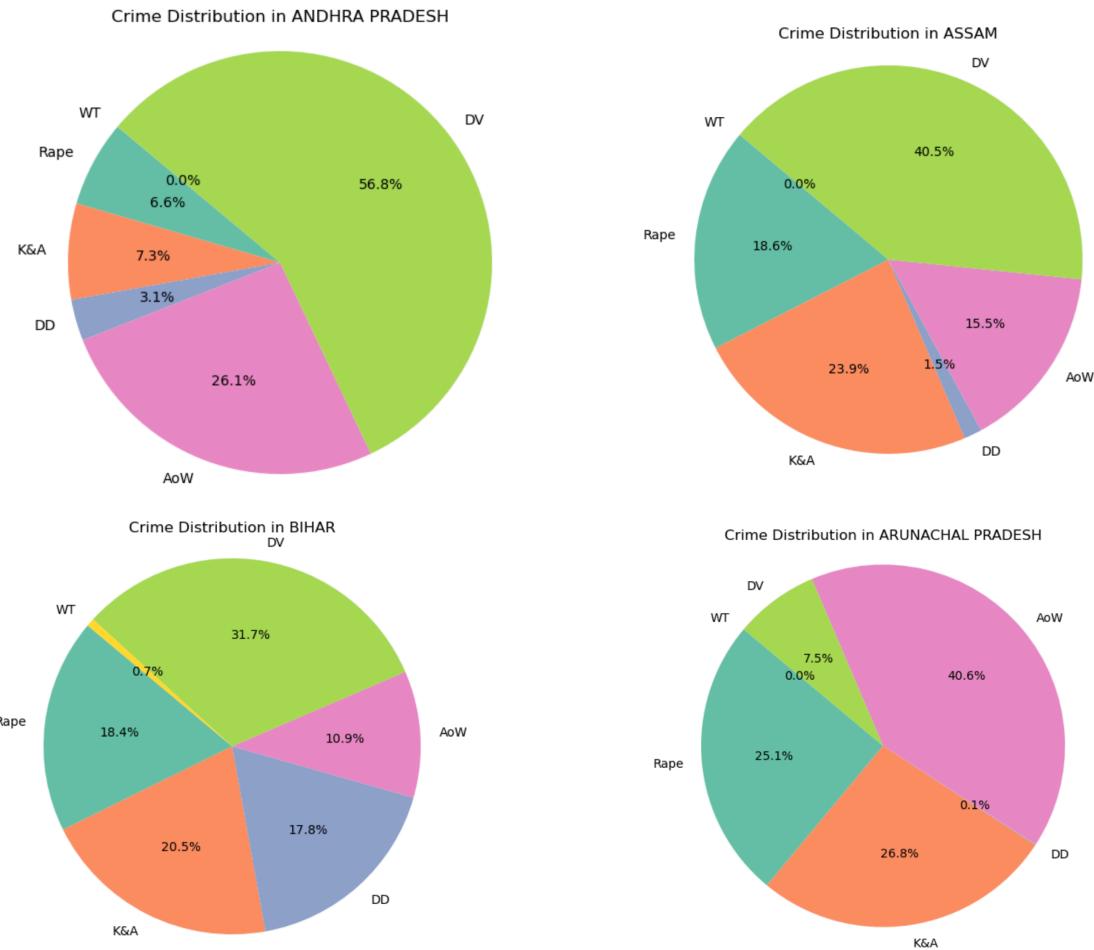
Dowry Deaths (green line):

- Dowry deaths appear relatively stable over the years with a slight decline, hovering between 6,000 and 8,000 cases annually.

Women Trafficking (pink line):

- Women trafficking cases have stayed quite low compared to other crimes, with minor fluctuations across the years, peaking around 2013 and then decreasing.

5. State wise Crime distribution over the year.



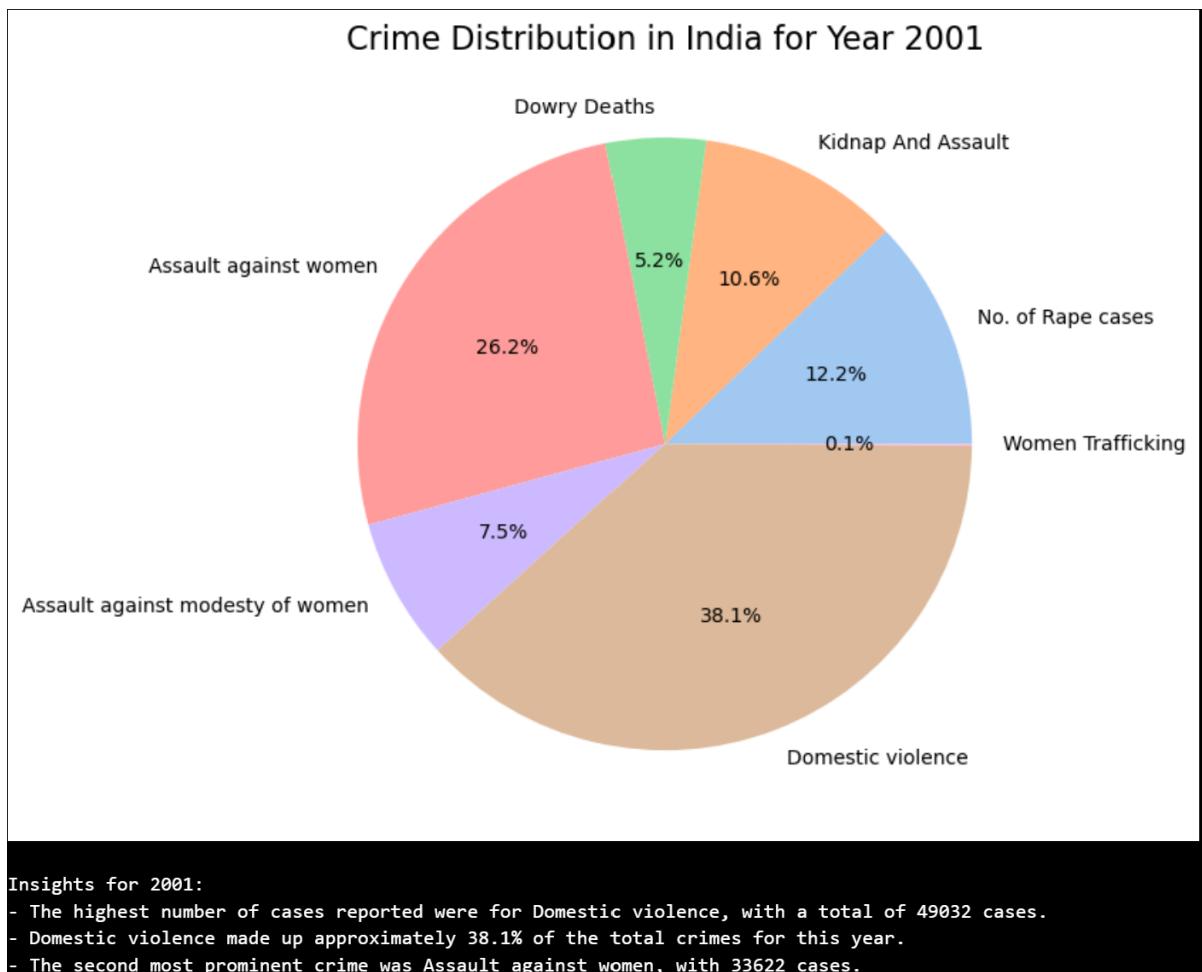
And son on.....

Likewise we have similar visualization for all the States/Uts crime distribution over the years. For example insights for Andhra Pradesh crime distribution over the year has been mentioned below.

- **Domestic Violence (DV)** is overwhelmingly the most prevalent crime in Andhra Pradesh, accounting for **56.8%** of the total crime distribution. This suggests that domestic violence is a significant concern in the state and requires focused attention for prevention and intervention.
- **Assault on Women (AoW)** follows next with **26.1%**, indicating a substantial portion of crimes involving direct physical harm to women. This aligns with broader trends in India, where assault cases remain a critical issue.
- **Kidnap & Abduction (K&A)** represents **7.3%** of the crimes, showcasing a smaller but notable concern, likely involving issues such as human trafficking or forced marriages.
- **Dowry Deaths (DD)** make up **3.1%** of the crimes, reflecting ongoing issues related to dowry disputes, which still remain an unfortunate cultural problem.
- **Rape** accounts for **6.6%** of the crime distribution. While this is not the highest percentage, it is still a critical issue considering its devastating impact on victims and communities.

- **Women Trafficking (WT)** is marked as **0.0%**, which could indicate underreporting or a minimal occurrence of such cases, though it might still require further investigation to ensure accuracy in reporting.

