Capston project

TOPIC: Analysis of crimes in India

Dataset Used: Crimes in India (https://www.kaggle.com/datasets/rajanand/crime-in-india)

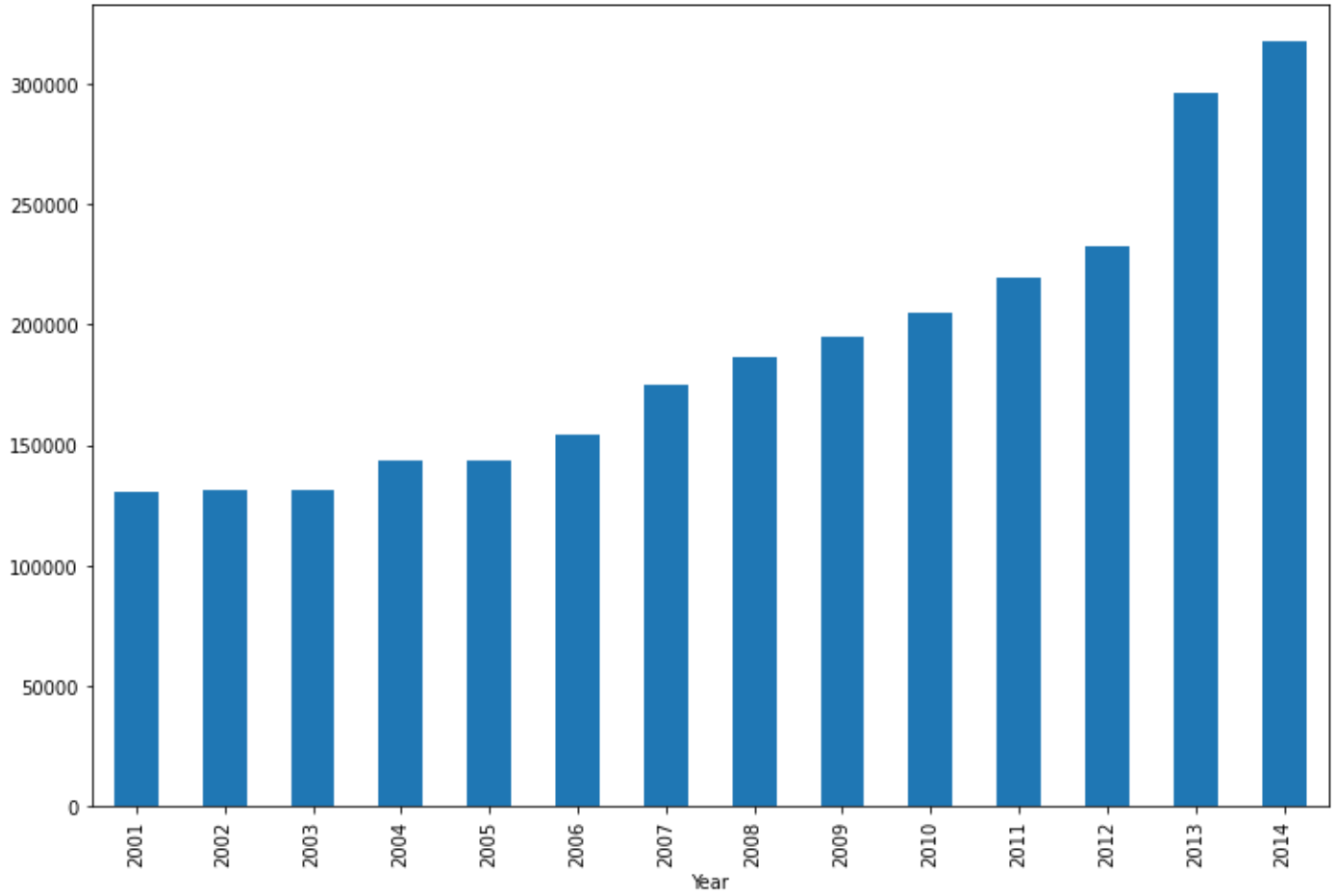
The Dataset chosen is extremely varied and hence extensive EDA has been performed as it is predominantly a data analysis oriented dataset. Predictive analysis has been performed on a selected dataset (IPC).

The EDA includes analysis of Crimes against women of SC and ST, Crimes against Children, General IPC dataset.

The various visual plots have detailed legend and substantial inference which can be derived from them. Important and non obvious inference with values have been mentioned below each plot.

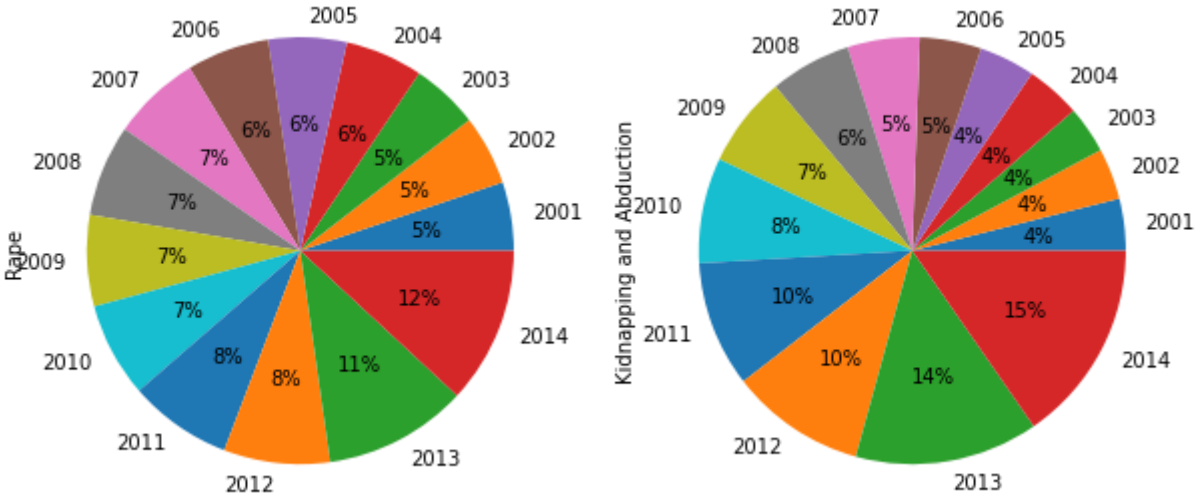
Exploratory data analysis (EDA)

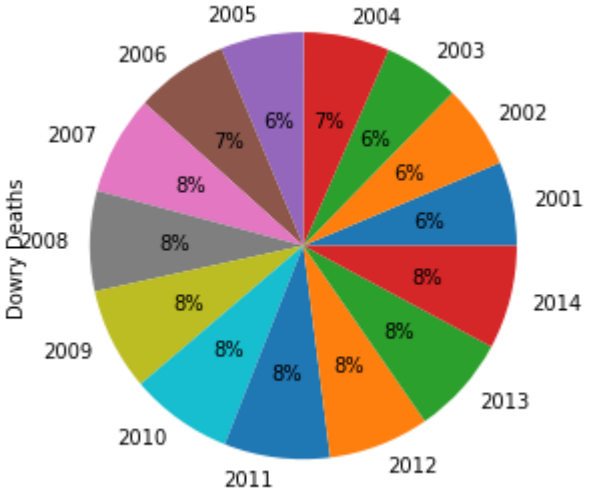
Yearly Crime Rate Analysis

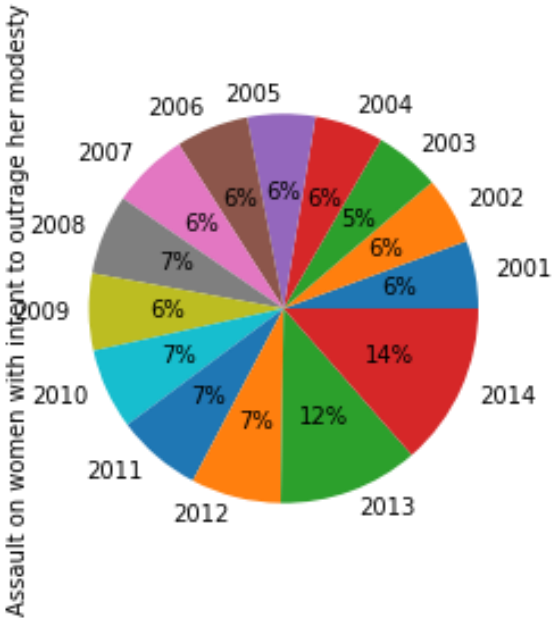


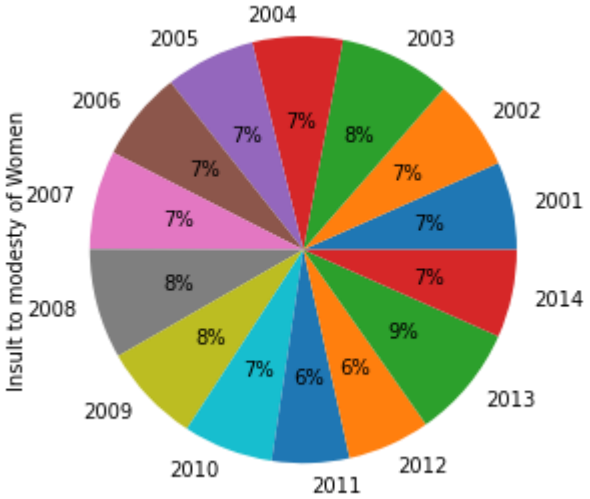
- We can see that the crime rate has been steadily increasing as the years pass by.

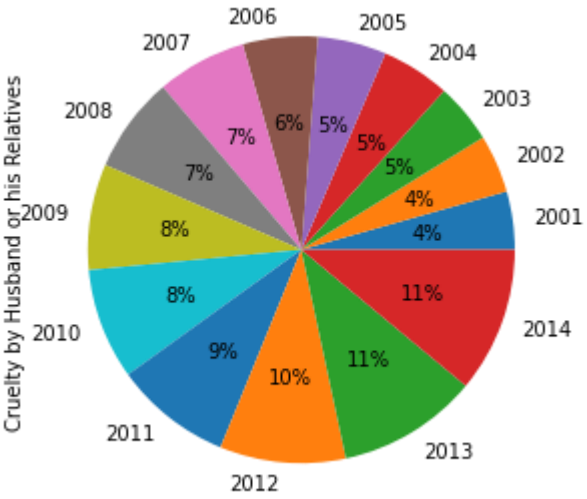
Percentage Distribution of each Crime based on Year

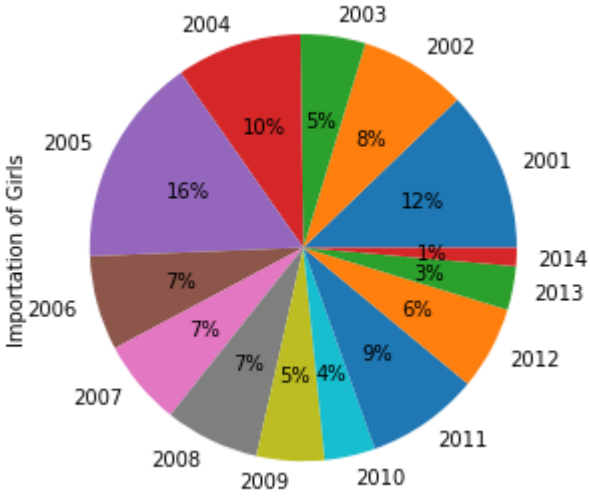


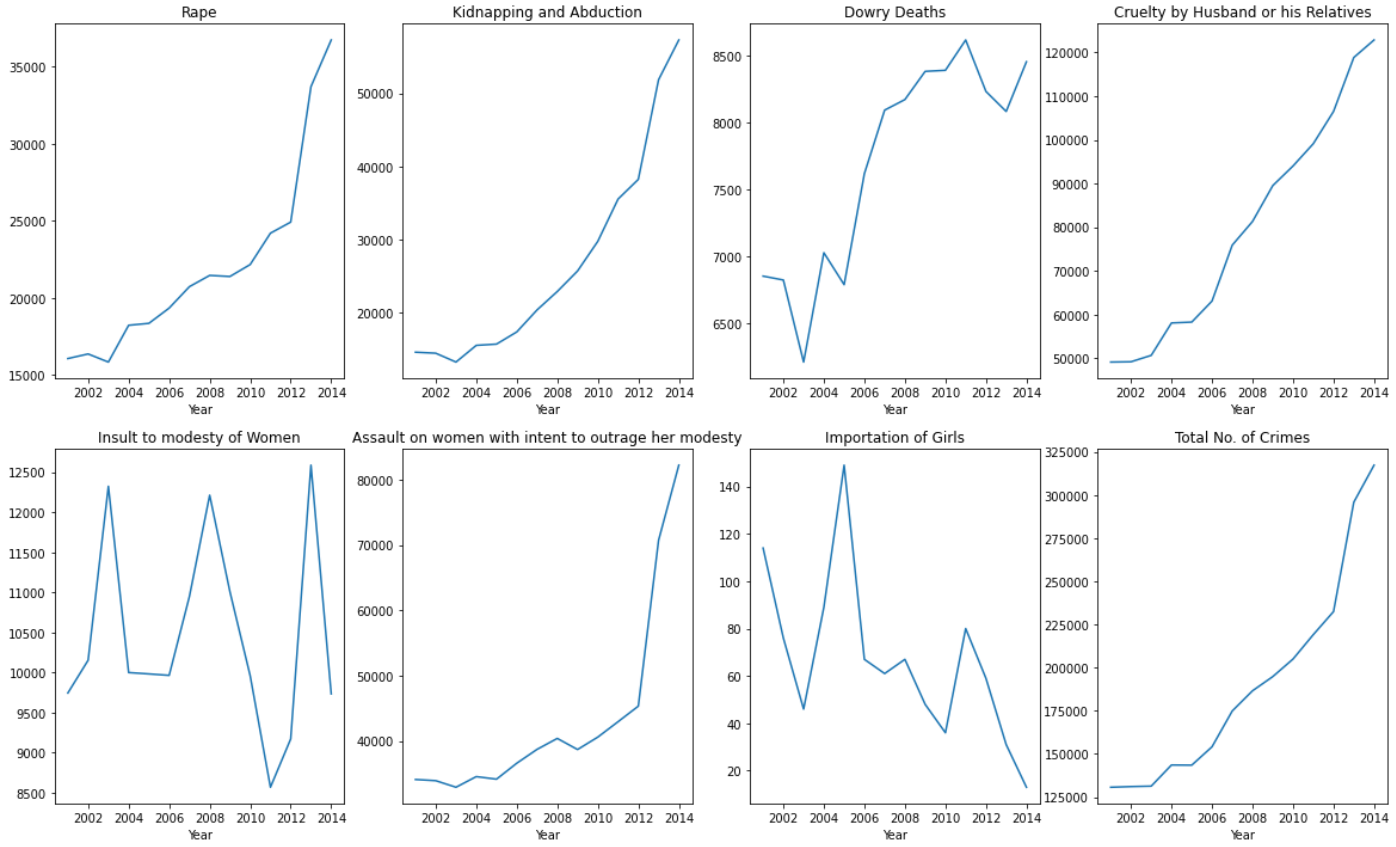

Dowry Deaths
2004 7%
2003 6%
2002 6%
2001 6%
2014 8%
2013 8%
2012 8%
2011 8%
2010 8%
2009 8%
2008 8%
2007 8%
2006 7%
2005 6%

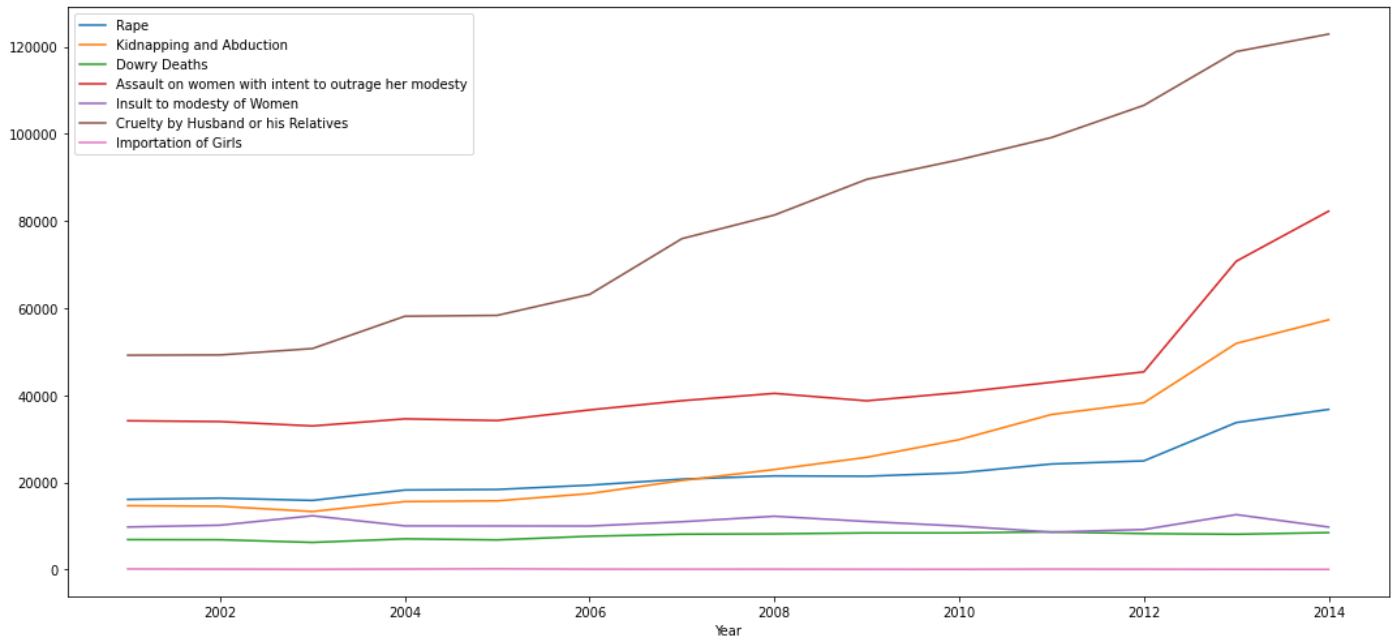

Assault on women with intent to outrage her modesty
2004 6%
2003 5%
2002 6%
2001 6%
2014 14%
2013 12%
2012 7%
2011 7%
2010 7%
2009 6%
2008 7%
2007 6%
2006 6%
2005 6%


