

# INDIAN CRIME ANALYSIS

Work by:  
A S Naveena  
GitHub : [naveena-as](#)

# OBJECTIVE

- The rate of crimes has obviously been increasing in the recent times. But what was the situation of the crime rate just before this decade? To analyse the same, we have the Indian Crime Data over the years 2001-2012 for all the states of India.
- Our objective is to analyse the trends in the count, gender based and age wise distribution of people involved in crimes which were committed over a span of 12 years (2001-2012), in the country of India. This can help us analyse the crime status of each of the states / UT of India and hence identify which are the safest / most unsafe of them.

# ABOUT THE DATASET

- Source : <https://www.kaggle.com/shishir349/indian-crime-analysis>
- File name : crimes.csv
- Contains details about the crimes committed in India (state-wise) over the years 2001-2012
- Details included : state, year, purpose of crime, differentiation based on age group, gender
- At the end of each year, for every state, there is a row with the cumulative sum of number of people involved in all the crimes committed in that year

# A GLIMPSE OF THE DATASET

	STATE/UT	YEAR	Purpose	Total No. of cases reported	Male upto 10 years	Female upto 10 years	Male 10-15 years	Female 10-15 years	Male 15-18 years	Female 15-18 years	Male 18-30 years	Female 18-30 years	Male 30-50 years	Female 30-50 years	Male above 50 years	Female above 50 years	Total Male	Total Female	Grand Total
0	Andhra Pradesh	2001	For Adoption	8	3	1	0	0	0	0	0	4	0	0	0	0	3	5	8
1	Andhra Pradesh	2001	For Begging	2	0	0	2	0	0	0	0	0	0	0	0	0	2	0	2
2	Andhra Pradesh	2001	For Camel racing	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	Andhra Pradesh	2001	For Illicit intercourse	78	0	2	0	25	0	24	0	25	0	2	0	0	0	78	78
4	Andhra Pradesh	2001	For marriage	339	0	0	0	73	1	164	6	91	0	4	0	0	7	332	339
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
5861	Puducherry	2012	For Selling body parts	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5862	Puducherry	2012	For Slavery	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5863	Puducherry	2012	For unlawful activity	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5864	Puducherry	2012	Others	8	2	4	1	1	0	0	0	0	0	0	0	0	3	5	8
5865	Puducherry	2012	Total	19	2	4	1	1	0	10	0	1	0	0	0	0	3	16	19

5866 rows x 19 columns

# STRUCTURE OF THE DATASET

- 5866 rows and 19 columns

```
df_crimes.shape
```

```
(5866, 19)
```

```
df_crimes.columns
```

```
Index(['STATE/UT', 'YEAR', 'Purpose', 'Total No. of cases reported',  
      'Male upto 10 years', 'Female upto 10 years', 'Male 10-15 years',  
      'Female 10-15 years', 'Male 15-18 years', 'Female 15-18 years',  
      'Male 18-30 years', 'Female 18-30 years', 'Male 30-50 years',  
      'Female 30-50 years', 'Male above 50 years', 'Female above 50 years',  
      'Total Male', 'Total Female', 'Grand Total'],  
      dtype='object')
```

# ABOUT THE VISUALISATION

## Packages used:

- pandas, numpy, seaborn, plotly, matplotlib, wordcloud

## Techniques used:

- Bar graphs and its variants

- Line graphs and stacked line plots

- Pie, donut, sunburst charts

- Scatter plots and matrices

- Boxplots and Violinplots

- Treemaps, Parallel plots, Word Cloud

# ANALYSIS



For Prostitution  
For marriage  
For unlawful activity  
For Revenge  
Others  
For Illicit intercourse

For Sale  
For Ransom  
For Stalking  
For Adultery

# ANALYSIS

- 1) An overall visualisation of the total crimes committed in all the states/UT included in the dataset

(Bar graph) - Uttar Pradesh leads, followed by Bihar. UT's have the least number of cases, followed by states like Mizoram and Sikkim.

- 2) Which gender is most involved in crimes?

(Line chart, grouped bar graph) – females were more involved in crimes compared to men. Highest contribution is in Uttar Pradesh, followed by Rajasthan and Assam.