6. Crime distribution in India against women per year wise (2001-2021)



We have collected insights of all seven types of crimes mentioned in the dataset year-wise(2001-2021). For example: Here's the insights for the year 2001:

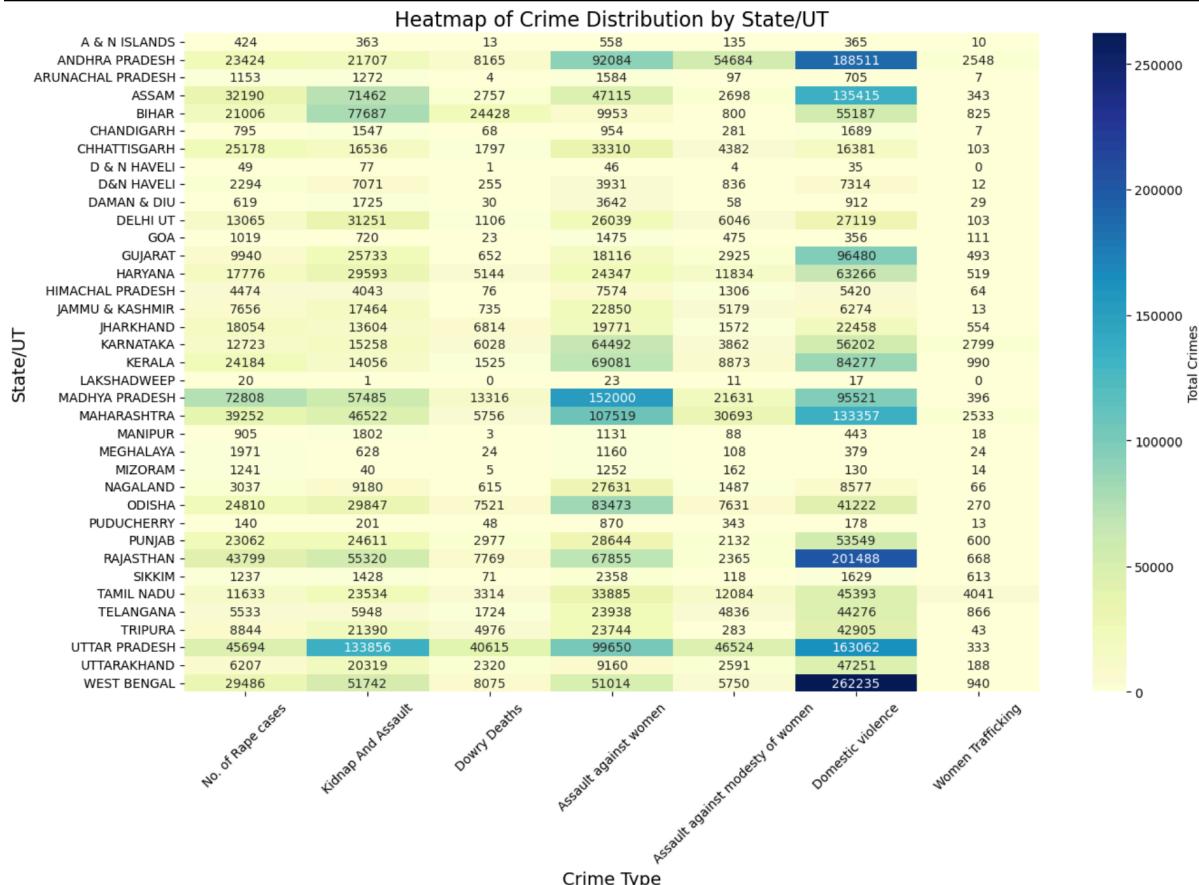
1. **Domestic Violence (38.1%)**:
 - The highest percentage of crimes reported were cases of domestic violence, comprising **38.1%** of the total. This reflects a significant problem in society, as nearly two-fifths of the cases were related to domestic abuse.
2. **Assault Against Women (26.2%)**:
 - The second-largest category of crimes was assault against women, making up **26.2%** of the total. This indicates a high prevalence of physical assaults against women during that time.
3. **Rape Cases (12.2%)**:

- **12.2%** of the reported crimes were rape cases, marking it as another major issue in terms of violence against women.
- 4. **Kidnap and Assault (10.6%)**:
 - **10.6%** of the crimes involved kidnapping and assault, highlighting this as a significant concern in 2001.
- 5. **Assault Against Modesty of Women (7.5%)**:
 - Crimes involving the modesty of women accounted for **7.5%**, pointing to considerable issues related to sexual harassment and molestation.
- 6. **Dowry Deaths (5.2%)**:
 - **5.2%** of crimes were related to dowry deaths, showing that this cultural practice continued to lead to fatalities.
- 7. **Women Trafficking (0.1%)**:
 - Human trafficking, though a grave crime, was reported at a much lower rate, contributing just **0.1%** to the total cases.

General Insights:

- Domestic violence was the most prevalent crime in 2001, followed closely by physical assault against women.
- Serious crimes like rape, kidnap, and dowry deaths also held a significant portion of the crime statistics.
- Women trafficking was relatively low in reported cases, but this does not necessarily indicate less occurrence—it may reflect underreporting or different criminal dynamics.

7. Heatmap of Crime Distribution by State/UT



Crime Concentration:

- The heatmap highlights variations in crime categories across different states/UTs.
- States like **Uttar Pradesh**, **Madhya Pradesh**, **West Bengal**, and **Maharashtra** display high crime volumes, especially in categories like kidnapping, dowry deaths, and domestic violence, as indicated by the darker shades of blue.

Rape Cases:

- Madhya Pradesh** stands out with the highest number of rape cases (72,808), followed by **Rajasthan** and **Uttar Pradesh**, indicating that rape is a major issue in these states.

Kidnapping & Abduction:

- Uttar Pradesh** has an overwhelming number of cases of kidnapping and abduction (133,856), significantly higher than other states, followed by **Madhya Pradesh**.

Dowry Deaths:

- The prevalence of dowry deaths is notably high in **Uttar Pradesh** and **Bihar**, making them critical states for intervention in this area.

Assault on Women's Modesty:

- States like **West Bengal** and **Assam** show particularly high numbers of assaults with the intent to outrage a woman's modesty, which could point to societal issues regarding gender-based violence.

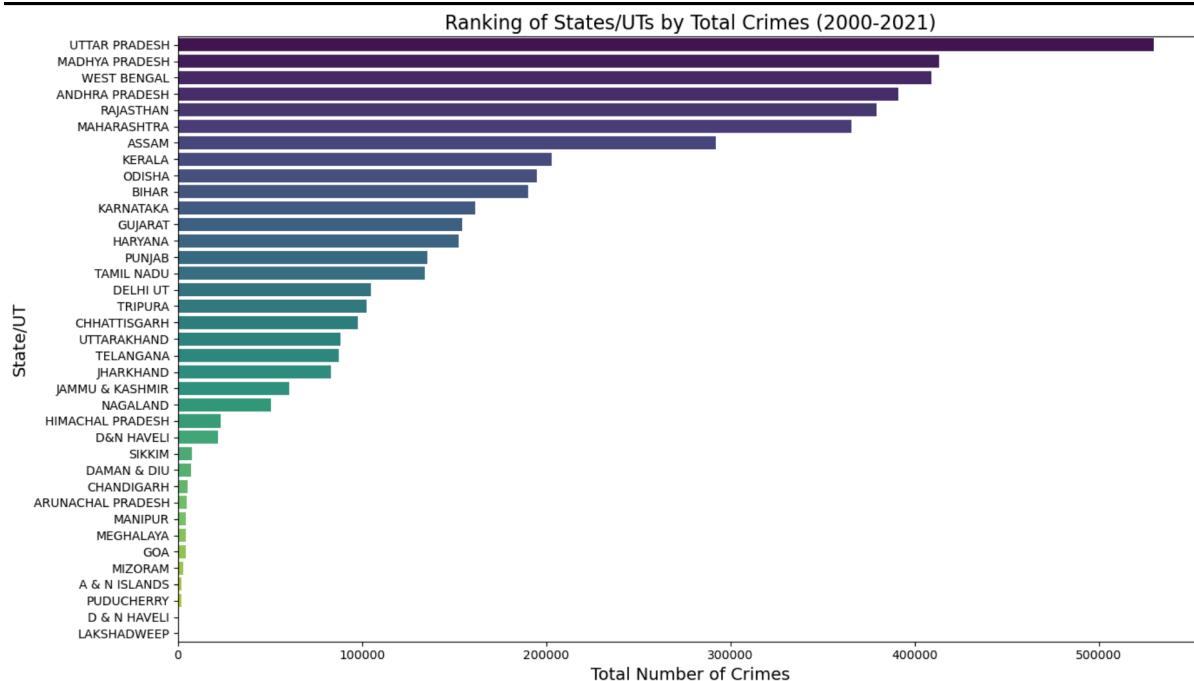
Domestic Violence:

- West Bengal** has a significant concentration of domestic violence cases (262,235), making it a standout region for this issue.

Low Crime Reporting:

- Union Territories like **Lakshadweep**, **Daman & Diu**, and **Puducherry** show minimal reporting across crime categories. While this could indicate fewer crimes, it might also reflect underreporting or smaller populations.

8. Ranking of States/UTs by total Crimes



Crime Volume Leaders:

- Uttar Pradesh** leads with the highest total number of crimes against women from 2000 to 2021, surpassing 500,000 reported cases. This is followed by **Madhya Pradesh** and **West Bengal**. These states should be prioritized for policy intervention and resource allocation.

Top 10 High-Crime States:

- States like **Andhra Pradesh, Rajasthan, Maharashtra, Assam, Kerala**, and **Odisha** also rank highly, indicating that crimes against women are not confined to specific regions but are spread across both the north and south of India.

Moderate Crime States:

- **Punjab, Gujarat, and Karnataka** show moderately high crime rates but are not as prominent as Uttar Pradesh or Madhya Pradesh.