Insult to modesty of Women
2004 7%
2003 8%
2002 7%
2001 7%
2014 7%
2013 9%
2012 6%
2011 6%
2010 7%
2009 8%
2008 8%
2007 7%
2006 7%
2005 7%


Cruelty by Husband or his Relatives
2004 5%
2003 5%
2002 4%
2001 4%
2014 11%
2013 11%
2012 10%
2011 9%
2010 8%
2009 8%
2008 7%
2007 7%
2006 6%
2005 5%

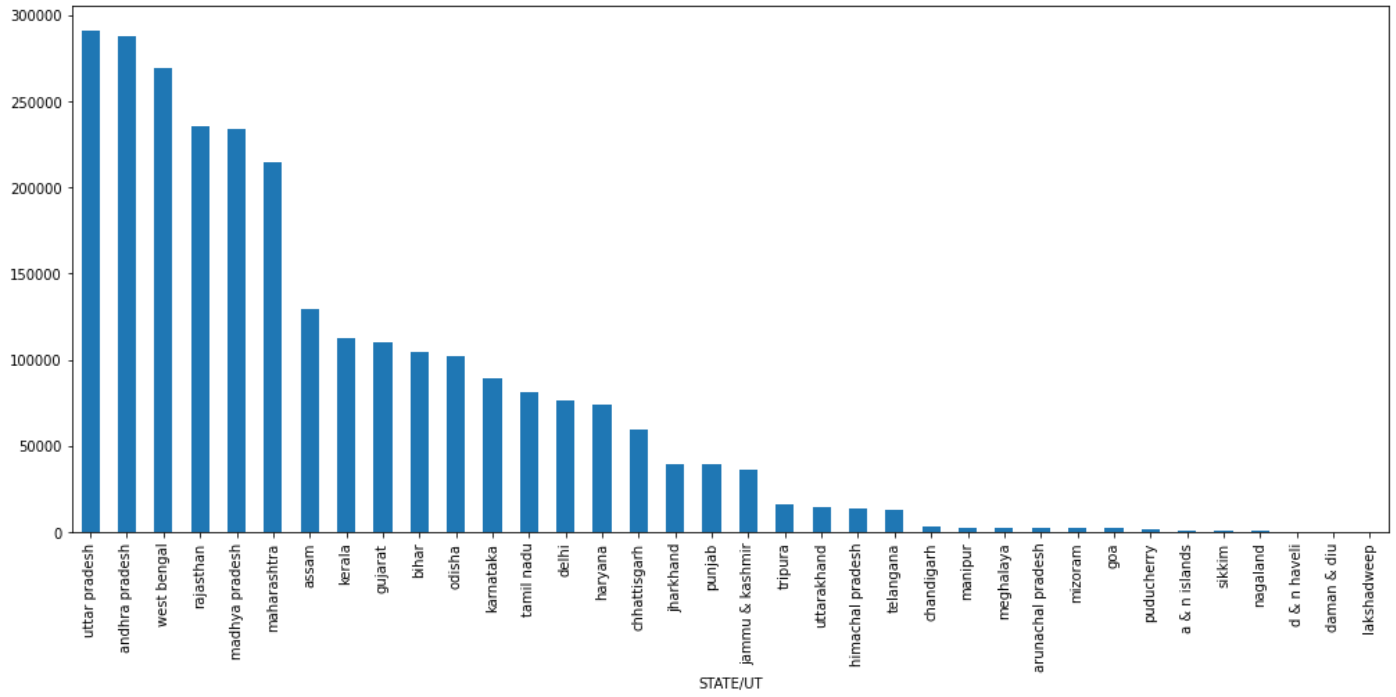

Importation of Girls
2004 10%
2003 5%
2002 8%
2001 12%
2014 1%
2013 3%
2012 6%
2011 9%
2010 4%
2009 5%
2008 7%
2007 7%
2006 7%
2005 16%





- It can be seen that the importation of girls is the only type of crime that has been reducing over the years.

State / UT wise Analysis of Crime



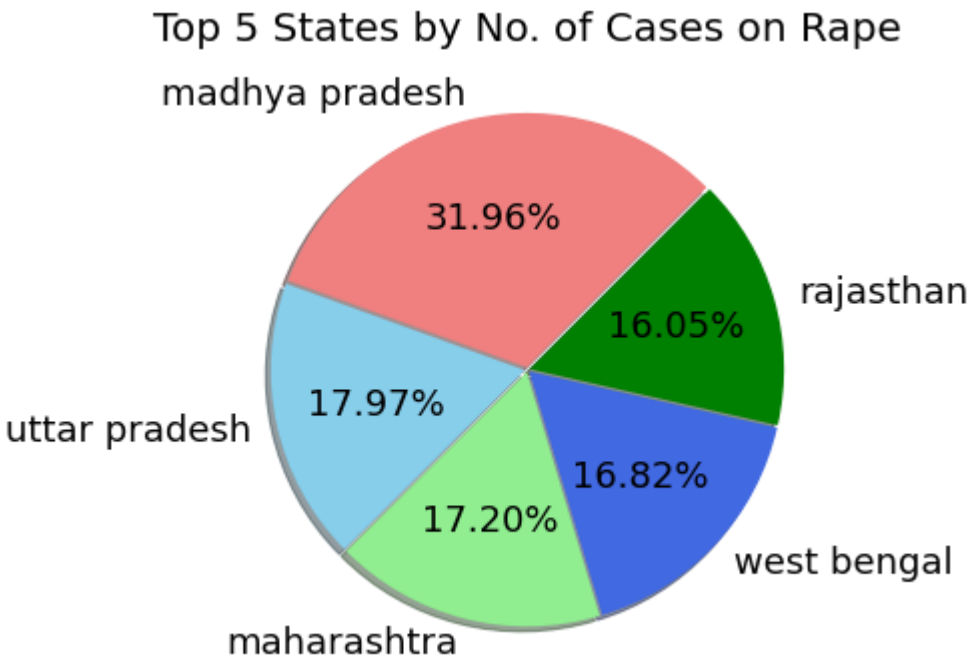
Top 3 States with Higher Number of Crimes

STATE/UT	Rape	Kidnapping and Abduction	Dowry Deaths	Assault on women with intent to outrage her modesty	Insult to modesty of Women	Cruelty by Husband or his Relatives	Importation of Girls	Total
uttar pradesh	25575	67953	28628	45606	26565	96869	3	291199
andhra pradesh	16075	17252	6922	63476	43482	140453	17	287677
west bengal	23938	30579	6154	33454	2674	172062	127	268988

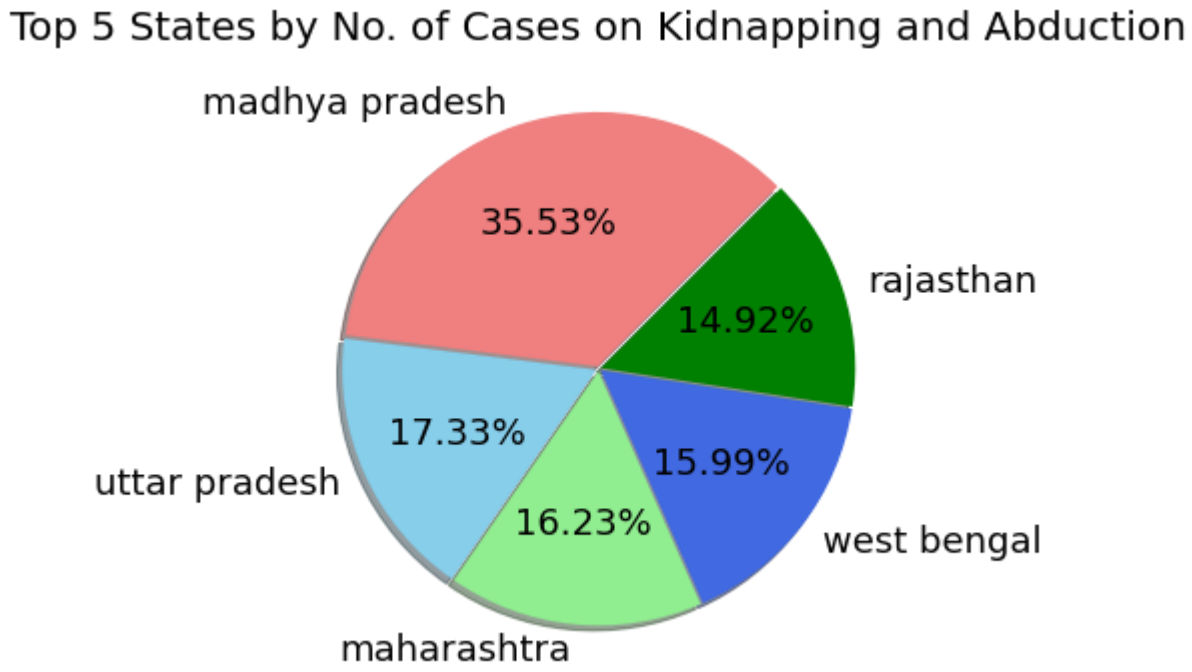
3 States with lower Number of Crimes

STATE/UT	Rape	Kidnapping and Abduction	Dowry Deaths	Assault on women with intent to outrage her modesty	Insult to modesty of Women	Cruelty by Husband or his Relatives	Importation of Girls	Total
d & n haveli	66	112	1	53	6	45	0	283
daman & diu	30	22	3	20	2	38	0	115
lakshadweep	10	1	0	7	2	7	0	27

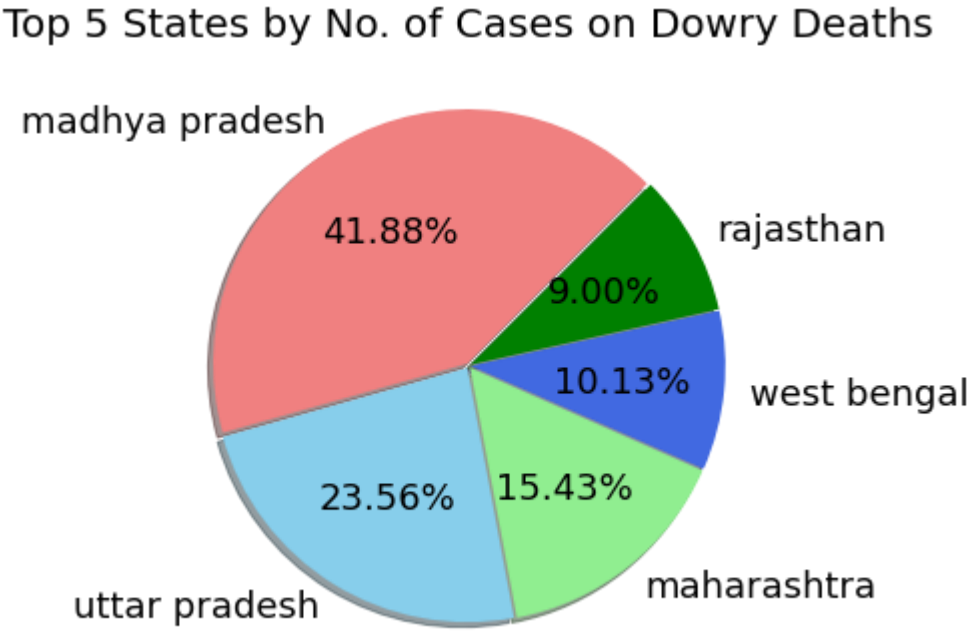
Top 5 States having higher Rape Cases



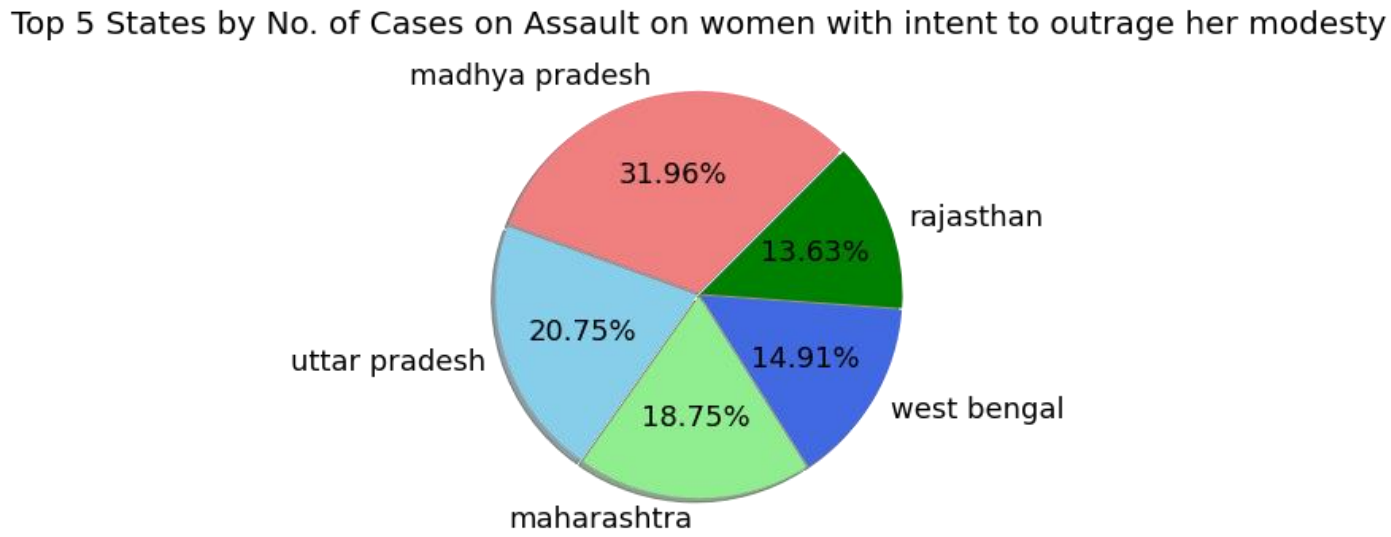
Top 5 States having higher Kidnapping and Abduction



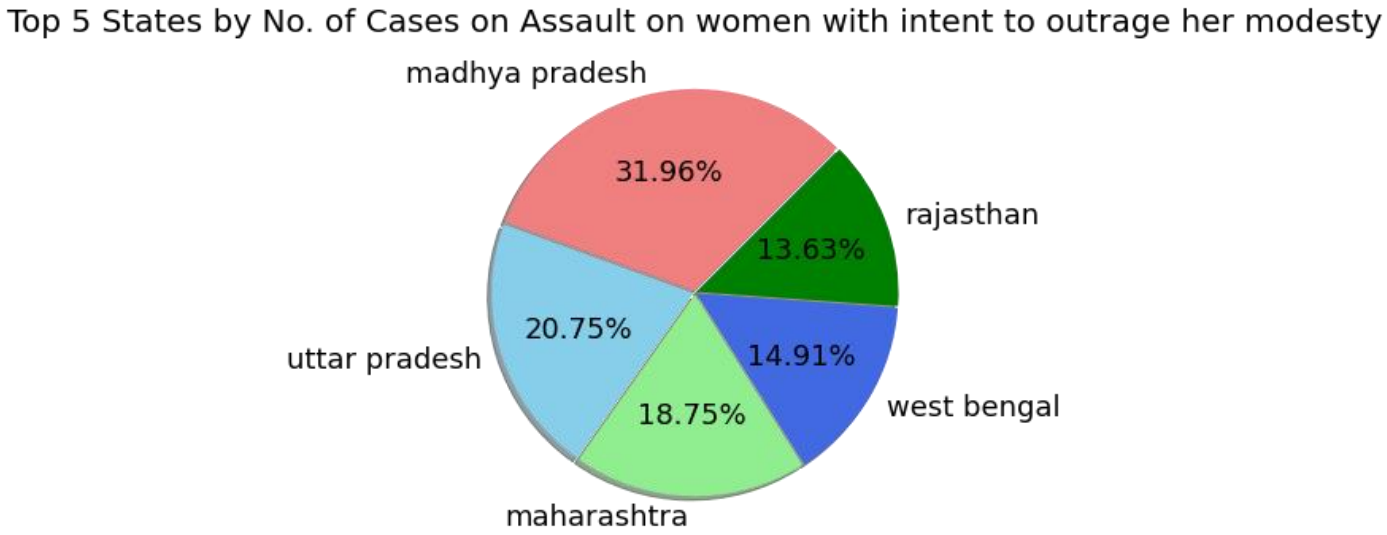
Top 5 States having higher Dowry Death's



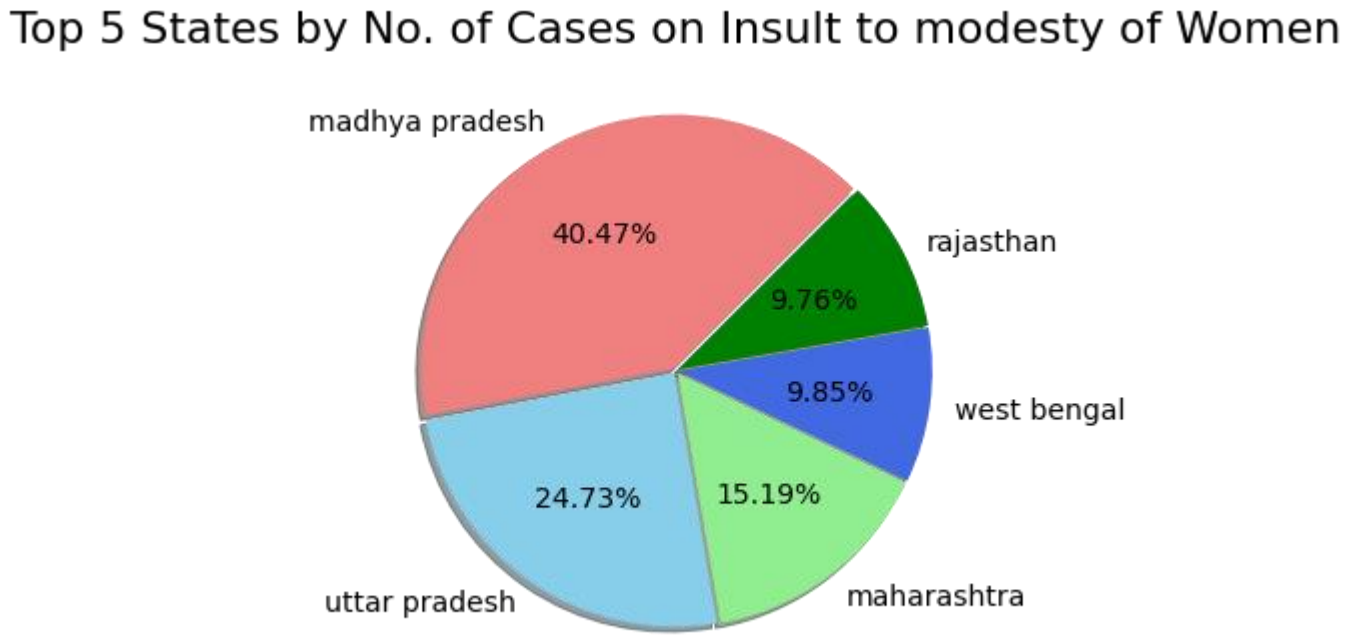
Top 5 States having higher Assault on women with intent to outrage her modesty