- 3) Visualise the crimes in top 5 safest states to live in

(Bar graph) – Lakshadweep, Daman & Diu, Sikkim, A&N Islands, Mizoram, D&N Haveli, Puducherry, Goa, Nagaland, Chandigarh. Safest UT is Lakshadweep, state is Sikkim.

- 4) Which age (& gender) group committed the most crimes in Kerala in 2010?

(Pie chart) – females in the age of 18-30 (30%), followed by those of age 15-18 (28.1%)



# ANALYSIS (CONTD.)

5) Which crime was committed the most in 2001?

(Horizontal bar graph) – crimes for marriage was committed the most, followed by a collective group of crimes for other purposes. This was followed by crimes related to illicit intercourse.

6) What is the trend of total crimes committed in Gujarat over the years?

(Tree map) – the number of crimes has increased over time. But throughout, the crimes related to marriage contributed the most, followed by the collective group of other crimes and illicit intercourse switching between the second and third highest.

7) Which year recorded the most number of crimes?

(Line plot with points) – has been increasing constantly since 2005. Hence, 2012. (95.184k)

8) Analyse the trend in crimes related to marriage, illicit intercourse, ransom in the decade (2001-2010)

(Stacked bar chart) – throughout the decade, crimes related to marriages was committed the most, followed by those for illicit intercourse. For the first two purposes, UP tops the list and for the third, Bihar.

# ANALYSIS (CONTD.)

9) Range of number of females involved in marriage related crimes in the year 2009?

(Boxplot) – most states have a total in the range (0,650). There are only 2 states with count between 1500 and 2000, and only 1 with greater than 3500.

10) Analyse the state-wise male v/s female trend in crimes in South India (KL, TN, KA, AP)

(Scatterplot) – Kerala has the least cases, AP the most (count of males and females involved has increased). In TN, there was a rapid increase in number of females involved, but not males.

11) Visualise the range of males and females involved in crimes in South India (KL, TN, KA, AP)

(Scatterplot + Boxplot) – Males in range (69,731) with majority in the range (133,461). Females in range (88,1773) with majority in range (262,1139). On average, males = 260 and females = 697

12) Variation across age of females involved in crimes in North East India?

(Parallel coordinates) – in females upto 18, almost all states have a significant number of cases. Above 18, Tripura, Sikkim, Nagaland have very less number of cases. Major contributors are Arunachal, Assam, Manipur.

# ANALYSIS (CONTD.)

13) Trends in number of minor females (<18 years, age group wise) involved in crimes in the North Eastern states of India

(Scatter matrix) – no linear relationship. By analysing the scales we can say that age 15-18 is most involved (max. 451). Tripura and Assam stand out from the rest with the highest cases in all groups.

14) Analyse the trends in purpose for crimes in 2011 and 2012?

(Sunburst plot) – 2012 recorded a higher number than 2011 with marriage, others, illicit intercourse being the purpose in that order.

15) Which crime tops the chart?

(Donut chart) – On overall marriage crimes are the highest (45.9%) followed by the group of crimes for other purposes (36.4%) and illicit intercourse (8.32%).

16) Analyse the range of the total number of cases per year.

(Boxplot) – at least 39,384 and at most 95,184 in a year. Mostly between 45,314 and 72,300, with a median of 51,552.

# ANALYSIS (CONTD.)

17) Which crimes occur the most? Visualise with the words.

(Word cloud) – The size of the words show the frequency of occurrence. As seen previously, marriage crimes are the highest, followed by other purposes and illicit intercourse. Crimes for begging is the least significant by vision.

18) Analyse the range of number of males belonging to different age groups involved in crimes

(Boxplots) – age group 18-30 is the most spread, and >50 is the least spread.

19) Total male and female count ratio?

(Stacked line plot) – Females are visibly more involved than males since the area covered by them is more in the chart. Towards the end we can say that female count was about 4 times that of males.

20) Analyse the total crime rate difference in the beginning (2001) v/s the end (2012)

(Violin plot) – in 2001, the probability spread was limited to a small value ( $< 2000$ ), towards the bottom of the plot. But in 2012, this width started flattening and reaching up to 6000. The maximum has risen from ~4000 to ~8000. This implies an immense hike in the total number of crimes.



**THANK YOU !**