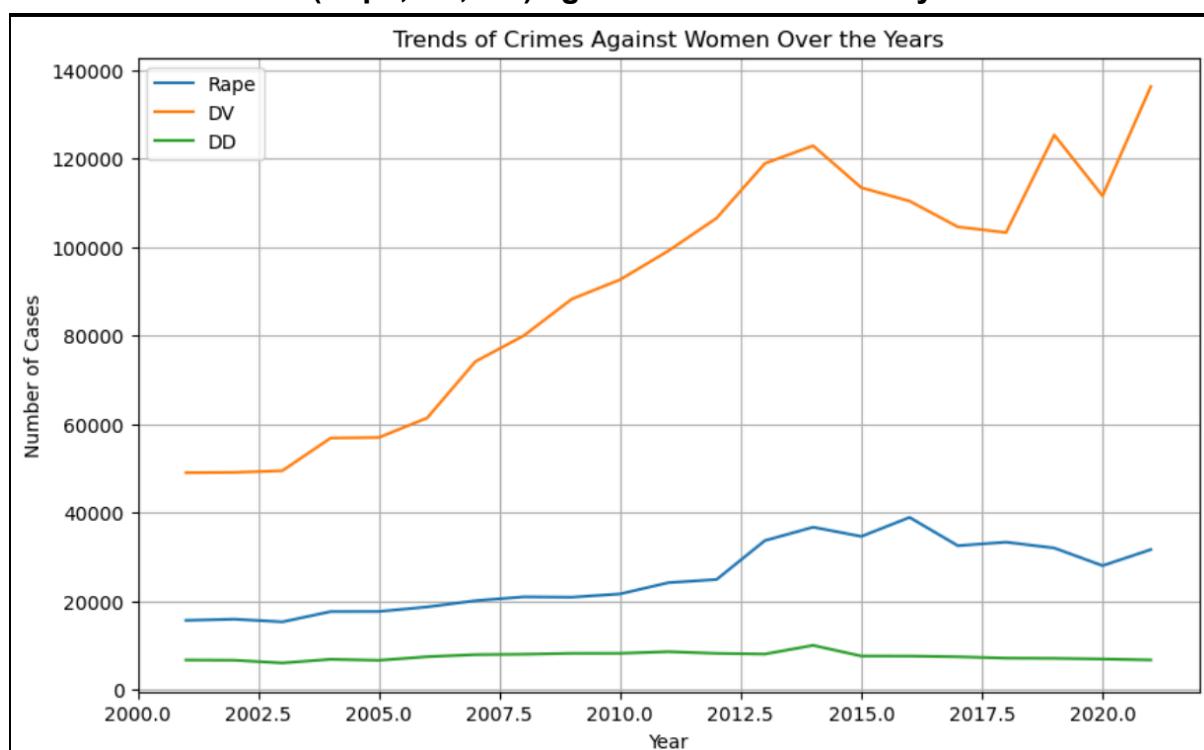
Low-Crime States and UTs:

- **Lakshadweep, Daman & Diu, and Mizoram** report very low crime numbers, possibly due to smaller populations and possibly underreporting.

Policy Implications:

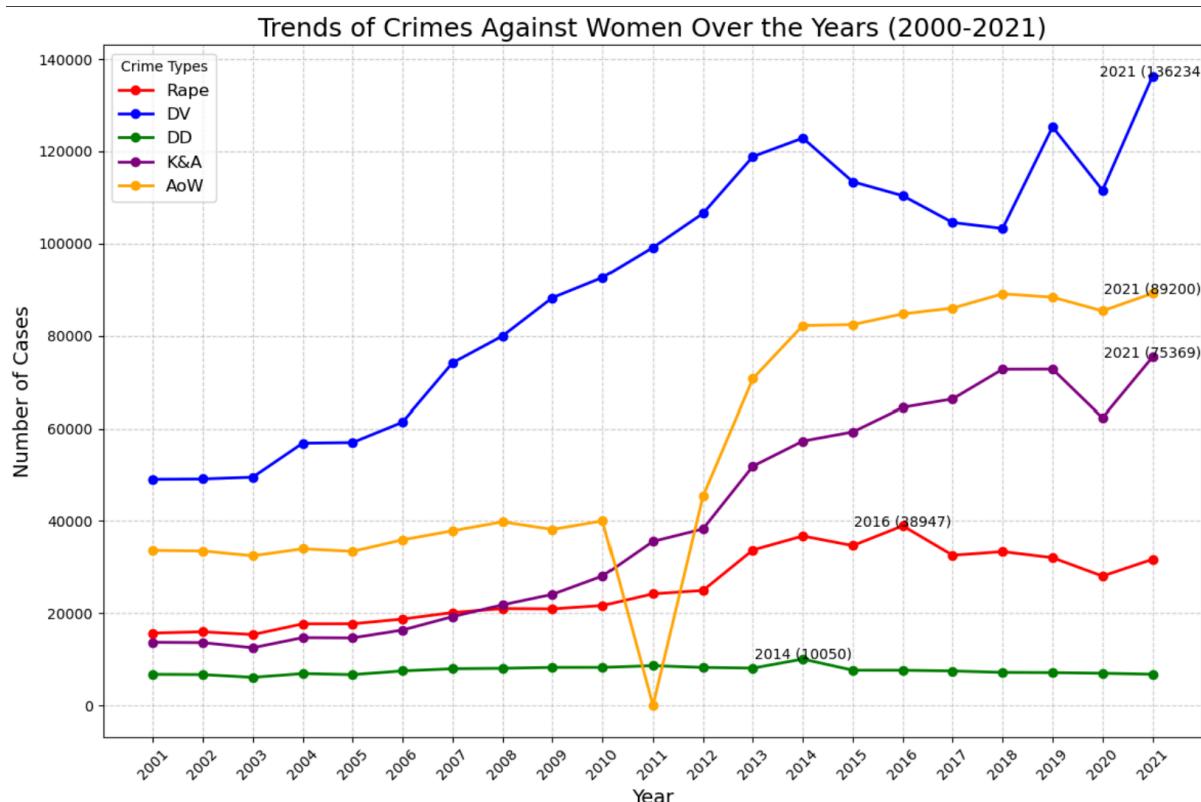
- The ranking of states provides clear targets for both localized and national-level interventions. States like **Uttar Pradesh** and **Madhya Pradesh** need the most urgent focus for improving women's safety, while states with low crime reports should be analyzed to ensure that crime is not being underreported.

9. Trends of Crimes (Rape, DV, DD) against women over the years



The graph displays the trends of three types of crimes: Rape, Domestic Violence (DV), and Dowry Death (DD) from 2000 to 2020.

- **Domestic Violence (DV)** cases (orange line) show a sharp and continuous rise over the years, peaking in 2020 with nearly 140,000 cases.
- **Rape** cases (blue line) show a relatively stagnant trend with a slight increase around 2015, fluctuating between 18,000 and 25,000 cases during the period.
- **Dowry Death (DD)** cases (green line) remain fairly steady with minimal change, showing a very small increase over time.

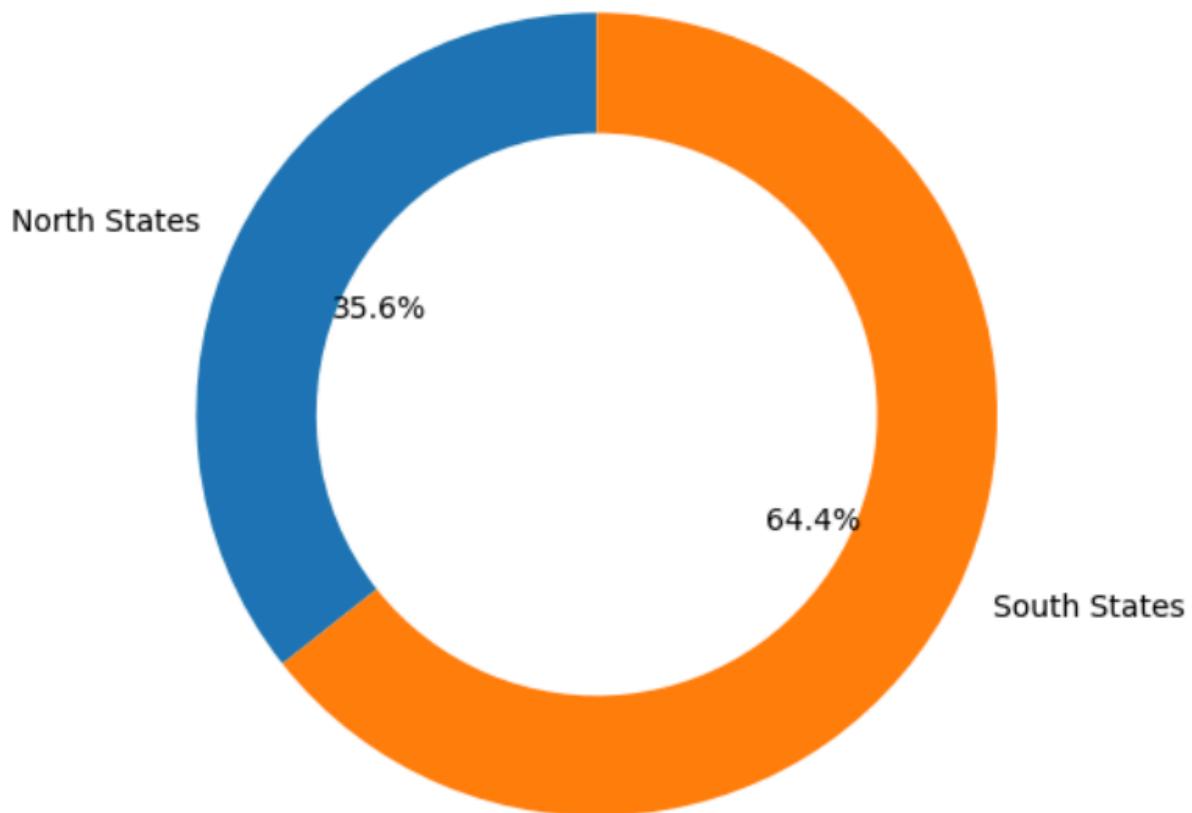


- Rape (red) remains relatively steady with a slight increase post-2010. The peak in 2021 is around 36,234 cases.
- Domestic Violence (DV, blue) shows a consistent rise, similar to the first chart. It peaks at 136,234 in 2021.
- Dowry Death (DD, green) remains stable with very minimal change. The lowest point is in 2014 (around 10,050 cases).
- Kidnapping & Abduction (K&A, purple) and Assault on Women (AoW, yellow) show significant increases, especially post-2010, with AoW peaking at 89,200 in 2021, while K&A reaches 53,369 in 2021.

10. Region wise comparison for total crimes over the years for each specific crime type: North and South

```
north_states = ['JAMMU & KASHMIR', 'HIMACHAL PRADESH', 'PUNJAB',  
'HARYANA', 'UTTARAKHAND', 'DELHI UT']  
  
south_states = ['ANDHRA PRADESH', 'TELANGANA', 'TAMIL NADU',  
'KARNATAKA', 'KERALA']
```

Rape Comparison: North vs South



Here's an example for Rape, likewise, we have visualization for each crime types mentioned in the dataset

Comparing North and South States/UTs crime distribution over the year for each types of crimes. For example, mentioned above image is Rape comparison of North and South States/UTs.

- This pie chart shows the distribution of rape cases between North and South states.