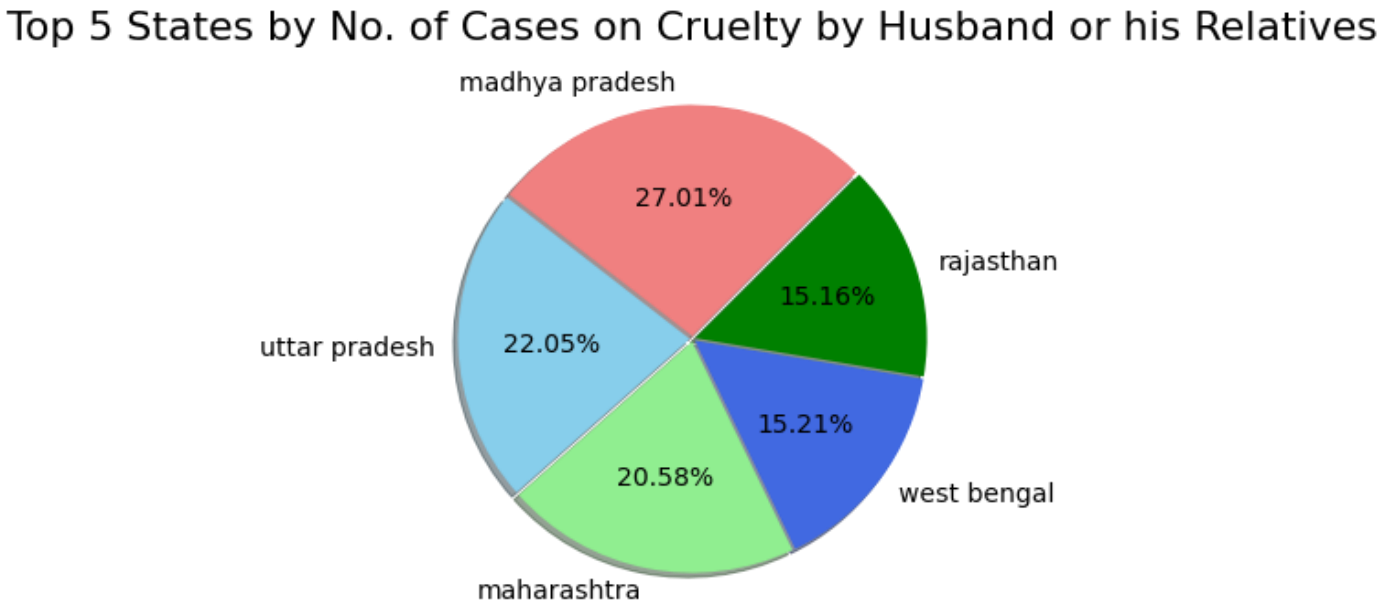
Top 5 States having higher Assault on women with intent to outrage her modesty



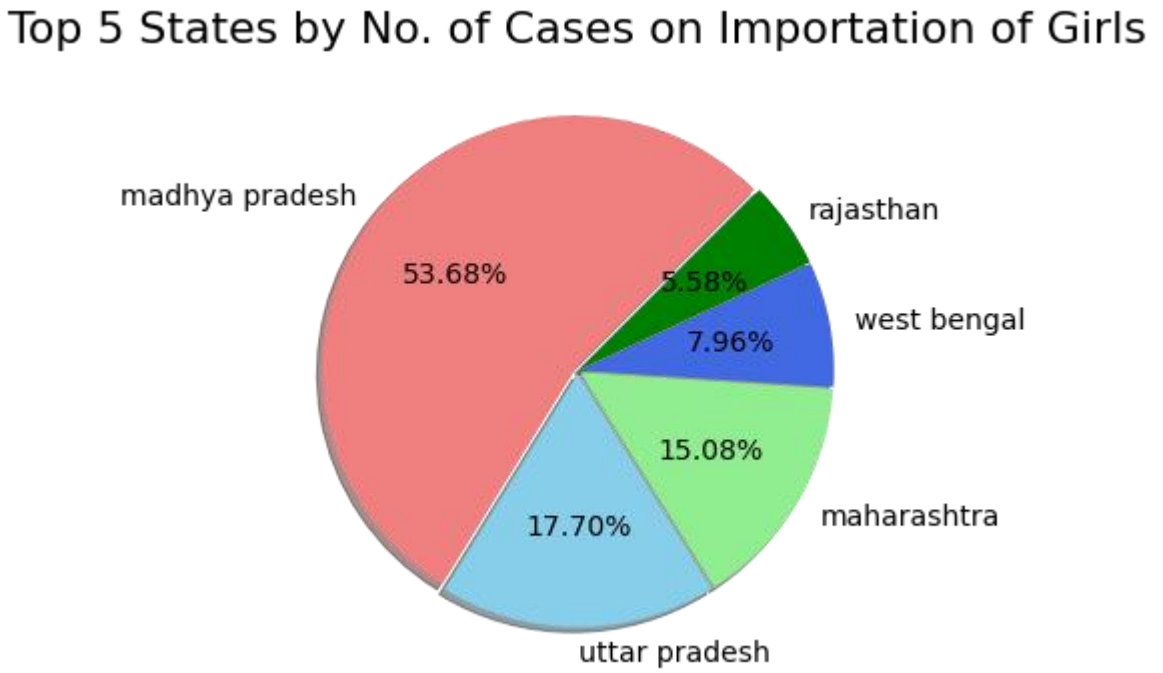
Top 5 States having higher Insult to modesty of Women



Top 5 States having higher cases on Cruelty by Husband or his Relatives



Top 5 States having higher cases on Importation of Girls



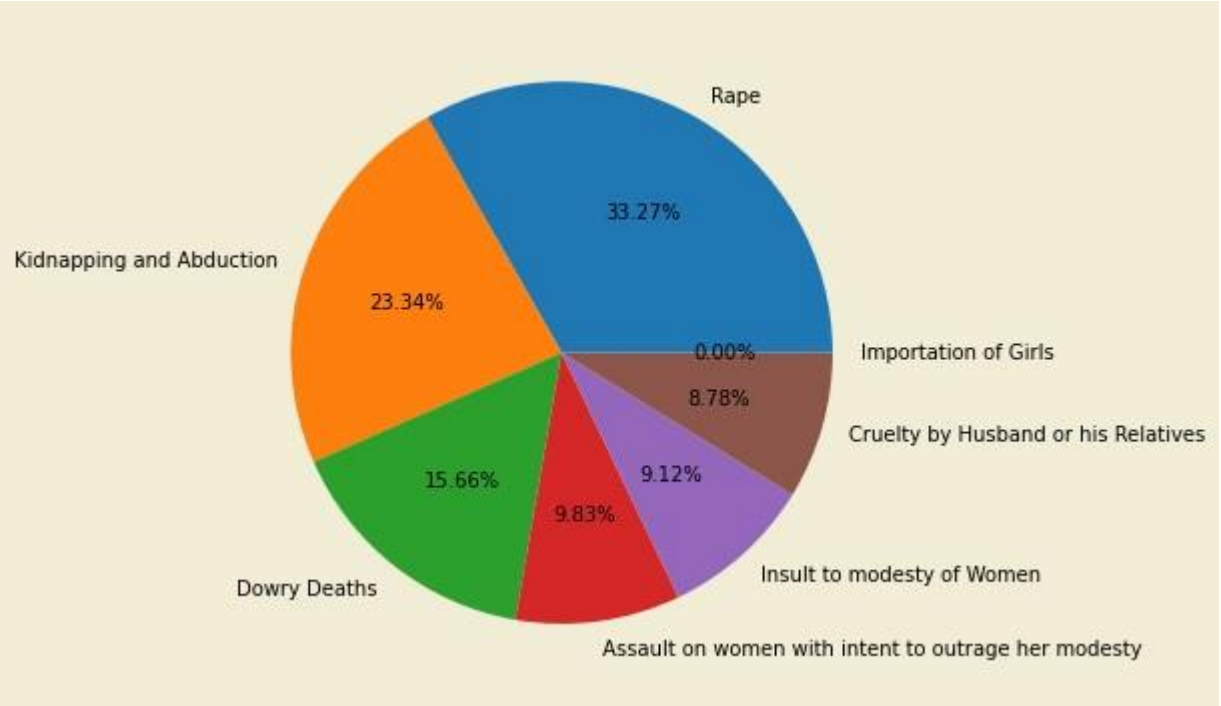
Analysis of Top 3 States

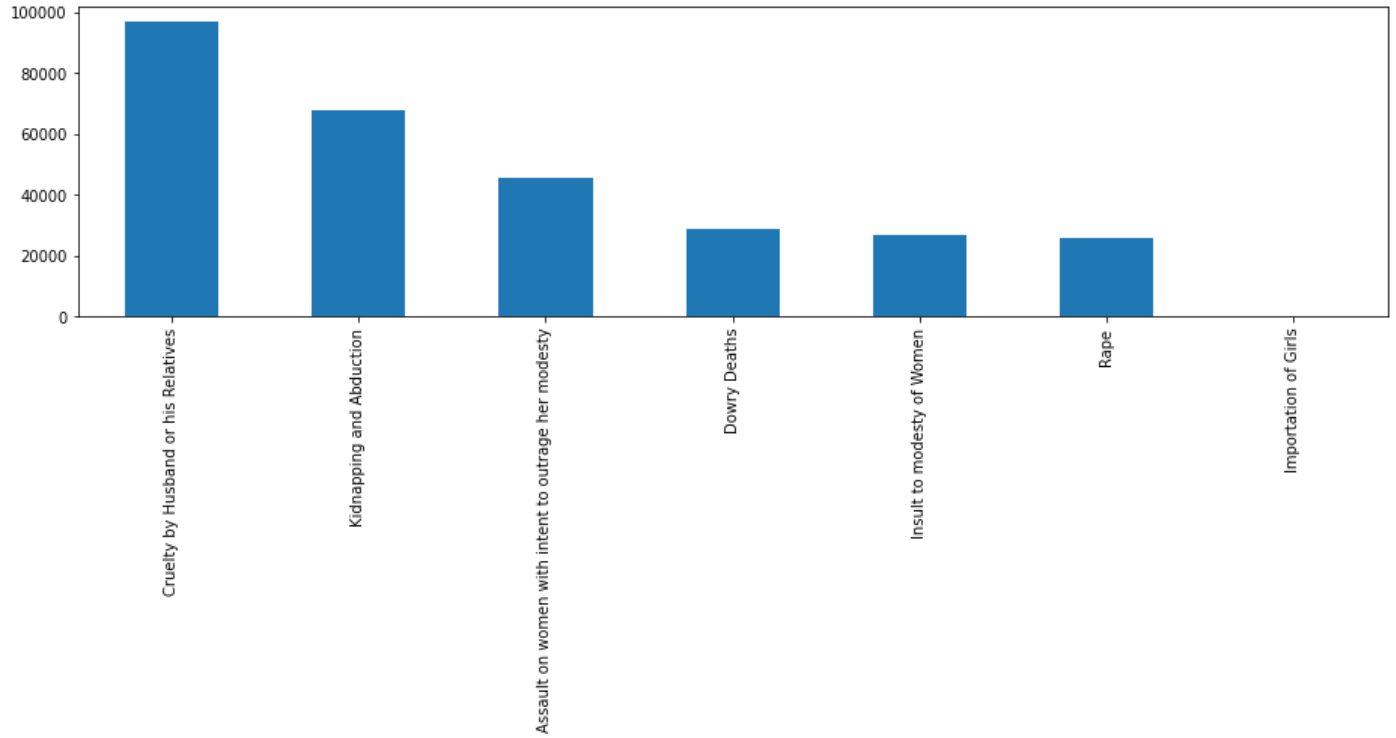
- Uttar Pradesh

Rape	25575
Kidnapping and Abduction	67953
Dowry Deaths	28628

Assault on women with intent to outrage her modesty	45606
Insult to modesty of Women	26565
Cruelty by Husband or his Relatives	96869
Importation of Girls	3

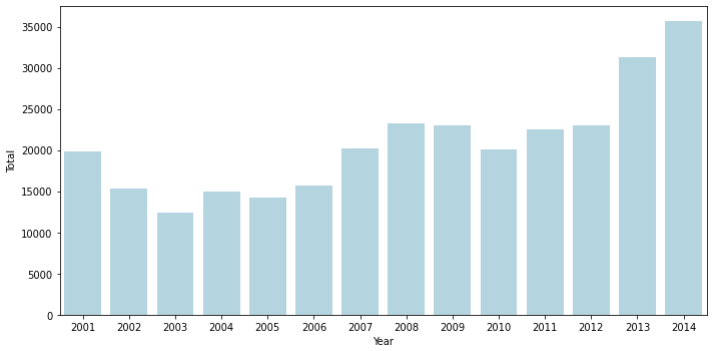
Percentage Distribution of Crimes in Uttar Pradesh from 2001 to 2014

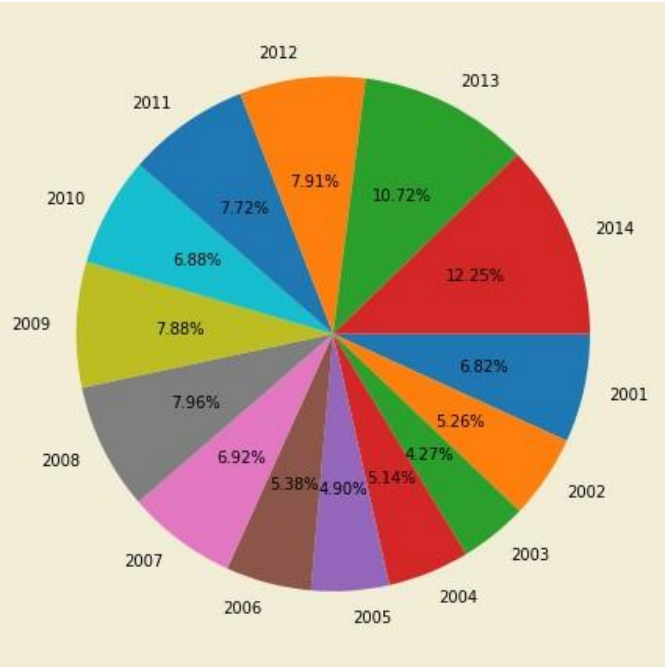




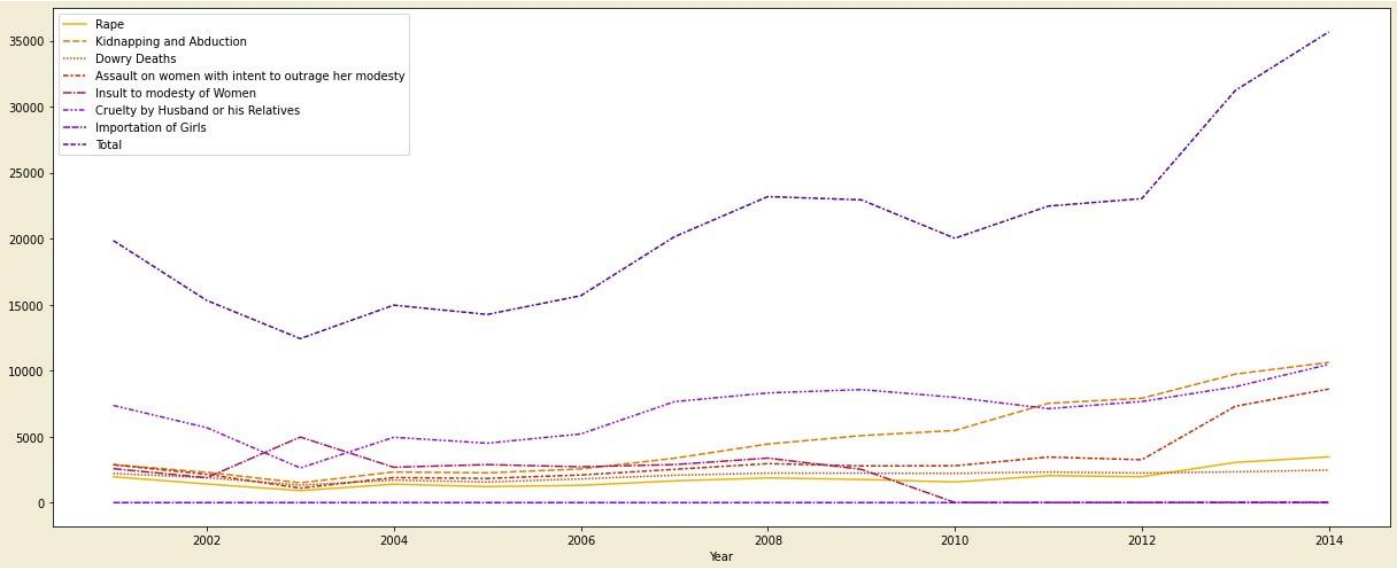
Year Wise Distribution of Crimes in Uttar Pradesh

Year	Rape	Kidnapping and Abduction	Dowry Deaths	Assault on women with intent to outrage her modesty	Insult to modesty of Women	Cruelty by Husband or his Relatives	Importation of Girls	Total
2001	1958	2879	2211	2870	2575	7365	0	19858
2002	1415	2298	1893	2145	1887	5679	0	15317
2003	911	1499	1322	1098	4970	2626	0	12426
2004	1397	2324	1708	1900	2682	4950	3	14964
2005	1217	2256	1564	1835	2881	4505	0	14258
2006	1314	2551	1798	2096	2714	5204	0	15677
2007	1648	3363	2076	2522	2882	7650	0	20141
2008	1871	4439	2237	2955	3374	8312	0	23188
2009	1759	5078	2232	2782	2524	8566	0	22941
2010	1563	5468	2217	2793	11	7978	0	20030
2011	2042	7525	2322	3455	3	7121	0	22468
2012	1963	7910	2244	3247	8	7661	0	23033
2013	3050	9737	2335	7303	25	8781	0	31231
2014	3467	10626	2469	8605	29	10471	0	35667





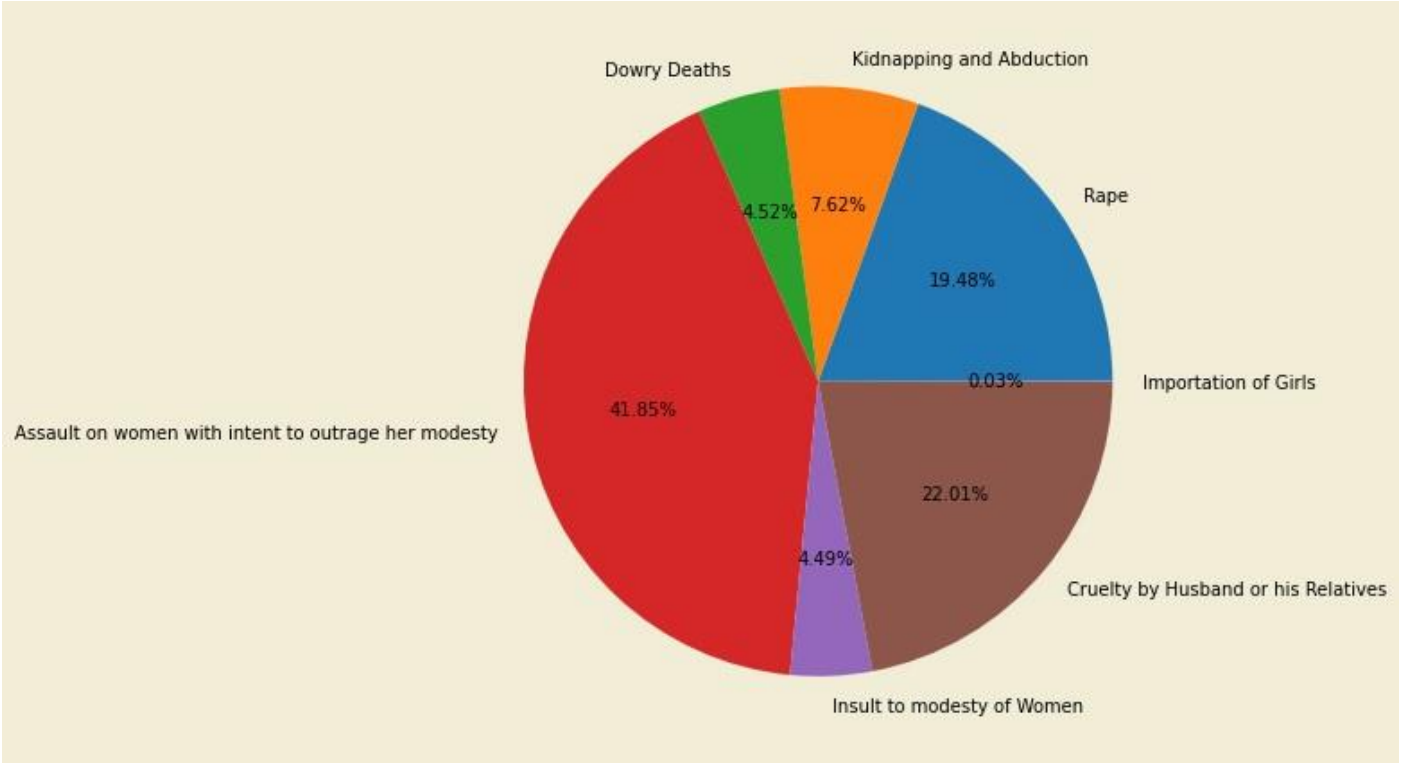
Timeline Analysis of Different Categories of Crime at Different Period of Time

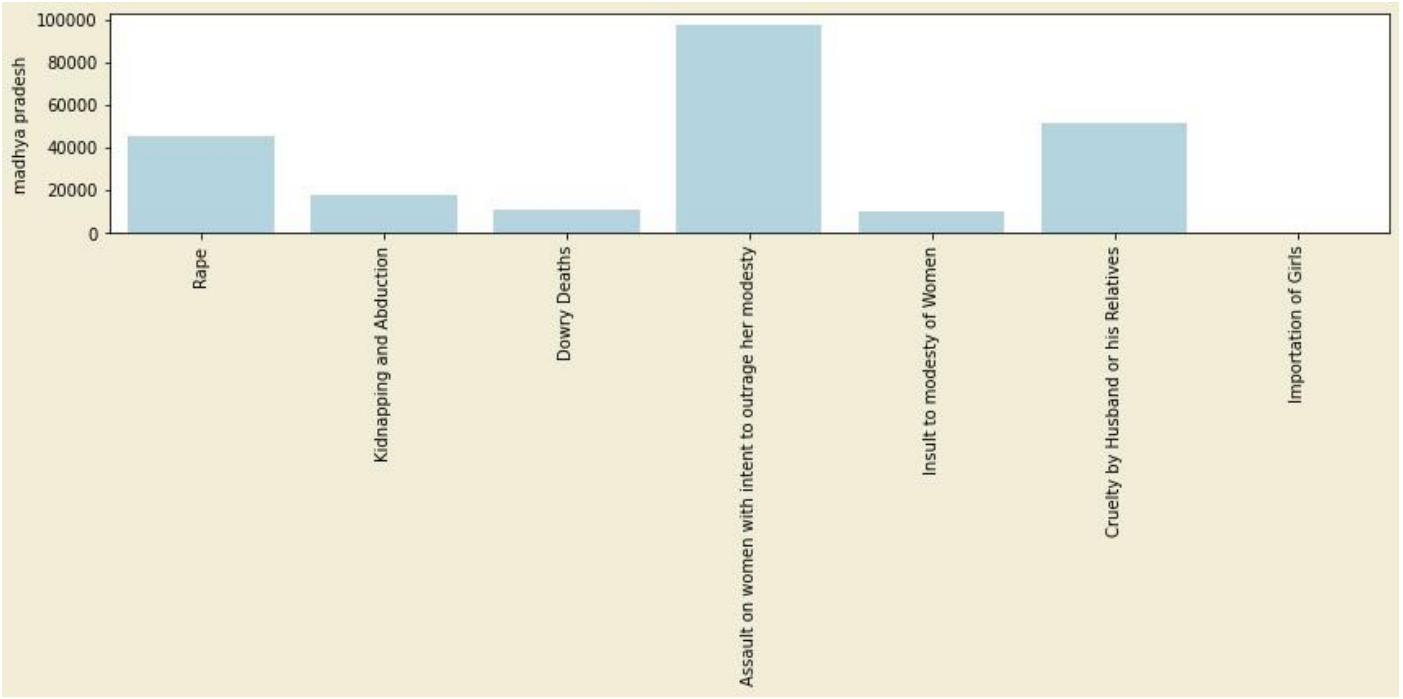


- **Madhya Pradesh**

Rape	45498
Kidnapping and Abduction	17804
Dowry Deaths	10545
Assault on women with intent to outrage her modesty	97739
Insult to modesty of Women	10482
Cruelty by Husband or his Relatives	51408
Importation of Girls	67

Percentage Distribution of Crimes in from 2001 to 2014

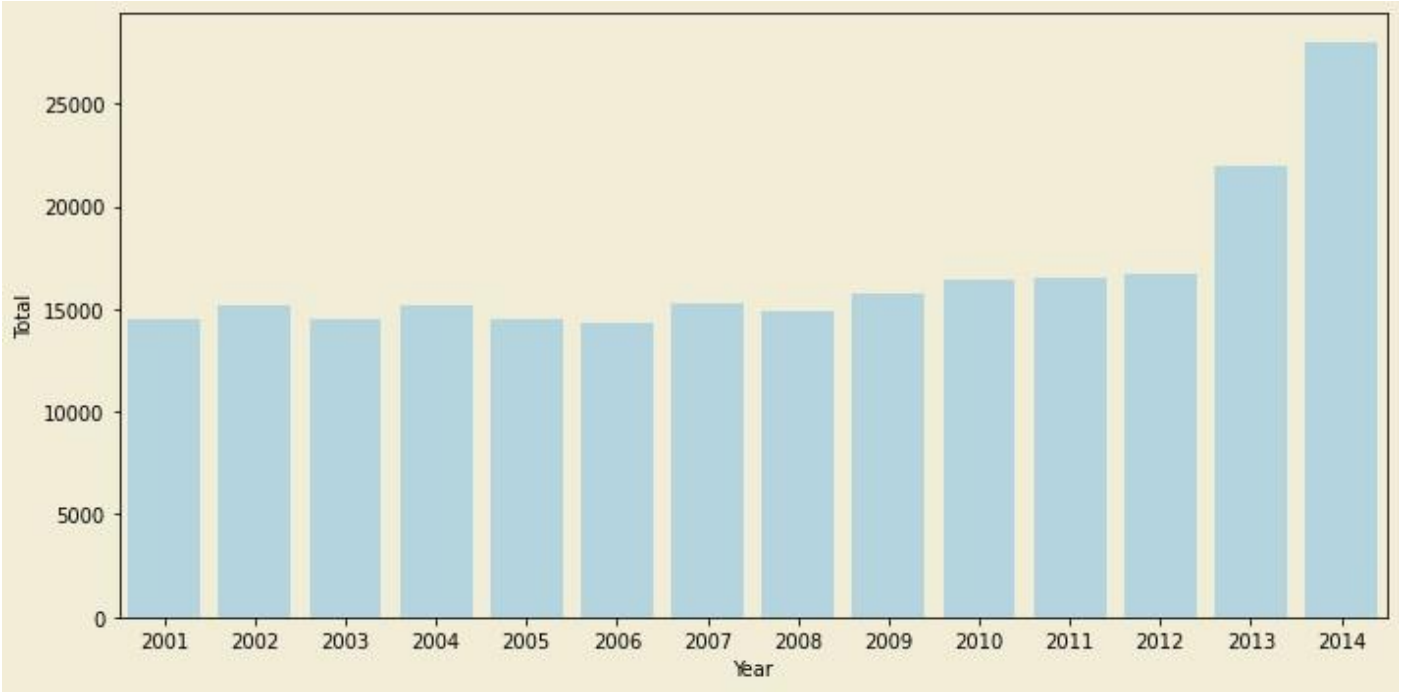


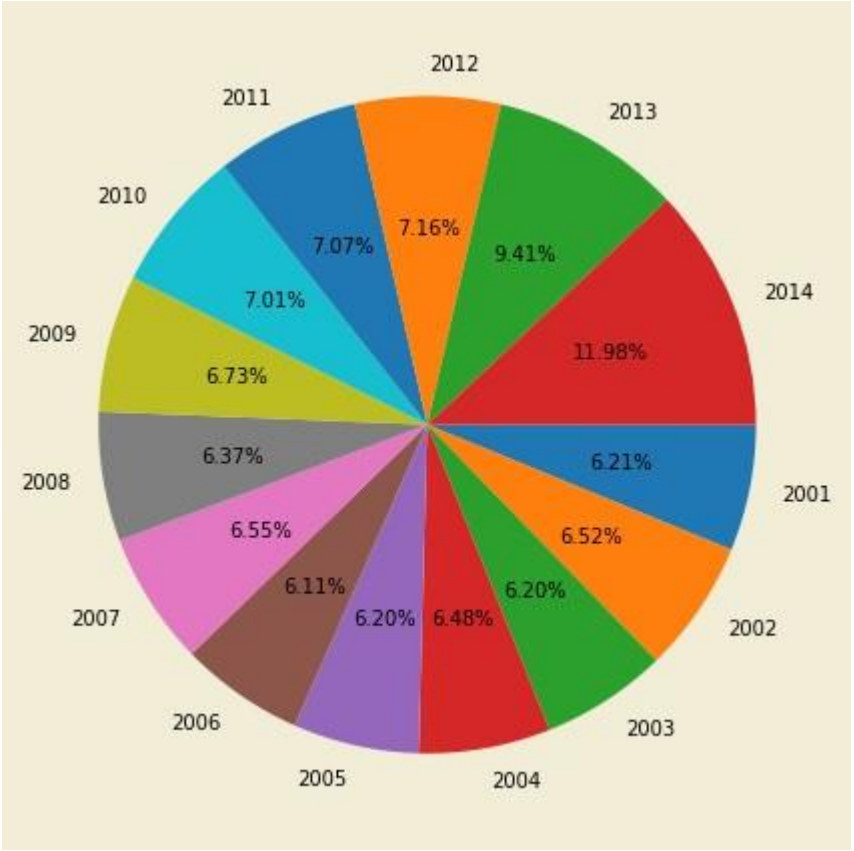


Year Wise Distribution of Crimes

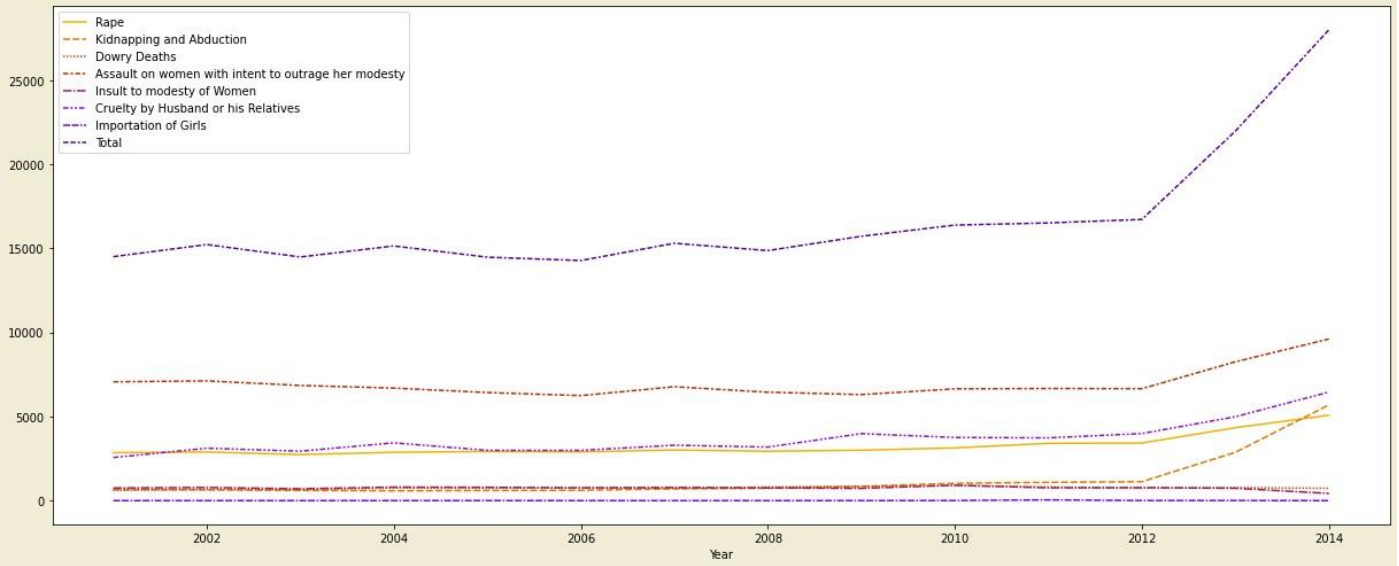
Year	Rape	Kidnapping and Abduction	Dowry Deaths	Assault on women with intent to outrage her modesty	Insult to modesty of Women	Cruelty by Husband or his Relatives	Importation of Girls	Total
2001	2851	668	609	7063	751	2562	0	14504
2002	2891	639	674	7118	783	3117	0	15222
2003	2738	608	648	6848	705	2938	0	14485
2004	2875	584	751	6690	804	3436	0	15140
2005	2921	604	739	6426	792	2989	3	14474
2006	2900	617	764	6243	762	2989	0	14275
2007	3010	701	742	6772	780	3294	0	15299
2008	2937	736	805	6445	758	3185	0	14866
2009	2998	841	858	6307	728	3983	1	15716
2010	3135	1030	892	6646	918	3756	5	16382

2011	3406	1088	811	6665	762	3732	45	16509
2012	3425	1127	743	6655	774	3988	6	16718
2013	4335	2873	776	8252	736	4988	7	21967
2014	5076	5688	733	9609	429	6451	0	27986

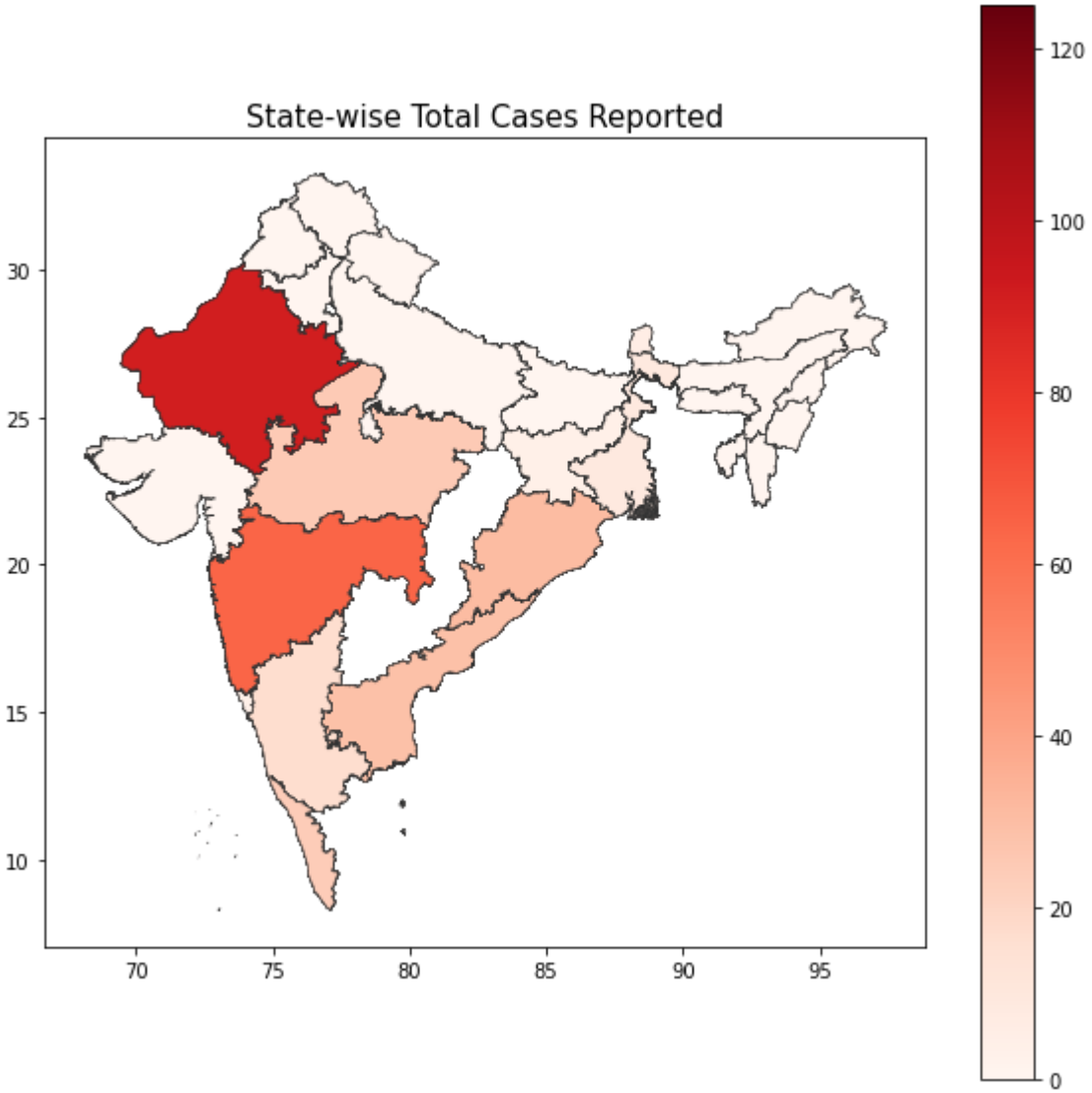

25000
20000
15000
10000
5000
0
Total
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
Year