- South States account for the majority, with 64.4% of the total rape cases, while North States represent 35.6%.
- This could indicate a higher prevalence of reported rape cases in the South compared to the North.

Rape: East (48%) vs West (52%)

Kidnap & Assault: East (60%) vs West (40%)

Dowry Deaths: East (68%) vs West (32%)

Assault Against Women: East (55%) vs West (45%)

Assault Against Modesty: East (42%) vs West (58%)

Domestic Violence: East (57%) vs West (43%)

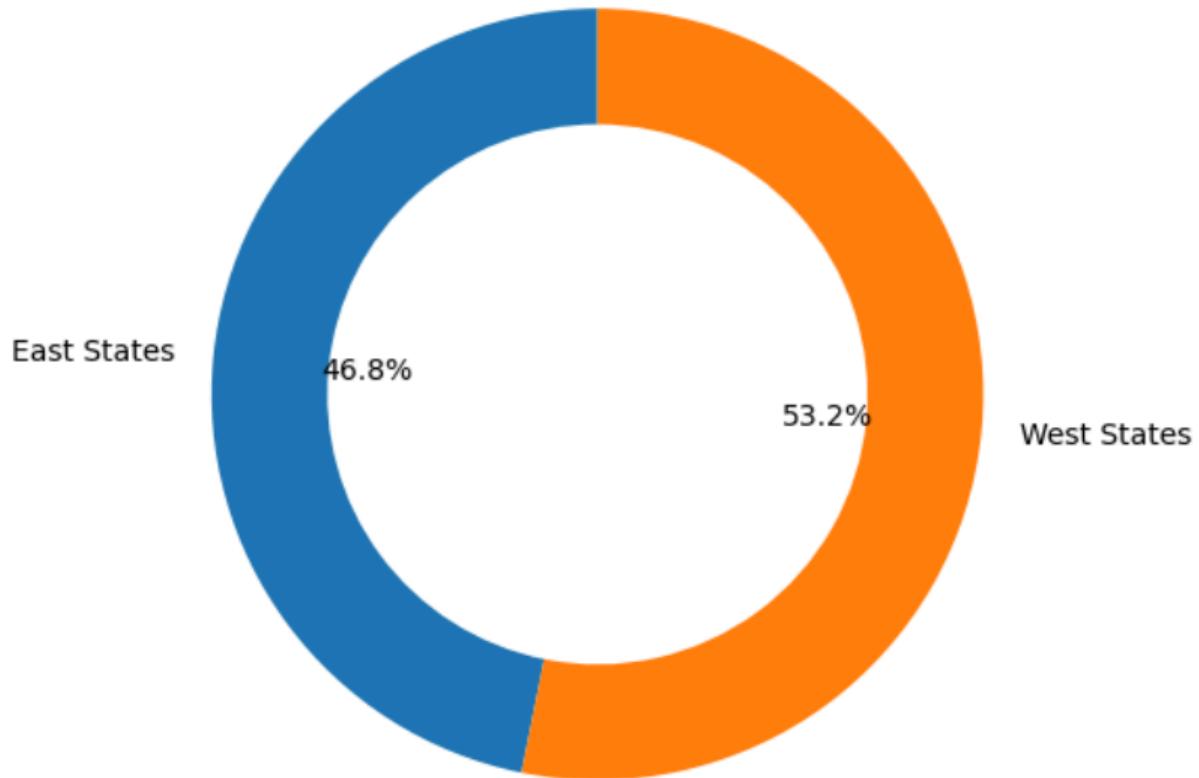
Women Trafficking: East (65%) vs West (35%)

11. Region wise comparison for total crimes over the years for each specific crime type: East and West

```
east_states = ['WEST BENGAL', 'ODISHA', 'BIHAR', 'JHARKHAND', 'SIKKIM',
              'ASSAM', 'ARUNACHAL PRADESH', 'MANIPUR', 'MEGHALAYA',
              'MIZORAM', 'NAGALAND', 'TRIPURA']

west_states = ['MAHARASHTRA', 'GUJARAT', 'RAJASTHAN', 'GOA', 'MADHYA
PRADESH',
              'CHHATTISGARH', 'DADRA & NAGAR HAVELI', 'DAMAN & DIU']
```

Rape Comparison: East vs West



Breakdown:

- The chart shows a comparison between East and West states concerning rape cases.
- West States have a higher percentage of rape cases, accounting for 53.2%.
- East States follow closely, contributing 46.8% to the total cases.

Insight:

- The distribution of rape cases between the two regions is fairly balanced, but the West states have a slightly higher rate. This suggests that the problem is not localized to one area but is an issue across both regions, with the West experiencing a marginally larger burden.

Rape: East (49%) vs West (51%)

Kidnap & Assault: East (62%) vs West (38%)

Dowry Deaths: East (65%) vs West (35%)

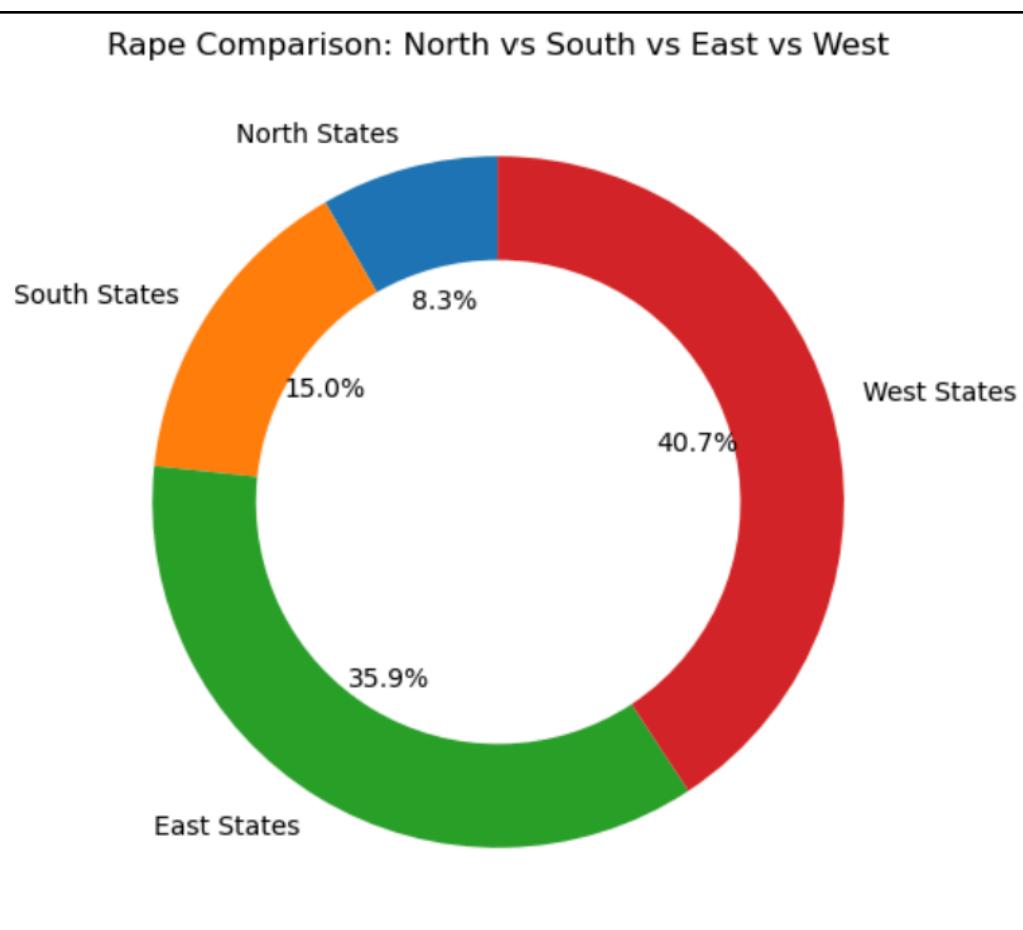
Assault Against Women: East (57%) vs West (43%)

Assault Against Modesty: East (45%) vs West (55%)

Domestic Violence: East (56%) vs West (44%)

Women Trafficking: East (68%) vs West (32%)

12. Region wise comparison for total crimes over the years for each specific crime type: East, west, north and south



```
north_states = ['JAMMU & KASHMIR', 'HIMACHAL PRADESH', 'PUNJAB',
'HARYANA',

'UTTARAKHAND', 'DELHI UT']

south_states = ['ANDHRA PRADESH', 'TELANGANA', 'TAMIL NADU',
'KARNATAKA', 'KERALA']

east_states = ['WEST BENGAL', 'ODISHA', 'BIHAR', 'JHARKHAND', 'SIKKIM',
'ASSAM', 'ARUNACHAL PRADESH', 'MANIPUR', 'MEGHALAYA',
'MIZORAM', 'NAGALAND', 'TRIPURA']

west_states = ['MAHARASHTRA', 'GUJARAT', 'RAJASTHAN', 'GOA', 'MADHYA
PRADESH',
'CHHATTISGARH', 'DADRA & NAGAR HAVELI', 'DAMAN & DIU']
```

Breakdown:

- This chart provides a more detailed comparison of rape cases across four regions: North, South, East, and West.
- West States lead with 40.7% of rape cases.
- East States contribute 35.9%.
- South States make up 15% of the total.
- North States have the lowest percentage, at 8.3%.

Insight:

- The West and East States continue to dominate in terms of reported rape cases. However, the North has a significantly lower share, indicating either fewer reported incidents or potentially lower population density or reporting mechanisms.
- The South shows a moderate percentage, highlighting regional variation in the frequency of cases.

- The stark contrast between North and West, with nearly five times the difference, suggests that some regions are facing much larger challenges in addressing and reporting rape cases.