2012
2011
2013
2010
7.07%
7.16%
9.41%
2014
7.01%
2009
11.98%
6.73%
6.37%
6.21%
2008
2001
6.55%
6.52%
6.11%
6.20%
6.20%
6.48%
2007
2002
2006
2003
2005
2004

Timeline Analysis of Different Categories of Crime at Different Period of Time

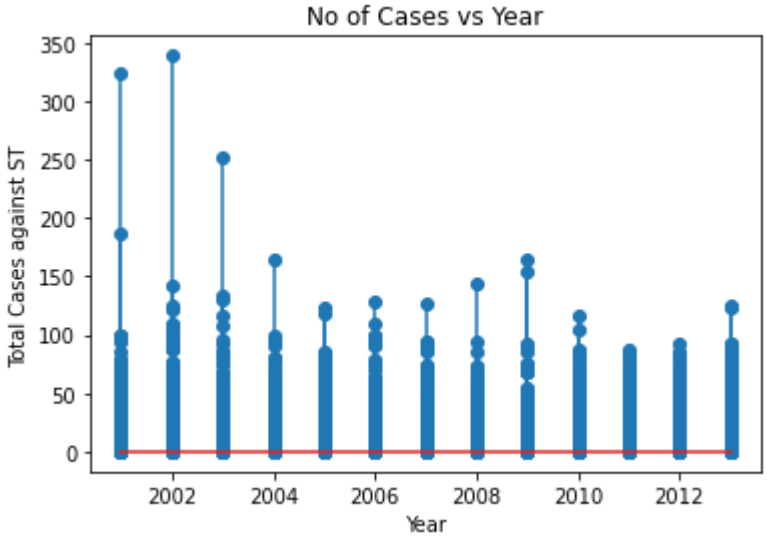


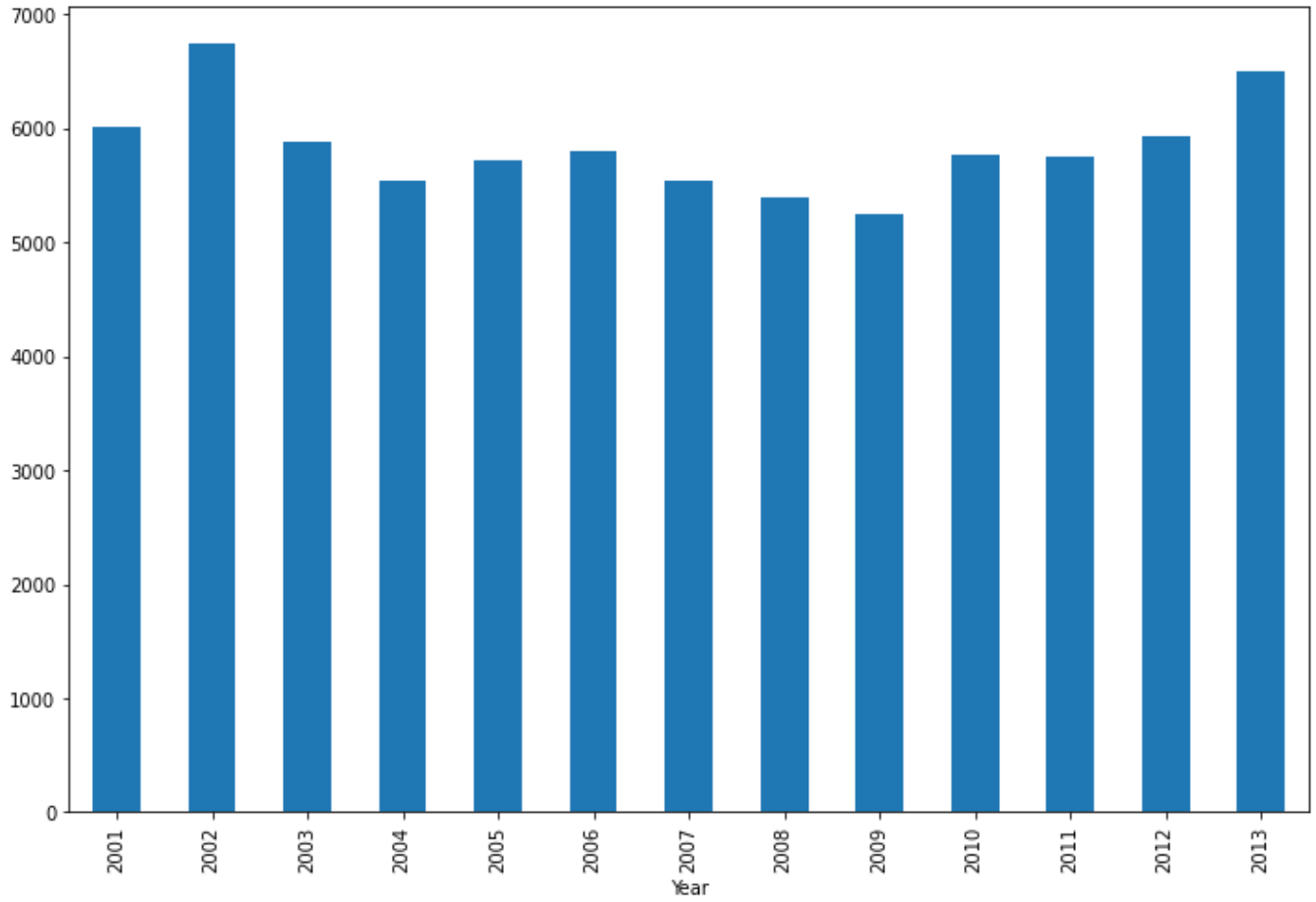
General Geomap Observation



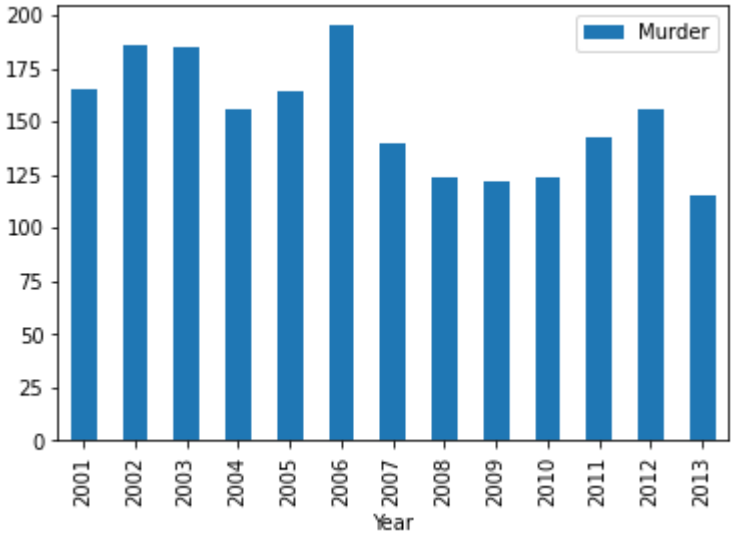
Yearly Analysis

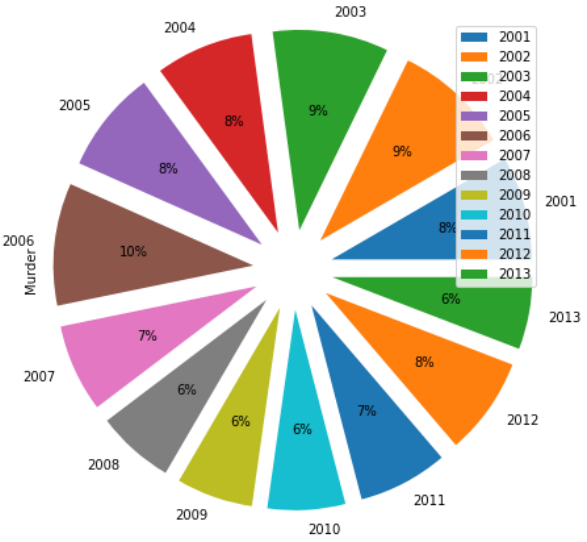
Stem and Leaf Plot

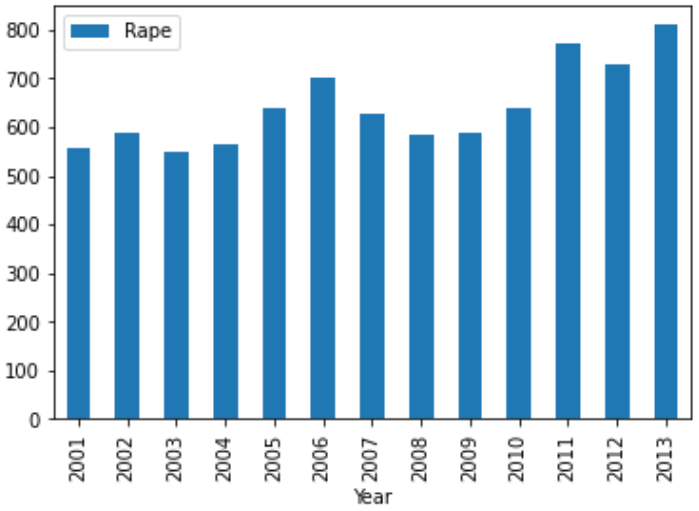


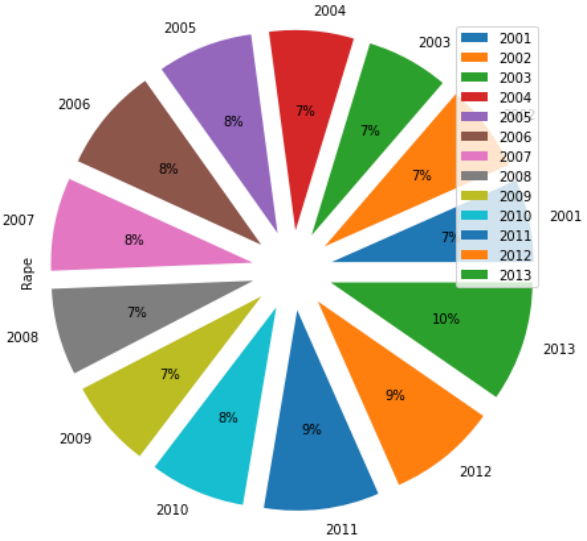


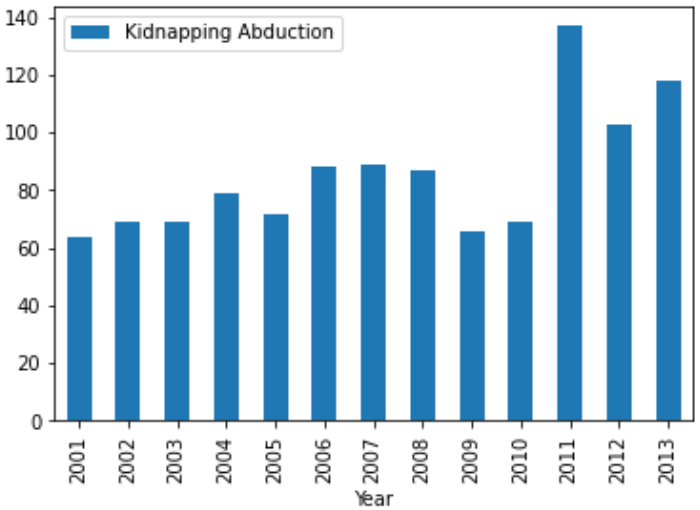
Type of crime wise Analysis

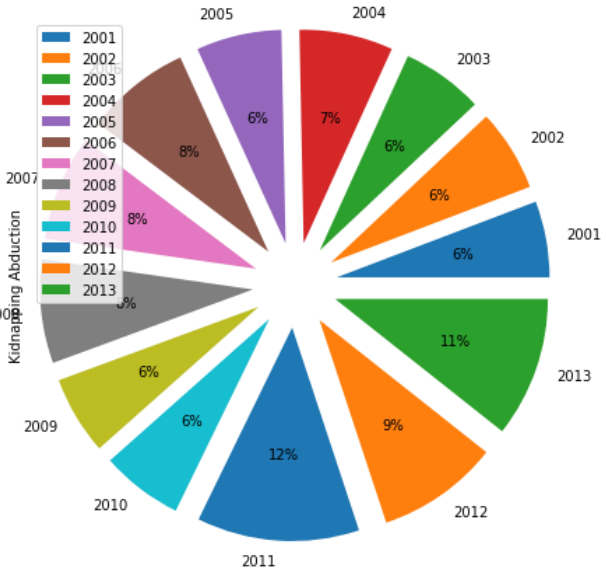


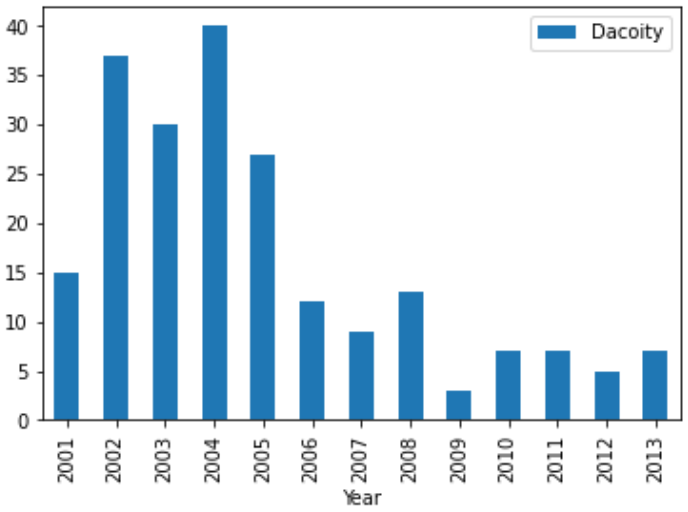


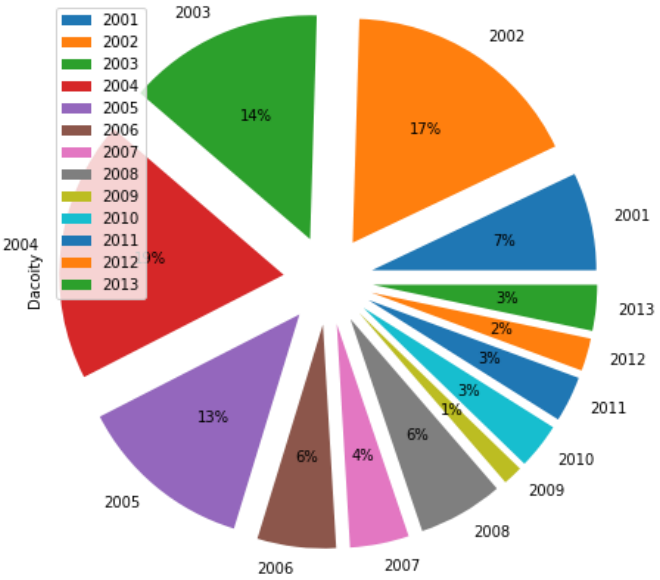


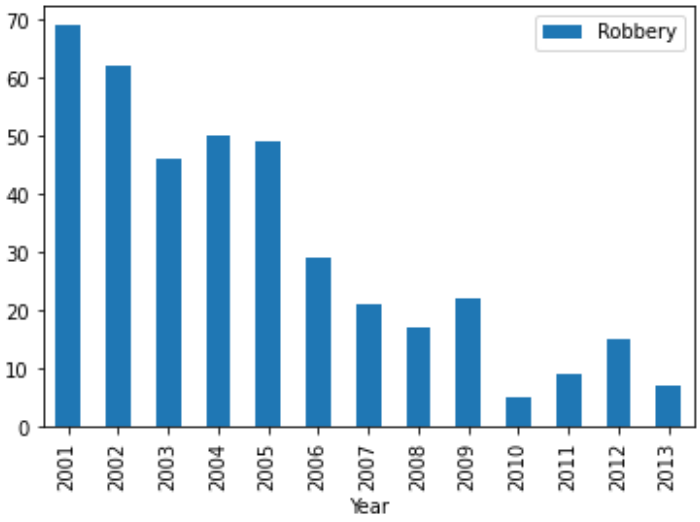


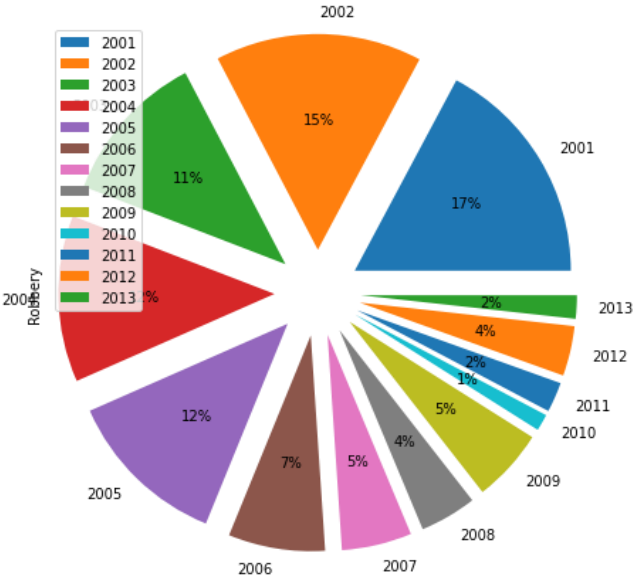


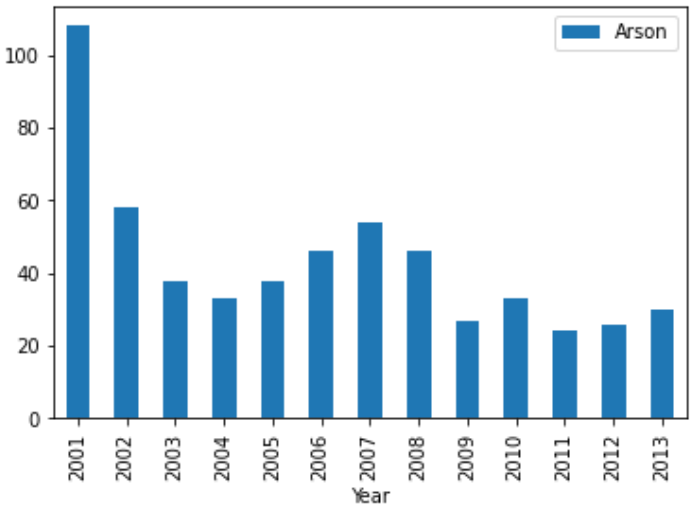


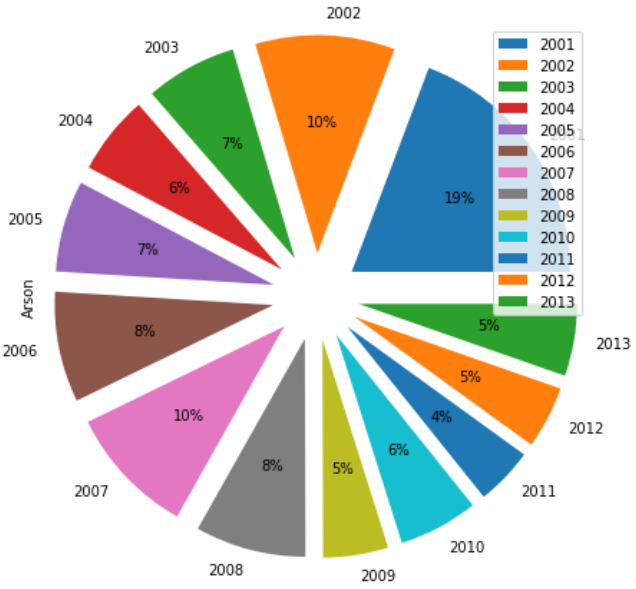

Dacoity
Year
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013

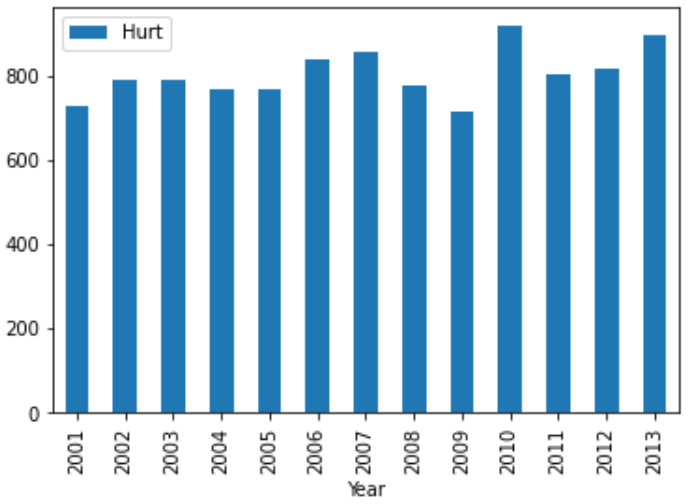

Dacoity
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
14%
17%
7%
3%
2%
3%
3%
1%
6%
4%
6%
13%

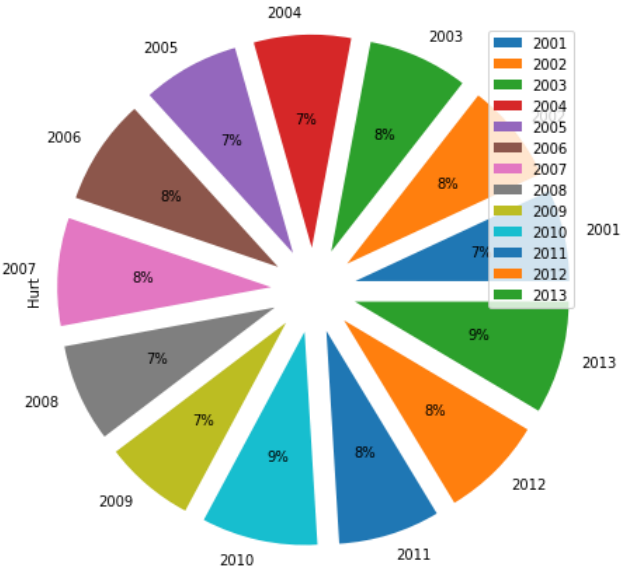

Robbery
Year
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013

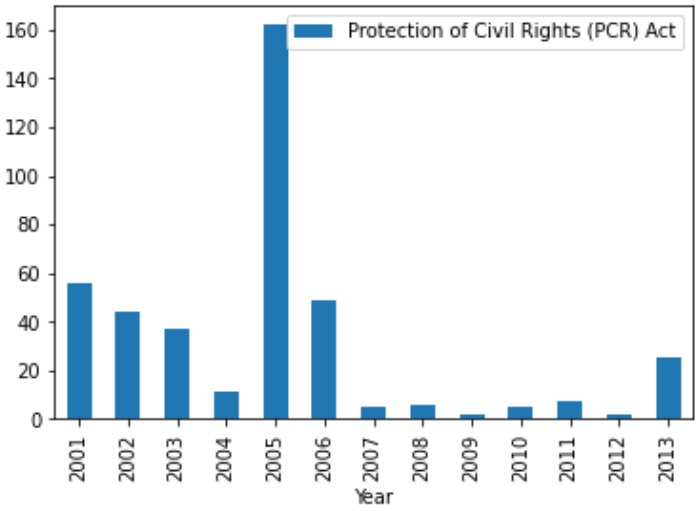