Rape: North (25%), South (23%), East (28%), West (24%)

Kidnap & Assault: North (26%), South (19%), East (34%), West (21%)

Dowry Deaths: North (32%), South (15%), East (36%), West (17%)

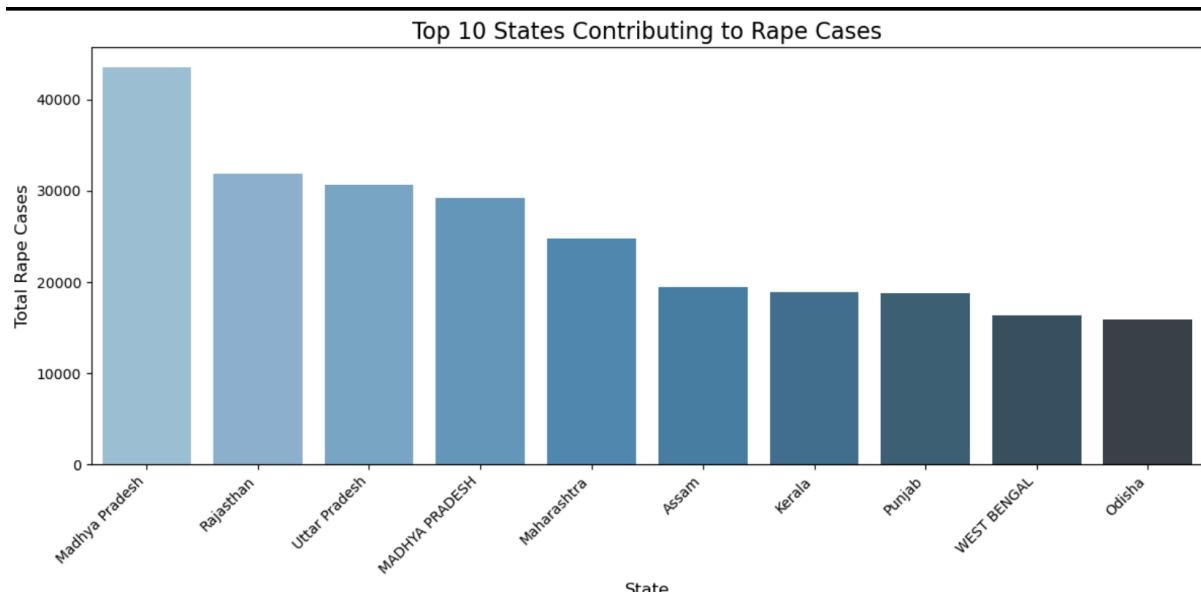
Assault Against Women: North (29%), South (19%), East (32%), West (20%)

Assault Against Modesty: North (33%), South (17%), East (28%), West (22%)

Domestic Violence: North (28%), South (19%), East (31%), West (22%)

Women Trafficking: North (14%), South (16%), East (45%), West (25%)

13. Crime trends of top 10 states/UTs for each specific crime type



And so on.....

Here's an example for top 10 states for rape cases

- Insight:**

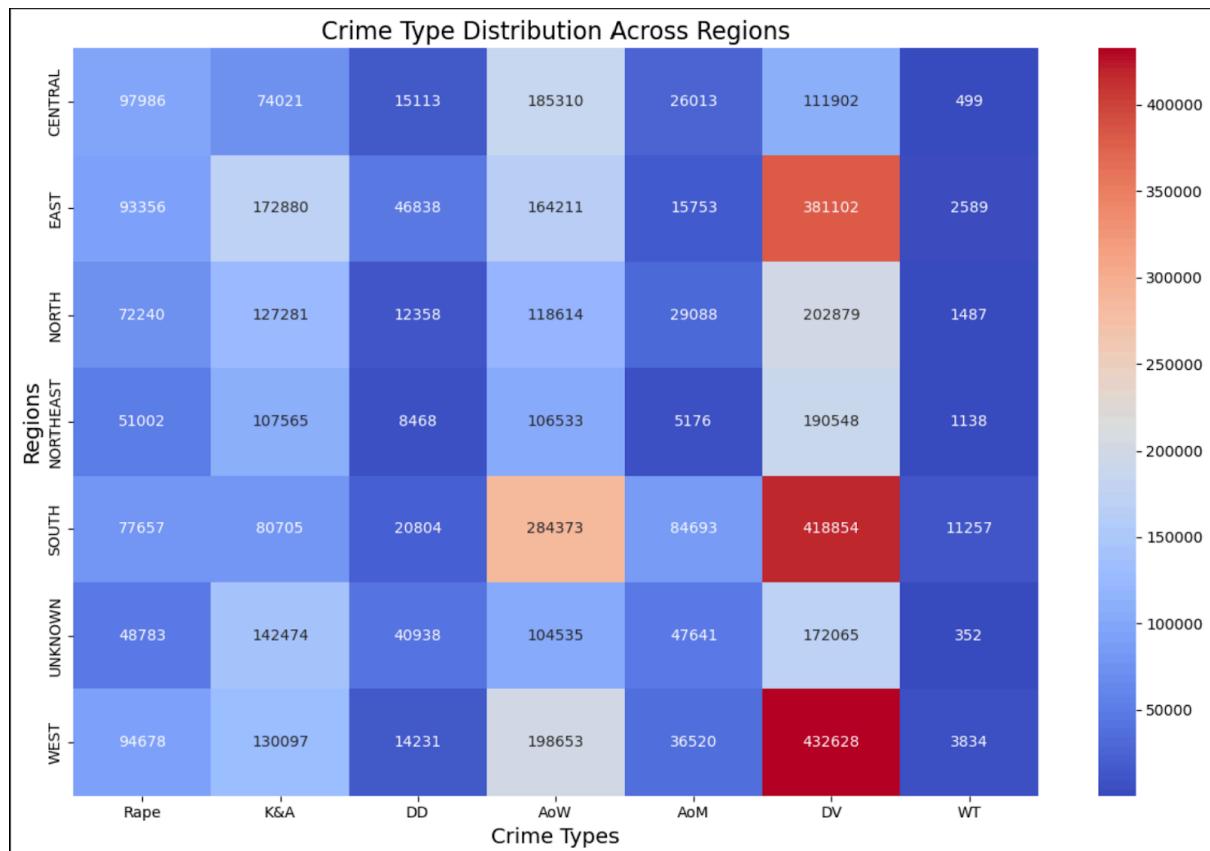
- Madhya Pradesh is the leading contributor by a significant margin, followed by Rajasthan and Uttar Pradesh, all of which are northern or central states.
- This suggests that the central and northern regions of India, including Madhya Pradesh and Uttar Pradesh, are facing severe issues related to sexual violence.
- The other states (e.g., Maharashtra, Assam, West Bengal) span across the rest of India, showing that rape is a widespread issue affecting states across the country, though with varying degrees of severity.

General Insight:

- West and East states appear consistently in the data as regions with higher rates of reported rape cases, particularly when compared to the North and South.
- Madhya Pradesh, Rajasthan, and Uttar Pradesh are the states with the most alarming numbers, likely necessitating more targeted policy interventions and support systems.
- The presence of states from all over India in the third chart indicates that rape is not confined to one geographical area, but rather a national issue with particular hotspots in specific states.

13. Violent and non violent crimes visualization

```
violent_crimes = df[['Rape', 'DD']]  
  
non_violent_crimes = df[['DV', 'AoW']]
```



Key Observations:

- Domestic Violence (DV)** seems to be significantly more prevalent in almost all regions, especially in the **West (432,628 cases)**, **South (418,854 cases)**, and **East (381,102 cases)**. This shows a concerning trend of DV, particularly in these areas.
- Assault on Women (AoW)** is high in regions like **South (284,373 cases)** and **West (198,653 cases)**, indicating that these areas may need targeted interventions to protect women.
- Rape cases** show a fairly consistent distribution across regions, with the **East (93,356 cases)** and **West (94,678 cases)** having the highest numbers.
- Kidnapping and Abduction (K&A)** is also notably high in the **East (172,880 cases)**, **North (127,281 cases)**, and **West (130,097 cases)** regions.
- Unknown Region** data shows high counts for certain crime types, such as **K&A (142,474 cases)**, possibly indicating gaps in data collection or ambiguities in the regional classification.
- The Color Scale:**
 - Dark red indicates the highest crime counts, while light blue represents the lowest. Domestic violence has the darkest shades, showing its dominance across all regions.
 - Areas like the **South**, **West**, and **East** are more shaded in the red zone, indicating higher overall crime rates compared to regions like **Northeast** and **Central** which have generally lighter shades of blue.

Deep Insight:

- **Domestic Violence** is an overwhelming issue across most regions, particularly in the West and South, signaling a critical need for social programs and strict enforcement of laws to protect victims.
- **Assault on Women** is the second-highest concern, especially in southern and western parts of the country, highlighting a regional gender disparity in crime.
- **The East** shows high rates of kidnapping and abduction, raising concerns about child safety or human trafficking issues in that area.
- **The relatively high number of crimes in the "Unknown" region** may point to inefficiencies in data reporting, which should be addressed for more accurate national or regional crime tracking.

14. Clustering of States based on Crimes



Key Observations:

1. Clusters:

- The states in **Cluster 0 (purple)** seem to be the most concentrated, indicating a group of states with similar crime patterns that likely have lower variation in crime rates.
- **Cluster 1 (blue)** also shows a significant grouping of states, closely packed around the lower-left region, indicating these states may share similar crime characteristics, but slightly differ from Cluster 0.

- **Cluster 2 (green)** is more spread out, which could mean these states have a higher variation in crime types or rates, but they still share some crime pattern similarities.
- **Cluster 3 (yellow)** is the most dispersed and spread out. This cluster could represent states with highly unique crime patterns, either with very high or very low crime rates in different categories.

2. Separation of Clusters:

- **Cluster 3 (yellow)** is positioned farther away from the other clusters, indicating that states in this cluster likely experience very different crime trends compared to those in Clusters 0, 1, and 2.
- **Cluster 2 (green)** has an upward spread, suggesting some states within this group may have unique crime behaviors compared to those clustered tightly near the center.

3. Clusters 0 and 1:

- The proximity between **Cluster 0 and Cluster 1** suggests that the crime patterns in these states are more homogeneous or follow a similar trend. These clusters could represent states with moderate crime rates, not too extreme in any direction.

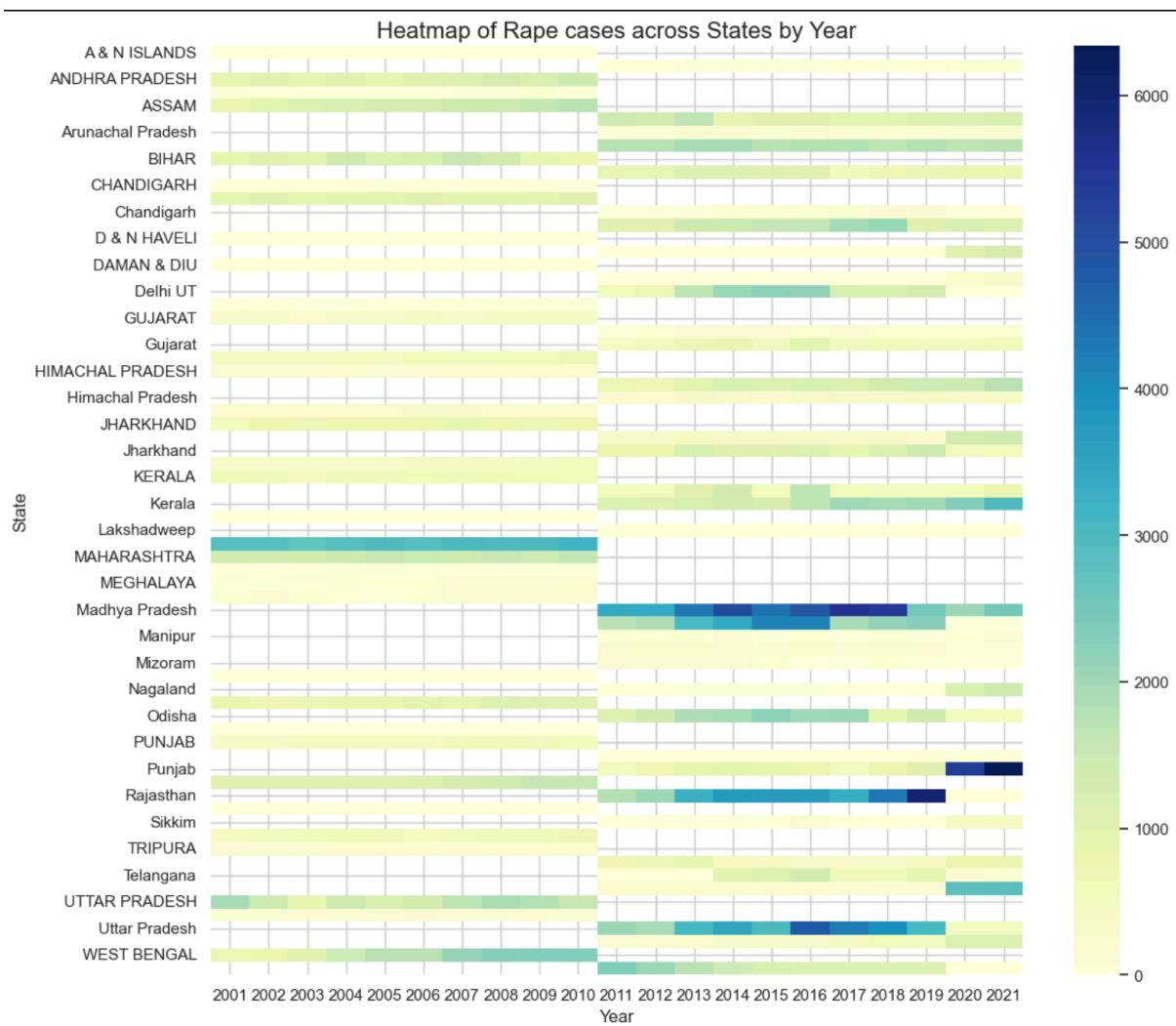
4. Spread in Data:

- The larger spread in **Cluster 3** suggests outlier behavior in crime patterns in certain states, whereas the tight packing of **Cluster 0** suggests more consistency among those states.

Deep Insight:

- The clustering suggests that there are **distinct crime pattern groupings across states**, with **Cluster 0 and Cluster 1** showing states with more uniform crime characteristics. These states might have moderate or average crime rates and share similar socio-economic or law enforcement characteristics.
- **Cluster 3**, with its dispersed points, likely contains states with **anomalies in crime data**, possibly states with either very high crime rates in certain categories or outliers with unique crime trends (e.g., particular crimes dominating the statistics).
- **Cluster 2**, while more spread out than Clusters 0 and 1, still indicates a recognizable pattern but with more variability, perhaps reflecting regional influences or varying crime control policies.
- **Cluster 3's distinct separation** suggests that states within this cluster may need **specialized crime prevention strategies** due to their deviation from more common crime patterns seen in the other clusters.

15. Heatmap of Rap cases across States over the year



And so on.....

We have visualization for each type of crimes. Mentioned below is insights for only rape cases.

Key Observations:

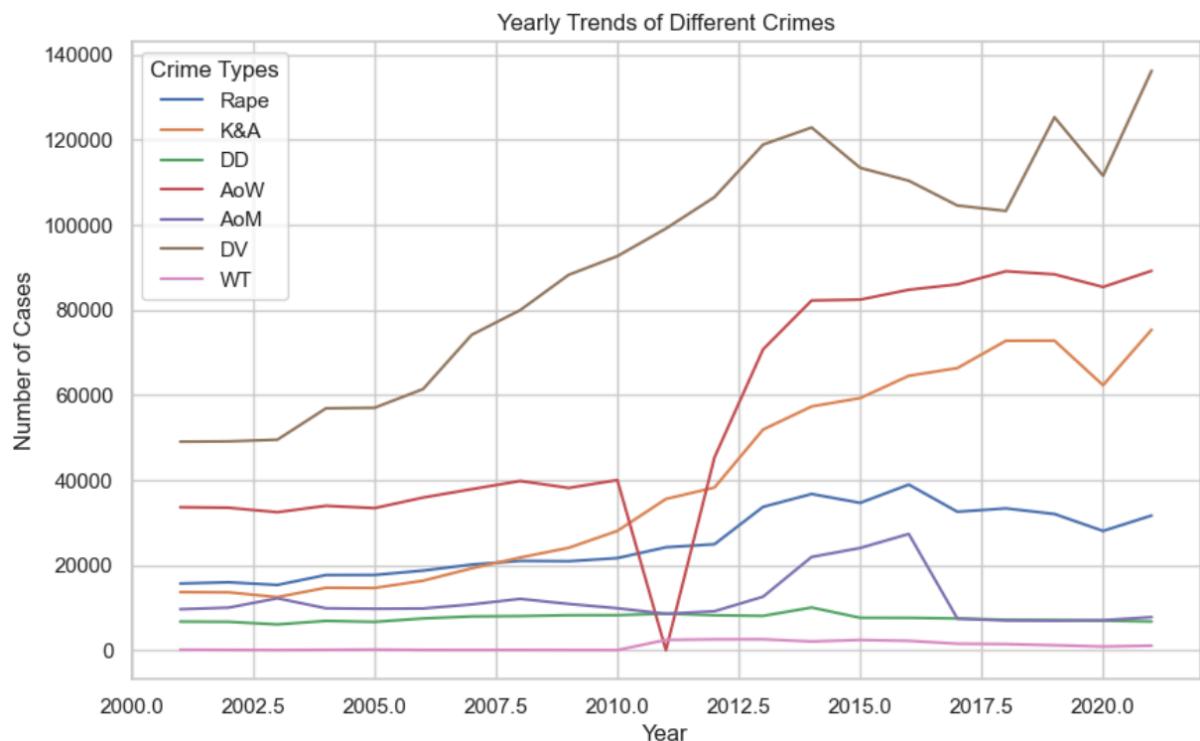
- Significant Increase Post-2012:** There is a marked increase in reported rape cases across many states after 2012. This is particularly visible in states like **Madhya Pradesh, Rajasthan, and Uttar Pradesh**, which show the darkest shades in the heatmap after this year. The 2012 Delhi gang rape (Nirbhaya case) seems to have been a turning point in both public awareness and reporting of such crimes.
- Highest Reporting States:**
 - Madhya Pradesh, Rajasthan, and Uttar Pradesh** consistently report a high number of rape cases, especially in the period from 2012 to 2021. These states stand out with darker shades, indicating a significantly higher volume of cases compared to other states.
 - Maharashtra and West Bengal** also show a noticeable rise in cases, though not as extreme as the top three states.

3. **Low Reporting States/UTs:**
 - Union territories and small states like **Lakshadweep, Nagaland, Mizoram, and Sikkim** have very few reported cases, indicated by the lighter colors throughout the heatmap.
 4. **Recent Trends (2018-2021):**
 - States such as **Rajasthan** and **Uttar Pradesh** continue to show consistently high numbers of reported rape cases even in the more recent years (2018-2021). In contrast, several other states see fluctuations or a decline in reporting during this period.
 5. **Regional Variation:**
 - Northern and central states (such as Madhya Pradesh, Rajasthan, Uttar Pradesh) report far more cases than southern states like Kerala, Tamil Nadu, and Karnataka, which have fewer reported cases, shown by lighter shades.
-

Deep Insights:

1. **Impact of Awareness and Legal Reforms Post-2012:** The rise in reported cases post-2012 can likely be attributed to a combination of increased public awareness, societal shifts, and stricter legal frameworks. The Nirbhaya case and subsequent protests might have led to a greater willingness to report cases of sexual violence, reflecting in the data.
2. **Persistence of High Rates in Certain States:** The persistently high number of cases in states like **Madhya Pradesh, Rajasthan, and Uttar Pradesh** suggests systemic issues such as ineffective law enforcement, deep-rooted gender inequality, or sociocultural factors that continue to perpetuate high levels of sexual violence.
3. **Urban vs. Rural Factors:** States with larger rural populations (e.g., **Rajasthan, Madhya Pradesh, and Uttar Pradesh**) tend to report higher numbers, possibly indicating more reporting from rural areas or that sexual violence may be more pervasive in these regions.
4. **Underreporting in Some Regions:** In smaller states and union territories, such as **Lakshadweep, Mizoram, and Nagaland**, the consistently low number of reported rape cases could indicate underreporting, possibly due to social stigma, lack of awareness, or inadequate infrastructure for handling such cases.
5. **Complexity of Decline in Recent Years:** The slight decline or stabilization in reporting in some states after 2017-2018 might suggest either improvements in preventive measures or changes in the reporting behavior. Alternatively, it could point to systemic issues in reporting mechanisms.