Robbery
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
11%
15%
17%
2%
4%
2%
1%
5%
4%
5%
7%
12%

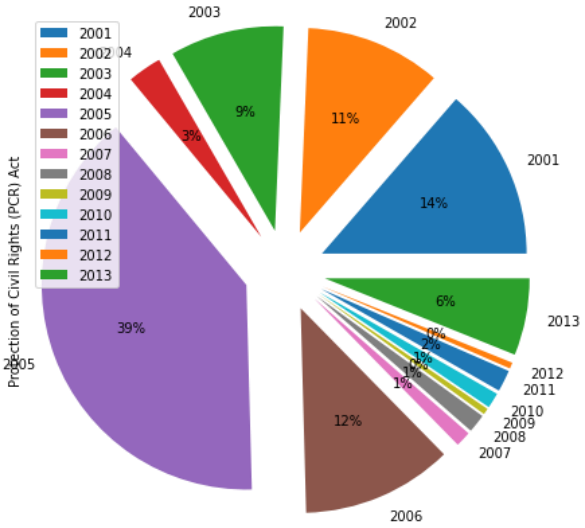

Arson
100
80
60
40
20
0
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
Year

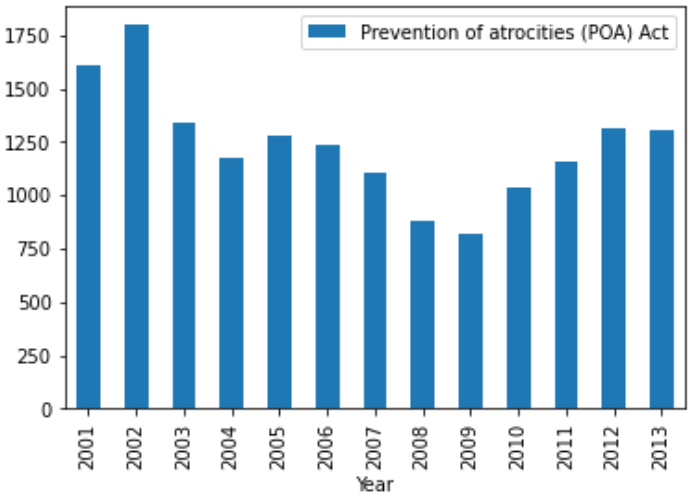

2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
Arson
19%
10%
7%
6%
7%
8%
10%
8%
5%
6%
4%
5%
5%

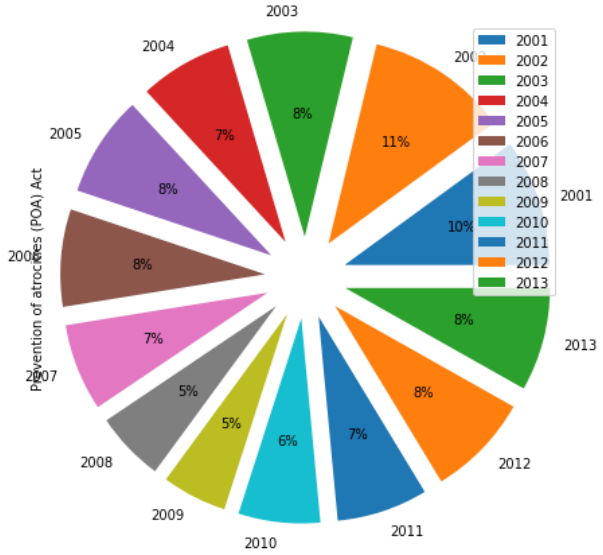

Hurt
0
200
400
600
800
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
Year

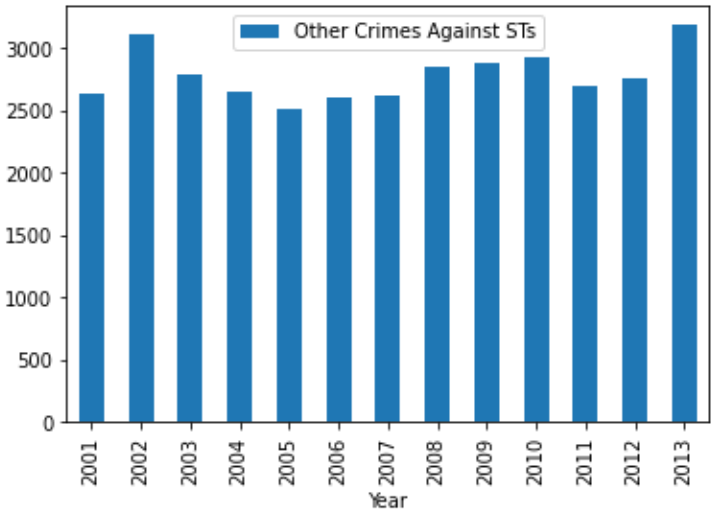

Hurt
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
7%
8%
8%
7%
7%
8%
8%
7%
9%
8%
8%
9%
7%

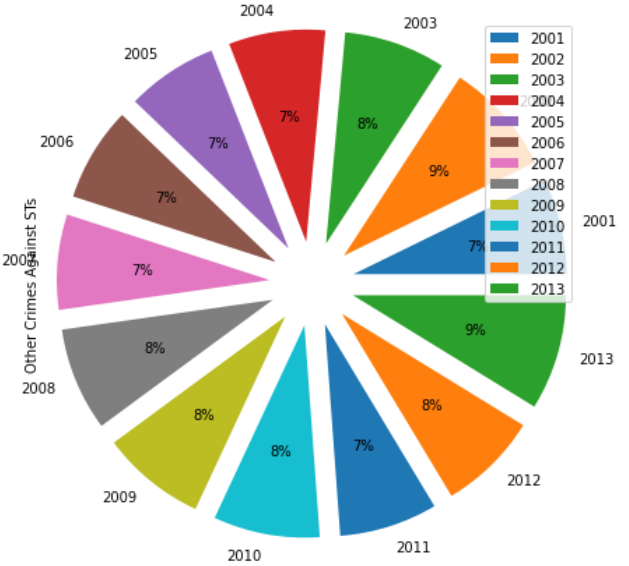

Protection of Civil Rights (PCR) Act
0
20
40
60
80
100
120
140
160
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
Year

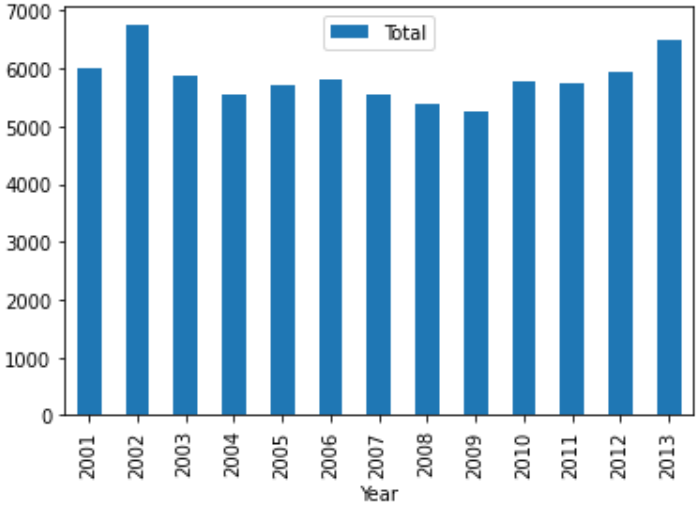

Protection of Civil Rights (PCR) Act
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
14%
11%
9%
3%
39%
12%
6%
2%
2%
1%
1%
1%
1%

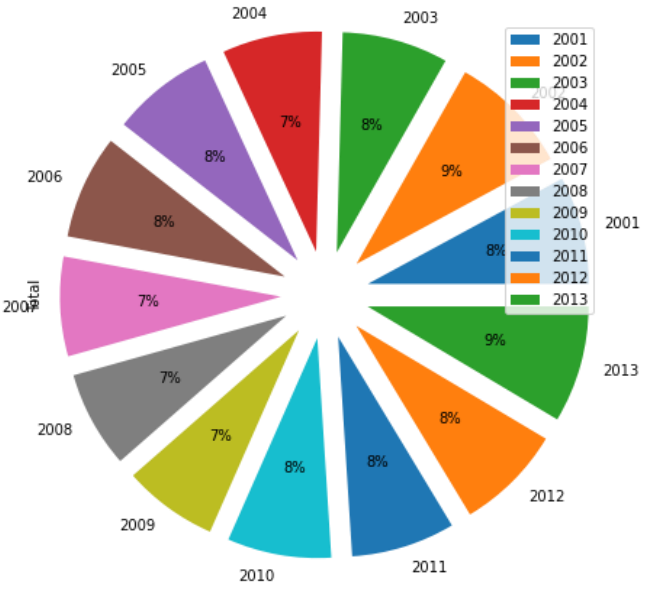

Prevention of atrocities (POA) Act
0
250
500
750
1000
1250
1500
1750
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
Year


Prevention of atrocities (POA) Act
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
10%
11%
8%
7%
8%
8%
7%
5%
5%
6%
7%
8%
8%