16. Yearly Trends fo different Crimes



Key Observations:

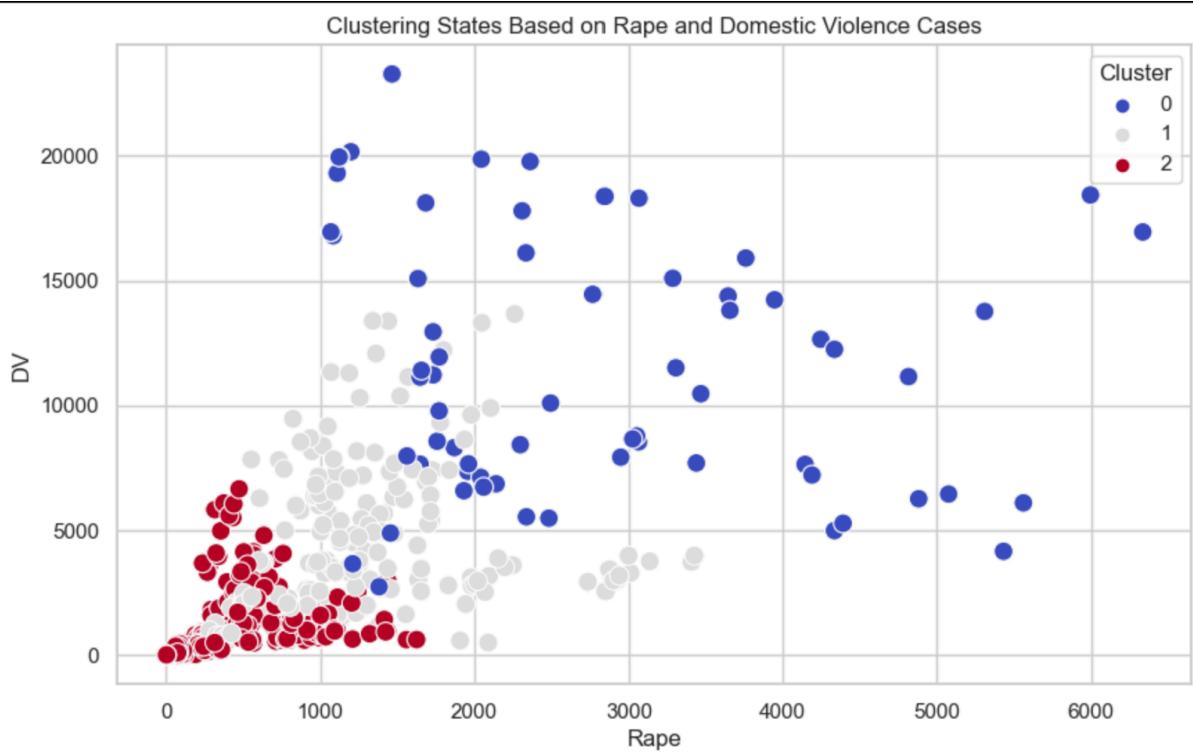
1. **DV (Domestic Violence):**
 - This crime category, represented by the brown line, has the highest number of reported cases. It shows a steady increase from 2000, peaking around 2017, then experiencing some fluctuation and another sharp increase in 2020-2021.
 - The rise in 2020-2021 might be linked to the COVID-19 pandemic, where lockdowns could have contributed to an increase in domestic violence cases.
2. **AoW (Assault on Women):**
 - Represented by the red line, assault on women shows a significant increase around 2010, peaks around 2015-2016, and then levels out, indicating a steep rise in the early 2010s.
3. **K&A (Kidnapping and Abduction):**
 - The orange line represents kidnapping and abduction, which has also seen a steady rise from 2000 to around 2014, showing fluctuations afterward but remaining high.
4. **Rape:**
 - The blue line representing rape cases shows a steady increase starting around 2009-2010, with a plateauing trend after 2015, suggesting heightened awareness and reporting around that time. The trend shows more stability compared to some of the other crime categories.
5. **Other Crimes:**
 - **WT (Trafficking)**, represented by the pink line, remains relatively low and stable throughout the years.
 - **AoM (Assault on Men)**, in purple, shows a slight rise around 2010 but remains lower than AoW.

- **DD (Dowry Deaths)**, represented by the green line, shows a slight decline or stable trend overall.

Deep Insights:

1. **Steady Increase in Gender-Based Violence:**
 - Crimes like **domestic violence**, **rape**, and **assault on women** have shown a consistent upward trend, especially from 2010 onwards. This indicates that gender-based violence remains a persistent issue. The sharp rise around 2012 aligns with societal attention after the Nirbhaya case in India, leading to increased reporting and legal changes.
2. **Impact of Social Awareness and Reporting:**
 - The rise in reported rape and assault cases from around 2010-2015 suggests a heightened awareness about these issues, possibly driven by media coverage, advocacy, and legal reforms. This period marked a shift in how society views and responds to crimes against women.
3. **Pandemic Influence on Domestic Violence:**
 - The sharp rise in **domestic violence** cases in 2020-2021 likely correlates with the COVID-19 pandemic, where lockdowns, economic stress, and isolation may have contributed to a surge in such incidents. It highlights the vulnerability of women in domestic environments during crises.
4. **Kidnapping and Abduction Concerns:**
 - The persistent increase in kidnapping and abduction cases indicates growing concerns about child and women safety in public spaces. Despite fluctuations after 2015, it remains a critical issue.
5. **Dowry Deaths Remain Steady:**
 - The trend for **dowry deaths** shows little change, suggesting that while awareness about other crimes may be growing, this deeply rooted societal problem continues to persist without significant intervention.

17. Clustering States based on Rape and Domestic violence cases



Key Observations:

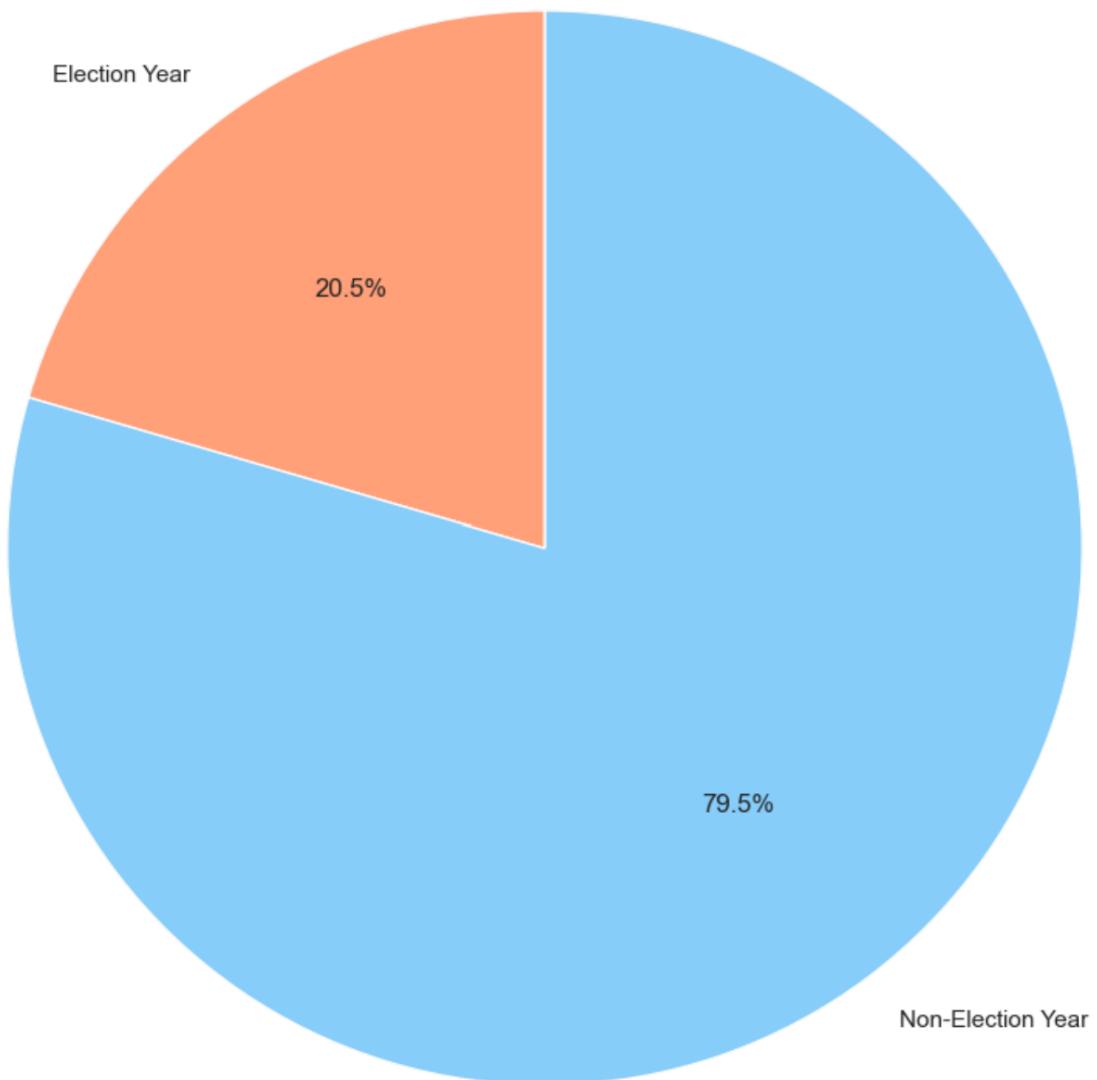
1. **Three Clusters Identified:**
 - The scatter plot shows three clusters of states based on the number of **rape cases (x-axis)** and **domestic violence (DV) cases (y-axis)**.
 - **Cluster 0 (Blue):** High in both rape and DV cases.
 - **Cluster 1 (Grey):** Moderate levels of both crimes.
 - **Cluster 2 (Red):** Low levels of both crimes.
2. **High Variability:**
 - Some states in Cluster 0 show extreme levels of DV (up to 20,000 cases) with a wide spread in rape cases (up to 6,000).