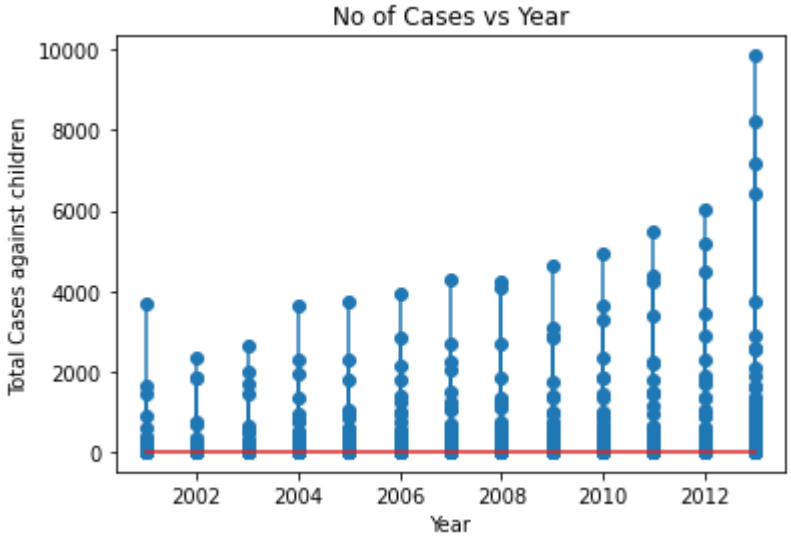


Categorical yearly crime rates

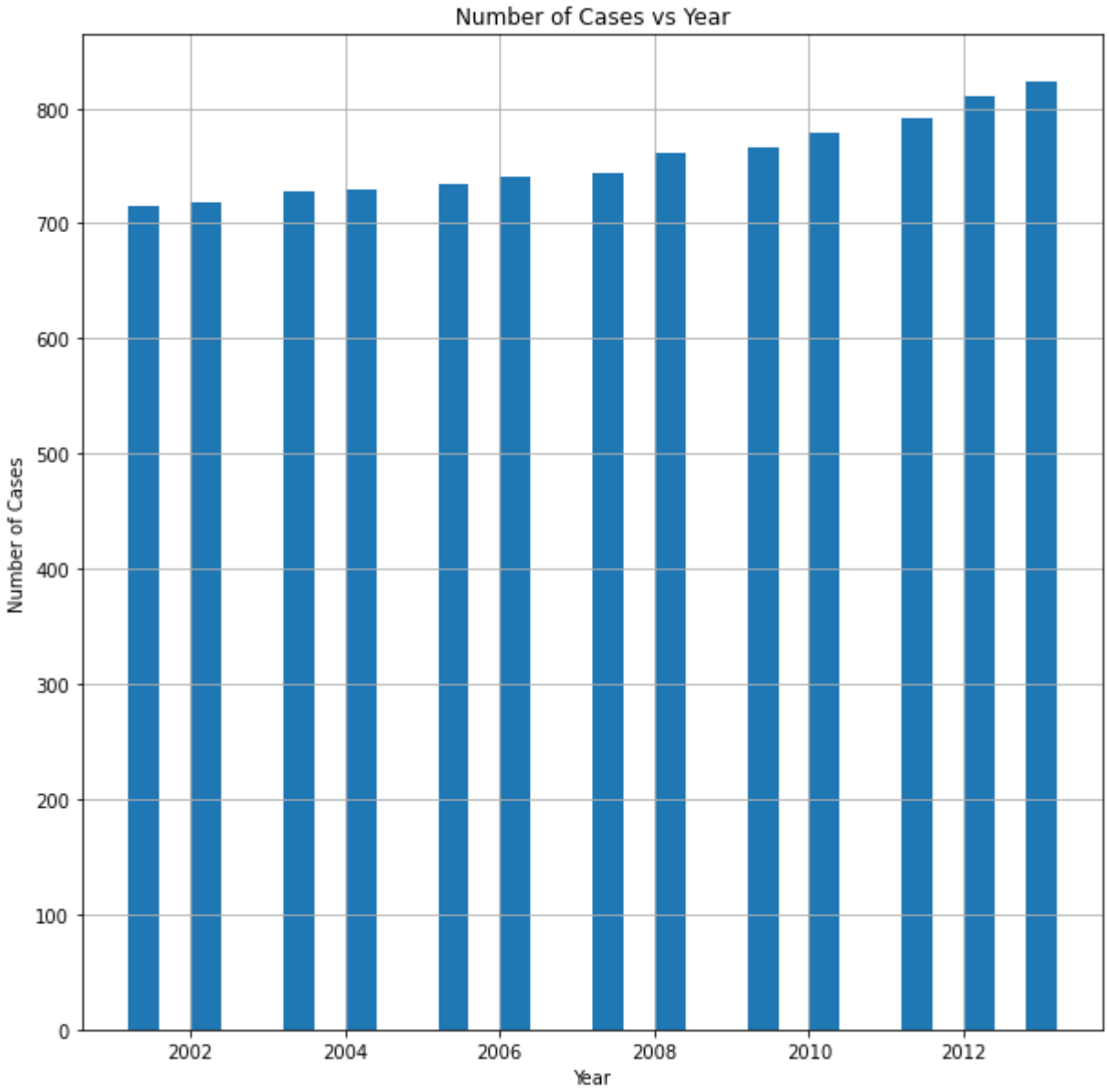
Year	Murder	Rape	Kidnapping Abduction	Dacoity	Robbery	Arson	Hurt	Protection of Civil Rights (PCR) Act	Prevention of atrocities (POA) Act	Other Crimes Against STs
2001	165	558	64	15	69	108	727	56	1609	2634

200 2	186	590	69	37	62	58	788	44	1797	3108
200 3	185	551	69	30	46	38	790	37	1340	2798
200 4	156	566	79	40	50	33	767	11	1175	2658
200 5	164	640	72	27	49	38	767	162	1283	2511
200 6	195	701	88	12	29	46	840	49	1232	2604
200 7	140	627	89	9	21	54	855	5	1104	2628
200 8	124	585	87	13	17	46	777	6	882	2855
200 9	122	589	66	3	22	27	715	2	822	2882
201 0	124	640	69	7	5	33	917	5	1037	2927

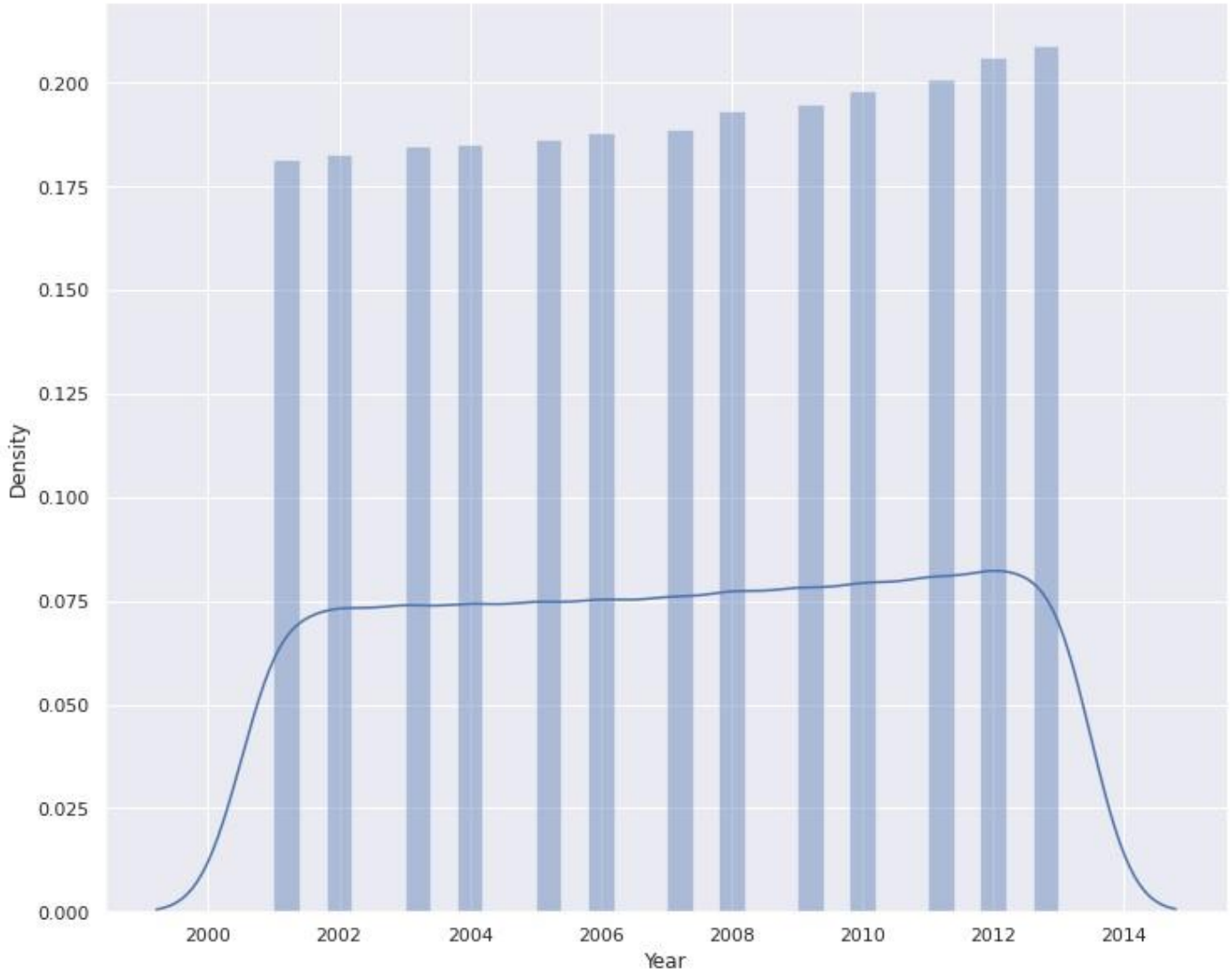
201 1	143	772	137	7	9	24	803	7	1154	2700
201 2	156	729	103	5	15	26	816	2	1311	2759
201 3	115	810	118	7	7	30	895	25	1309	3187



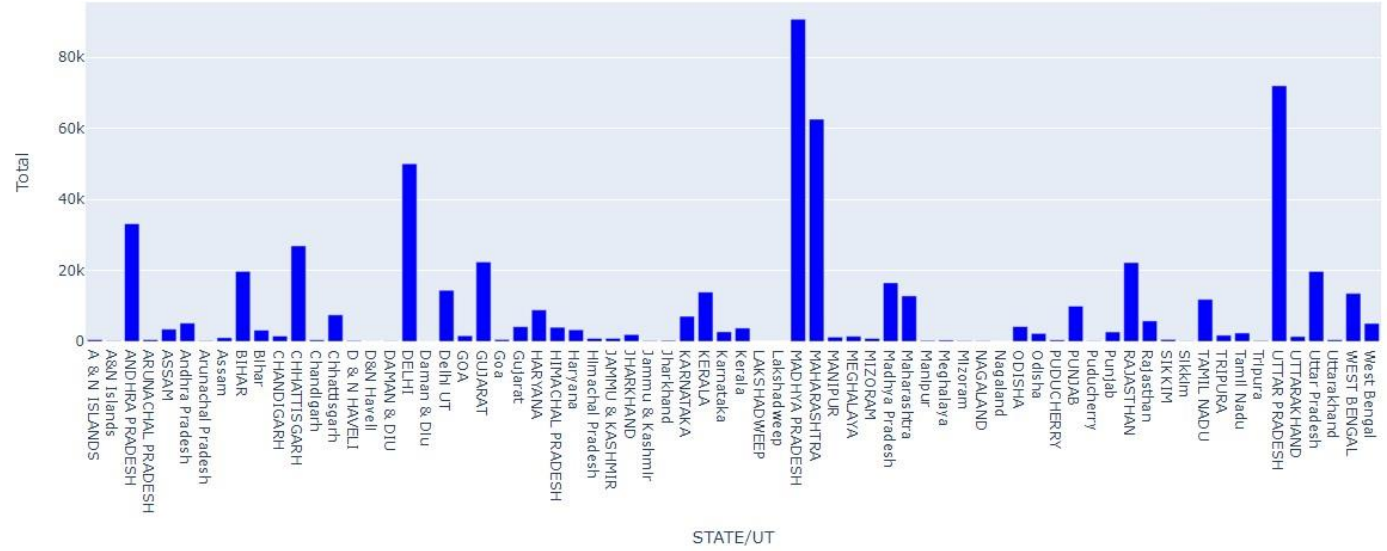
The cases against children is maximum in 2013 with almost 10000 cases in a state while the lowest was in 2002 when the maximum was less than 4000 cases against children in any state.



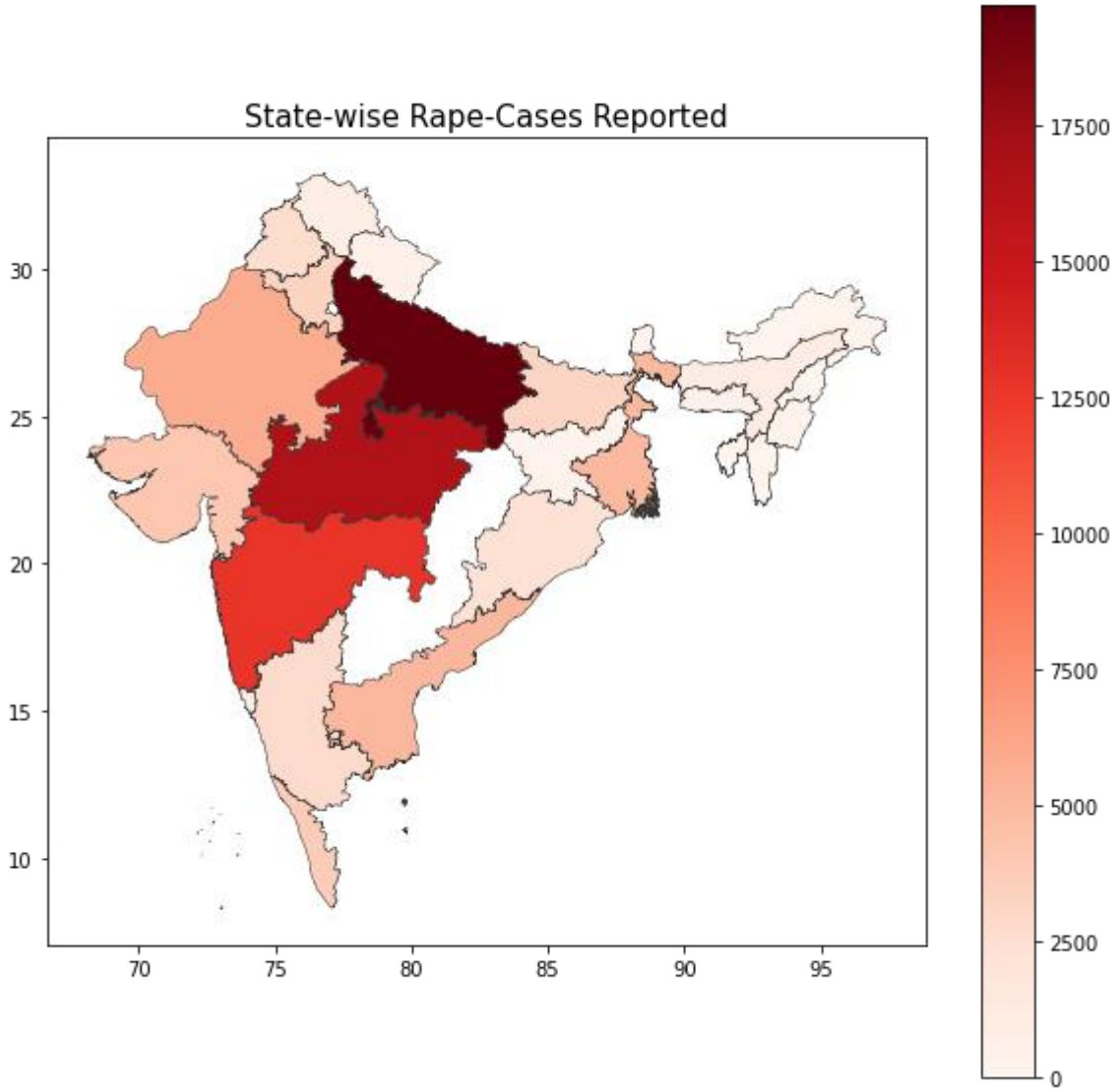
The total number of cases has seen a steady increase over 800k in 2013 and over 700k in 200.



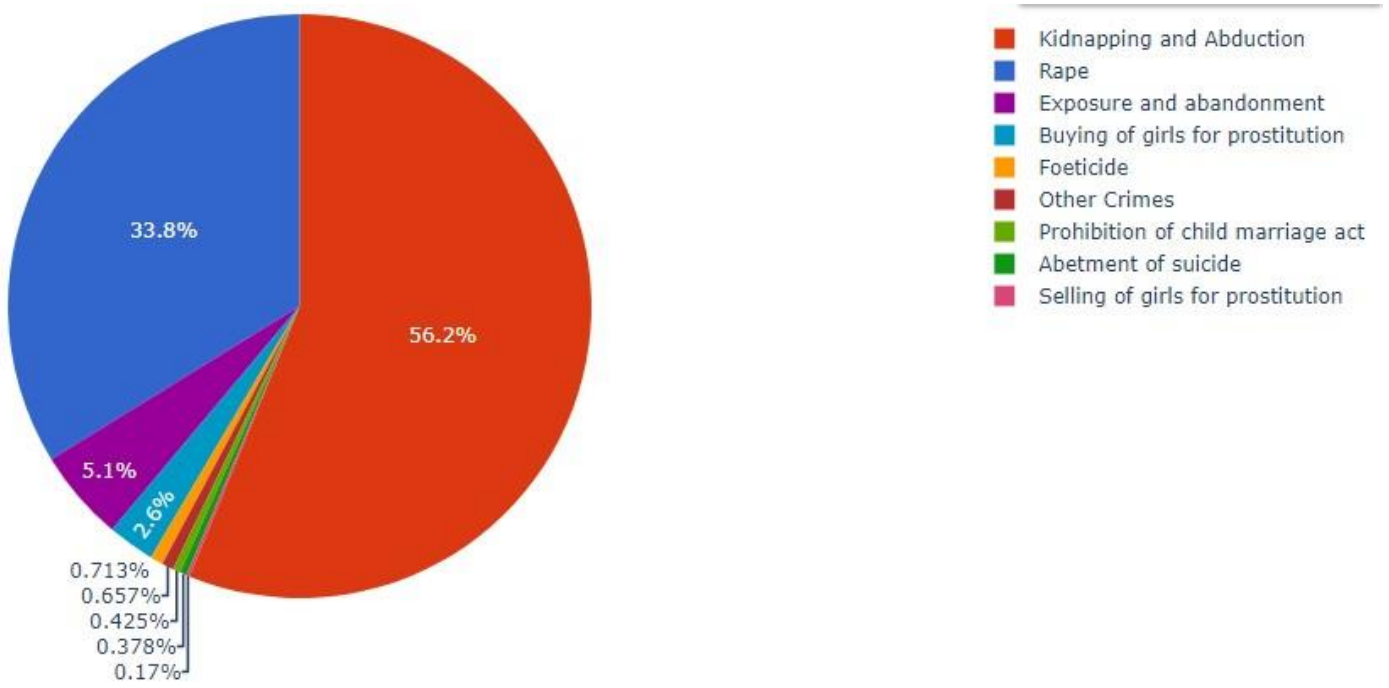
There has not been much difference in the total number of cases over the span of 13 years which can be understood from the density plot.



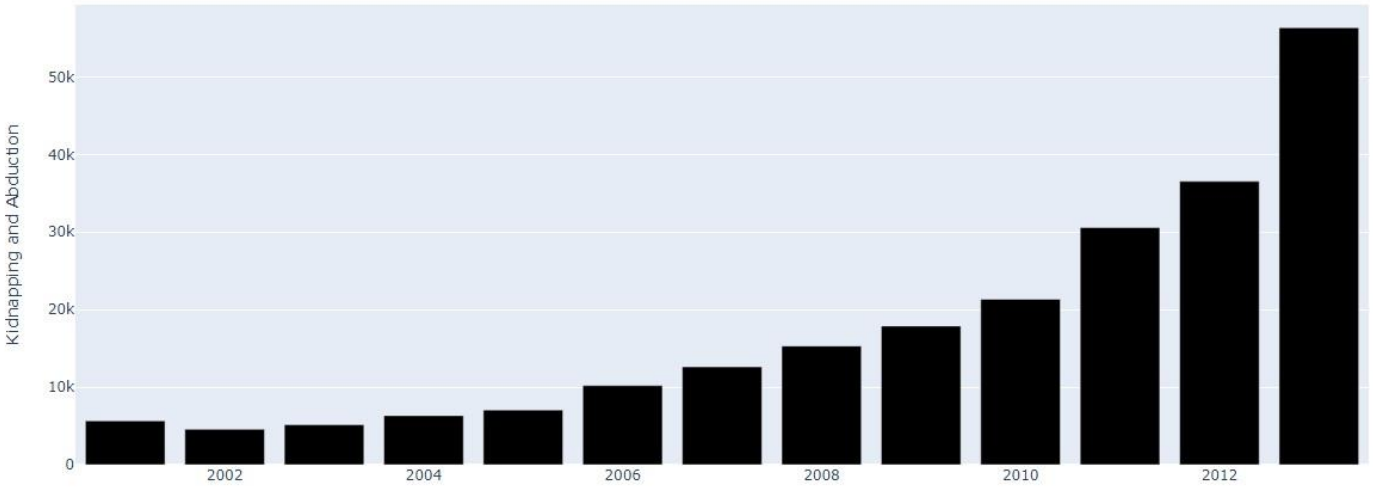
Madhya Pradesh had the maximum number of cases against children filed with more than 80000 cases.