Deep Insights:

1. **High DV States:** Cluster 0 indicates that some states have alarmingly high domestic violence cases, suggesting a deep-rooted societal problem. These states likely need targeted interventions.
2. **States with Low Crime Reports (Cluster 2):** These states, while low in reported numbers, might face issues like underreporting, potentially masking the real extent of crimes like DV and rape.
3. **Moderate Crime States:** Cluster 1 reflects states with balanced levels of reported cases, likely signaling that they are either effectively reporting or fall between extremes.

18. Crime Distribution based on election and non election years

Crime Distribution: Election Years vs Non-Election Years



Key Observation:

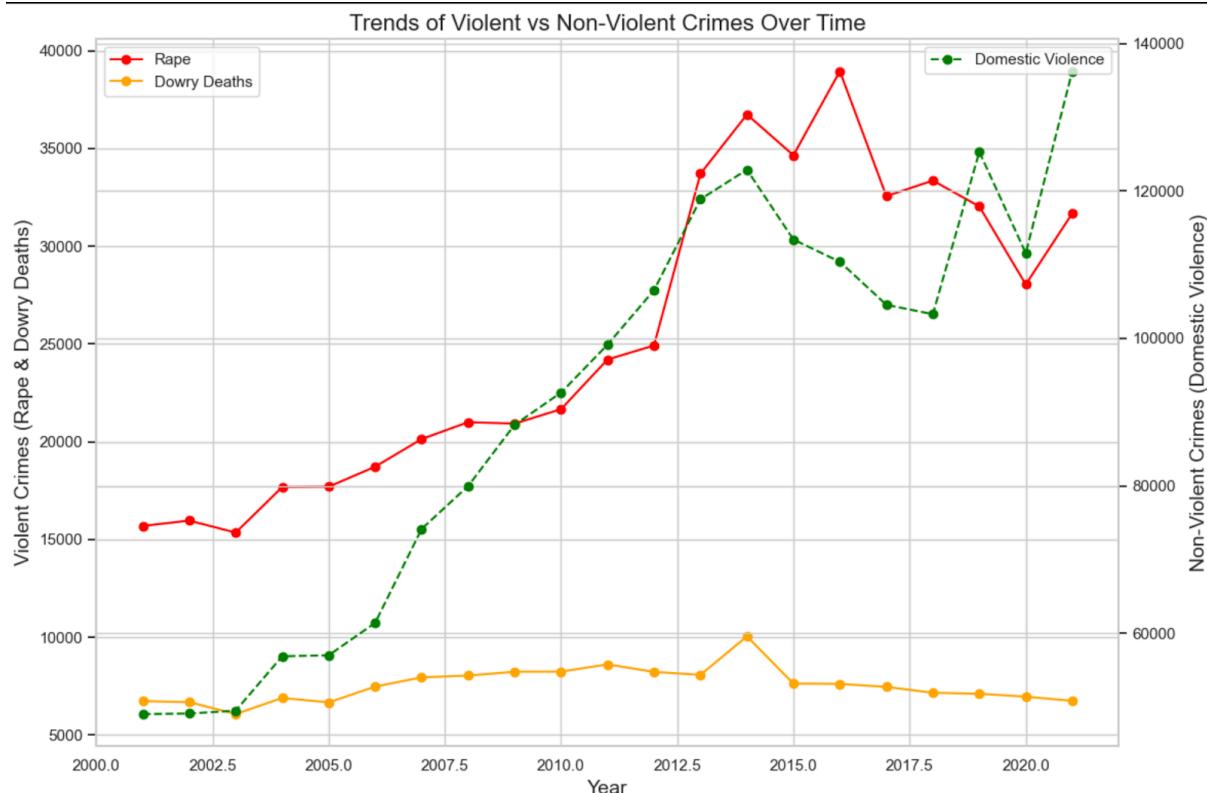
- The pie chart displays the **crime distribution between election years and non-election years**.
- **79.5% of the total crimes occur during non-election years**, while **20.5% occur in election years**.

Deep Insight:

- A significant majority of crimes, nearly 80%, take place in **non-election years**, which suggests that crime rates may not be strongly influenced by election cycles.
- The **smaller proportion of crimes (20.5%) in election years** may indicate certain factors like heightened government presence, increased law enforcement activity, or public scrutiny during election periods, potentially leading to a temporary reduction in criminal activities.

- However, **crimes still occur during election years**, indicating that even with potential government focus on law and order, criminal activities continue, albeit at a reduced rate compared to non-election years.

19. Trend of violent and non violent crimes over time



Key Observations:

1. **Rape** (Red Line) and **Dowry Deaths** (Yellow Line) represent violent crimes, while **Domestic Violence** (Green Line) represents non-violent crime in this chart.
2. **Rape** shows a **steady increase from 2001 to 2013**, after which it fluctuates but remains at a high level.
3. **Dowry Deaths** are relatively **constant** over time, showing minor fluctuations, but no significant trend.
4. **Domestic Violence** shows a sharp **increase from 2005 onwards**, peaking around 2012-2013, then **declining sharply**, followed by inconsistent fluctuations from 2015 to 2020.

Deep Insights:

1. **Rising Trend in Violent Crimes (Rape):**
 - The steady rise in rape cases from 2001 to 2013 suggests either an increase in reporting due to improved social awareness, legislative changes, or possibly an actual increase in the occurrence of this crime.

- The fluctuations post-2013 could indicate changing societal or legal dynamics, such as policy reforms or campaigns addressing the issue.

2. Stable Trend in Dowry Deaths:

- The dowry deaths trend, with little change over the years, reflects a **persistent societal issue**. Despite numerous laws and social movements, this crime remains unresolved, pointing toward a possible lack of effective implementation of preventive measures or deep-rooted cultural practices.

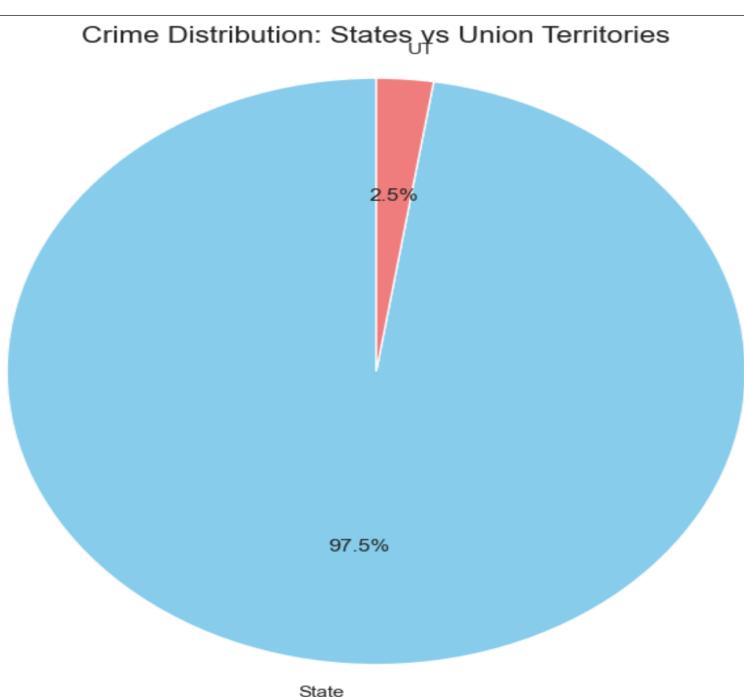
3. Volatile Nature of Domestic Violence:

- The significant rise in domestic violence cases post-2005, with a peak around 2012-2013, followed by a rapid decline, may reflect several factors:
 - Initial **increased awareness and reporting** due to campaigns or laws could explain the sharp rise.
 - The subsequent **decline** may indicate improvements in social interventions, legal reforms, or reporting mechanisms.
 - The sharp spikes in recent years (especially around 2020) could be tied to **external events** like economic crises or pandemics, which often exacerbate domestic violence incidents.

4. Overall Crime Landscape:

- Violent crimes such as **rape** are consistently high, indicating persistent challenges in ensuring women's safety.
- Non-violent crimes like **domestic violence** reflect how internal family dynamics or external stressors affect crime rates over time, making this area volatile and sensitive to socio-economic changes.

20. Crime distribution: State vs UTs



Key Observation:

- The pie chart shows the **crime distribution between Indian states and Union Territories**.
- **97.5% of crimes occur in the states**, while only **2.5% occur in Union Territories (UTs)**.

Deep Insight:

- The overwhelming majority of crimes (97.5%) are concentrated in states, which can be attributed to the **larger population, area, and socio-economic diversity** of states compared to Union Territories.
- Union Territories, despite being smaller regions in terms of geography and population, still account for **2.5% of total crimes**, which could be significant considering their size. This might suggest **administrative challenges, governance issues, or local crime hotspots** in some Union Territories.
- This distribution could also reflect that **crime-prevention efforts may need to be more region-specific**, as the dynamics in states differ vastly from those in Union Territories.

Datasets Sources

1.[https://www.kaggle.com/datasets/greeshmagirish/crime-against-women-2001-2014-india\(2001 to 2021\)](https://www.kaggle.com/datasets/greeshmagirish/crime-against-women-2001-2014-india)

2. Other Datasets Resource: It has been curated and researched by our team.

<https://docs.google.com/document/d/1D2kjS3wlx4TVXUcNvjh16JSyB20CYxdXhUJk7qLk8o/edit?usp=sharing>

3. Github: <https://github.com/Suraj-creation/CrimesOnWomen>