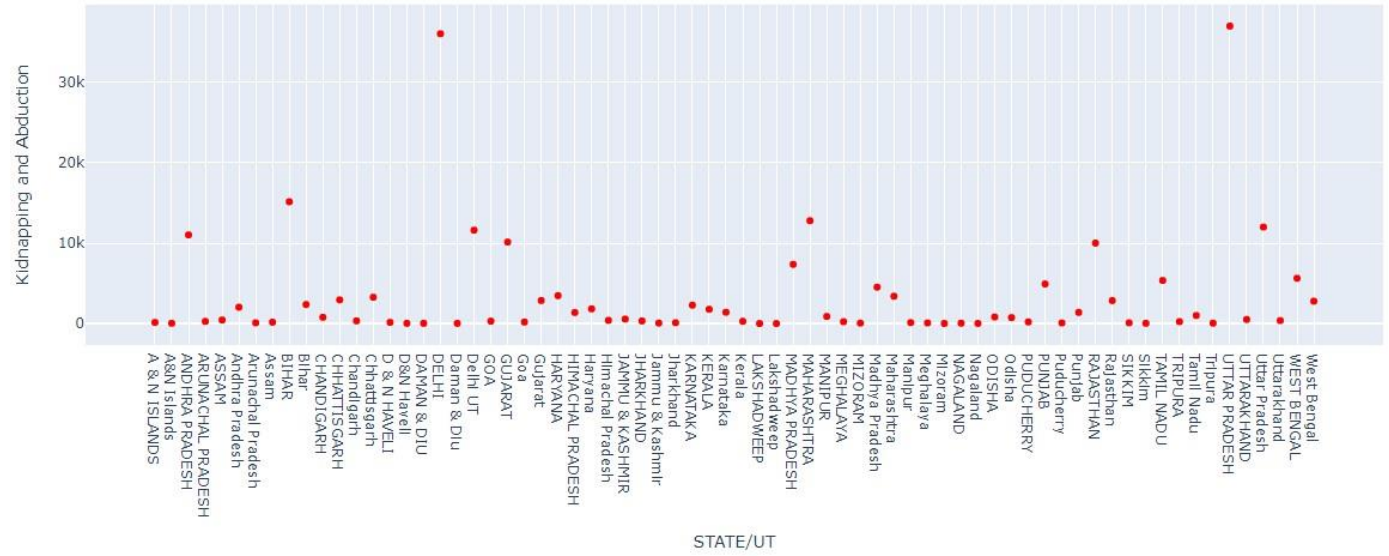
Uttar Pradesh with 19714 rape cases, Madhya Pradesh with 16494 rape cases and Maharashtra with 12820 were the three states with maximum rape cases against children.



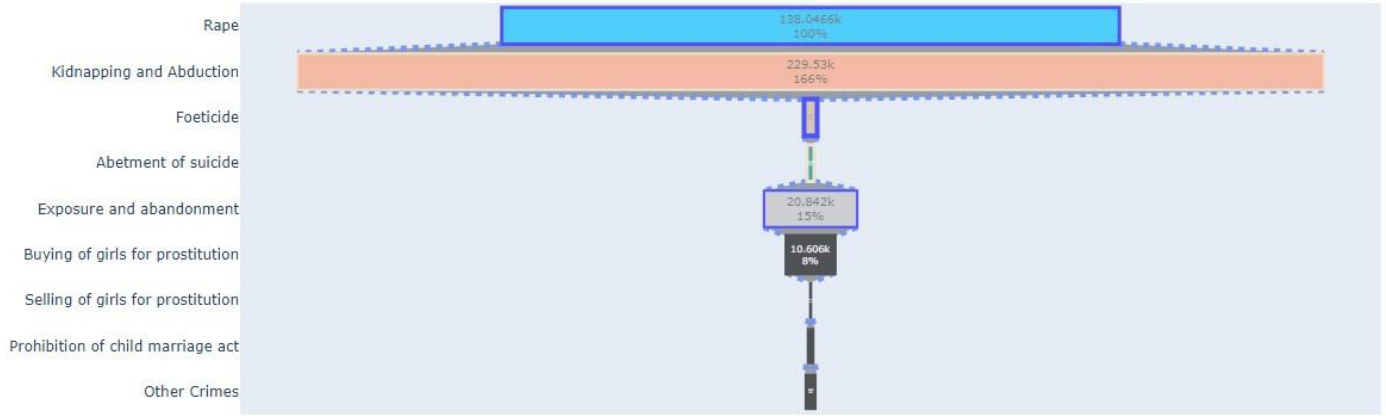
Kidnapping and Abduction of Children was the major crime with 56.2% of the total cases, followed by rape at 33.8% and exposure and abandonment at 5.1%.



The number of kidnapping and abduction has seen a steady increase in cases over the 12 years span with more than 50000 cases in 2013 alone.

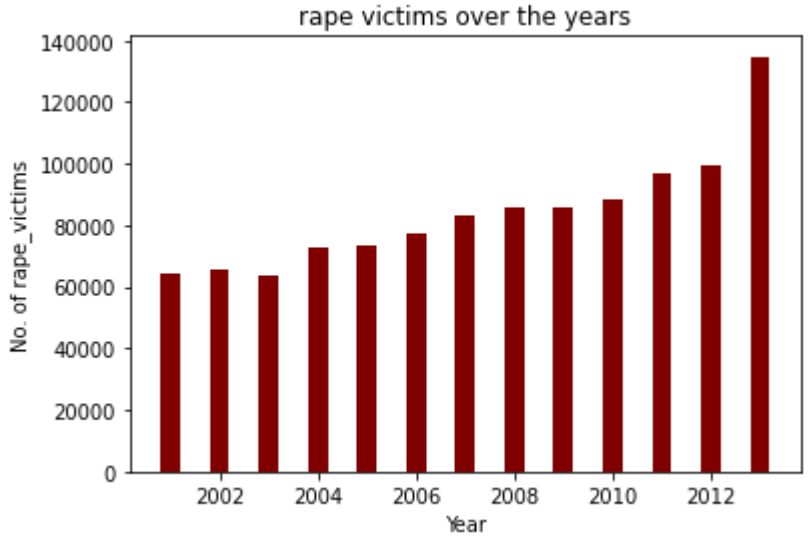


Uttar Pradesh has the highest number of kidnapping and abduction cases at 37012 followed closely by Delhi at 36055 cases and Bihar finishes third on this list with 15200 cases accounting to 78300 cases over the 12 years span from 2001 to 2013.

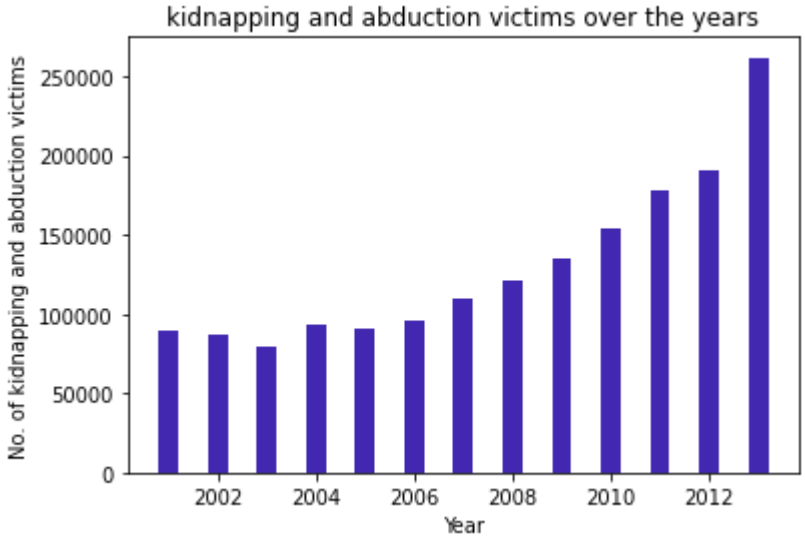


The funnel chart clearly indicates kidnapping and abduction as the maximum constituent of the total number of cases reported over the 13 year span.

IPC Dataset Analysis (General crimes non individual specific)



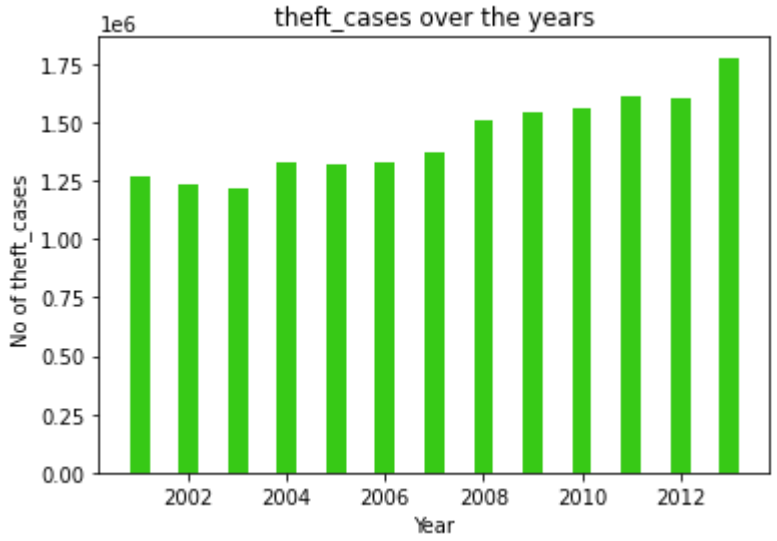
It can be inferred that the number of rape victims has increased by over 8000 in a span of 10 years.



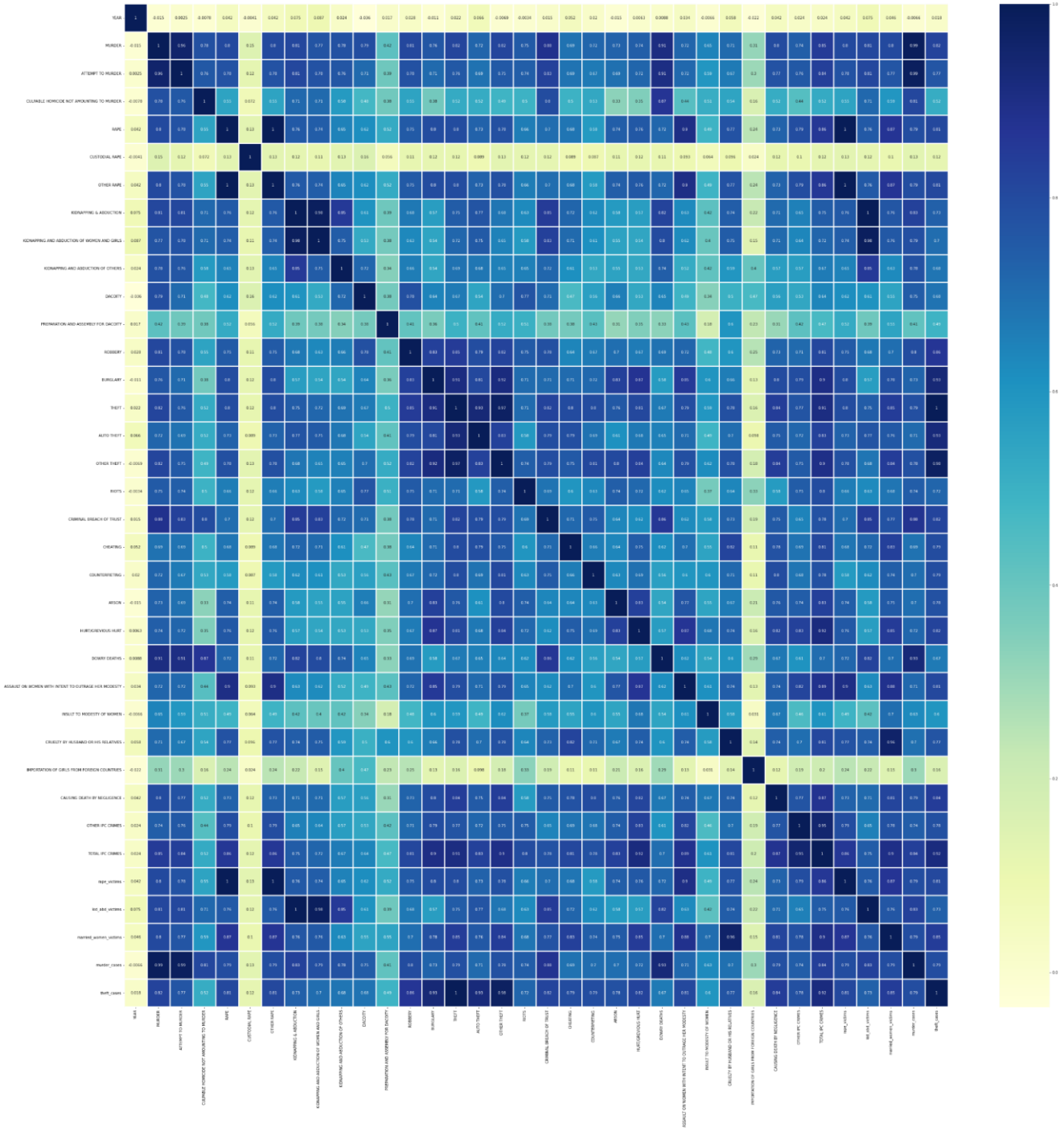
The total number of kidnapping and abduction which includes women, girls and children saw a small dip in 2003 but otherwise been a steady upward trend since then till 2013 with the number of cases having tripled from approximately 8000 to 24000.

A greater number of married women have become victims to marital rape with more than 40000 cases in 2013 as compared to around 20000 cases in 2003.

The number of murder cases every year over the 12 year span has not seen much reduction and has hovered around 12000 and more for over 10 years.



The number of thefts has increased by 150% from 2001 to 2013. (a gradual increase compared to rape, abduction etc)



- The heat map has been plotted with respect to correlation between the various categories of crimes against women and can be inferred that rape of women leads to murder 80% of the time with 70% by uncles and husbands.

- Kidnapping or abduction.can also be witnessed to have an 80% murder rate with 95% of such activities ocurring to girls.

- Victims who are married women are 95% subject to cruelty by their husbands with 86% of them having been raped and 76% of them are murdered.

- Almost 99% of all rape victims have been murdered over the 13 year span while 82% of those cases are due to kidnapping and abduction.

- 91% of the total cases reported are theft cases.

- The columns with less than 70% correlation were dropped to reduce the dataset and ensure greater accuracy.

Predictive Analysis (on IPC dataset)

The arreibutes that had a correlation above 70% were chosen to ensure a high accuracy.

The attributes are as follows:

1. MURDER
2. ATTEMPT TO MURDER
3. RAPE
4. OTHER RAPE
5. KIDNAPPING & ABDUCTION
6. KIDNAPPING AND ABDUCTION OF WOMEN AND GIRLS
7. ROBBERY
8. BURGLARY
9. THEFT
10. RIOTS
11. CRIMINAL BREACH OF TRUST
12. CHEATING
13. COUNTERFIETING
14. ARSON
15. HURT/GREVIOUS HURT
16. ASSAULT ON WOMEN WITH INTENT TO OUTRAGE HER MODESTY
17. CRUELTY BY HUSBAND OR HIS RELATIVES
18. CAUSING DEATH BY NEGLIGENCE

- The target attribute chosen was **THEFT**.

Regression Analysis has been performed as the dataset is more suited to it.

KNN - Regression

KNN - Regressor was trained with 2, 3 and 5 neighbours.

2 Neighbours

R^2 (coefficient of determination) regression score function for 2 neighbours was **0.9880**

3 Neighbours

R^2 (coefficient of determination) regression score function for 3 neighbours was **0.9878**

5 Neighbours

R^2 (coefficient of determination) regression score function for 5 neighbours was **0.9829**

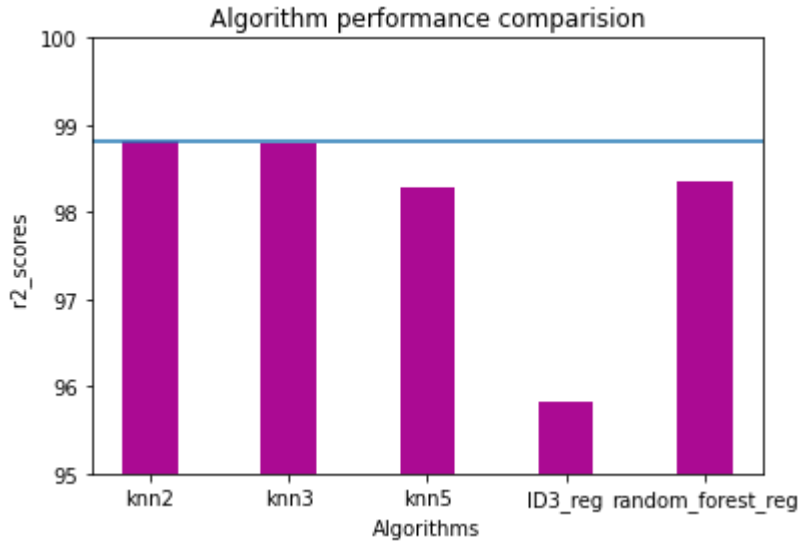
Decision Tree Regression

R^2 (coefficient of determination) regression score function was **0.9582**

Random Forest Regression

R^2 (coefficient of determination) regression score function was **0.9834**

Comparison of different regression models



The results clearly show that KNN with n=2 is the best suited of the regression algorithms implemented, with r2 score decreasing as n increases.

(these algorithms are not strictly restrictive and other regression methods can be applied based on the target attribute and dataset chosen)

END