CHILD CRIMES IN INDIA (2001-2012)

INTRODUCTION

India is home to 1.39 billion people, making it the second most populous country in the world. It accounts for 17.7% of world's population. 35.3% of the population of the country has been in the age group 0-14 years according to the Census of 2001. 41% of the population account for less than 18 years of age. Around 50% of this population is exposed to the risks of being victims to physical and emotional abuse.

The crimes in which children are the victims is considered as crime against children. The Juvenile Justice (Care and Protection of Children) is an act that caters to this bracket by giving them a room to protect themselves and exercise their rights.

This data consists of in total 35 States and Union Territories, spread across 12 years [2001-2012] in correspondence to number of criminals arrested under various Crimes Heads against children like:

- Infanticide
- Murder Of Children
- Rape Of Children
- Kidnapping And Abduction of Children
- Foeticide
- Abetment Of Suicide
- Exposure And Abandonment
- Procuration Of Minor Girls
- Buying Of Girls for Prostitution
- Selling Of Girls for Prostitution
- Prohibition Of Child Marriage Act
- Other Crimes Against Children.

The data explores the central idea:

- the states that lead in crime rates
- the states which are considerably safer for the youth of the nation
- the year wise trend for year 2001-2012
- the trend of crimes against children in nation
- A bifurcation in distribution between state and union territories

Further, as we deploy various data mining techniques to make predictions, we explore various novel patterns that might give rise to new interpretations and make predictions for upcoming year.

METHODS

1] Pre-processing and Cleaning of Data:

The data set consists of non-null values. However, if checked carefully, a lot of entries seem redundant.

1] The ‘STATES/UT’ column shows total 38 unique values. Of these ‘TOTAL (UTs)', 'TOTAL (ALL-INDIA)' and 'TOTAL (STATES) lead to double counting and hence are removed.

2] The unique values of Crime Head also include ' TOTAL CRIMES AGAINST CHILDREN', which needs to be removed to avoid the effect of the cumulative sum.

3] Another incongruency in data was TOTAL CRIMES AGAINST CHILDREN in above table showed 273475 as its value. But the data value sum indicates a sum of the values contributing to the same result as 273595. Hence to overcome this incongruency we will neglect the row depiciting TOTAL CRIMES AGAINST CHILDREN.

2] EDA:

- Statistical Explorations
- Visualisation
- Data Mining

- *STATISTICAL EXPLORATIONS*

State and UT distribution of crime

state	
LAKSHADWEEP	1
DAMAN & DIU	44
NAGALAND	60
D & N HAVELI	103
ARUNACHAL PRADESH	205
PUDUCHERRY	245
SIKKIM	260
A & N ISLANDS	305
MIZORAM	335
MANIPUR	356
JAMMU & KASHMIR	357
MEGHALAYA	447
CHANDIGARH	625
UTTARAKHAND	879
GOA	909
TRIPURA	932
JHARKHAND	1249
ASSAM	1614
HIMACHAL PRADESH	2027
ODISHA	2283
KARNATAKA	2962
HARYANA	4719
PUNJAB	5023
TAMIL NADU	5830
WEST BENGAL	6051
RAJASTHAN	7229
KERALA	7912
DELHI	10655
BIHAR	12038
GUJARAT	12075
CHHATTISGARH	15203
ANDHRA PRADESH	18407
MAHARASHTRA	35351
MADHYA PRADESH	58329
UTTAR PRADESH	58575

Name: count, dtype: int64

Year wise distribution of crimes

	year	count
0	2001	13401
1	2002	12507
2	2003	13524
3	2004	16663
4	2005	17353
5	2006	20870
6	2007	22432
7	2008	25766
8	2009	26050
9	2010	27485
10	2011	35427

Average Crimes / State & UT

state	
LAKSHADWEEP	0.006944
DAMAN & DIU	0.305556
D & N HAVELI	0.715278
PUDUCHERRY	1.701389
A & N ISLANDS	2.118056
CHANDIGARH	4.340278
DELHI	73.993056

Name: count, dtype: float64

state	
NAGALAND	0.416667
ARUNACHAL PRADESH	1.423611
SIKKIM	1.805556
MIZORAM	2.326389
MANIPUR	2.472222
JAMMU & KASHMIR	2.479167
MEGHALAYA	3.104167
UTTARAKHAND	6.104167
GOA	6.312500
TRIPURA	6.472222
JHARKHAND	8.673611
ASSAM	11.208333
HIMACHAL PRADESH	14.076389
ODISHA	15.854167
KARNATAKA	20.569444
HARYANA	32.770833
PUNJAB	34.881944
TAMIL NADU	40.486111
WEST BENGAL	42.020833
RAJASTHAN	50.201389
KERALA	54.944444
BIHAR	83.597222
GUJARAT	83.854167
CHHATTISGARH	105.576389
ANDHRA PRADESH	127.826389
MAHARASHTRA	245.493056
MADHYA PRADESH	405.062500
UTTAR PRADESH	406.770833

Name: count, dtype: float64

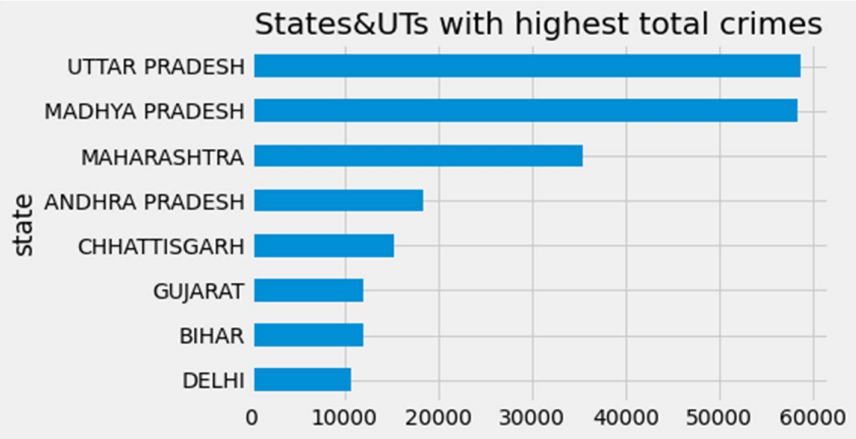
Selecting only top 5 crimes, except the "other crimes" and making its correlation matrix

crime	KIDNAPPING and ABDUCTION OF CHILDREN	MURDER OF CHILDREN	PROCURATION OF MINOR GIRLS	PROHIBITION OF CHILD MARRIAGE ACT	RAPE OF CHILDREN
KIDNAPPING and ABDUCTION OF CHILDREN	1.000000	0.936281	0.855784	-0.009293	0.965063
MURDER OF CHILDREN	0.936281	1.000000	0.789516	-0.258321	0.960549
PROCURATION OF MINOR GILRS	0.855784	0.789516	1.000000	-0.018643	0.802462
PROHIBITION OF CHILD MARRIAGE ACT	-0.009293	-0.258321	-0.018643	1.000000	-0.070326
RAPE OF CHILDREN	0.965063	0.960549	0.802462	-0.070326	1.000000

The strongest correlation is between Kidnapping and Abduction of Children and Rape of Children: **0.965063**. And the weakest correlation is between Kidnapping Kidnapping and Abduction of Children and Prohibition of child marriage act**: -0.0092**

- *VISUALISATIONS*

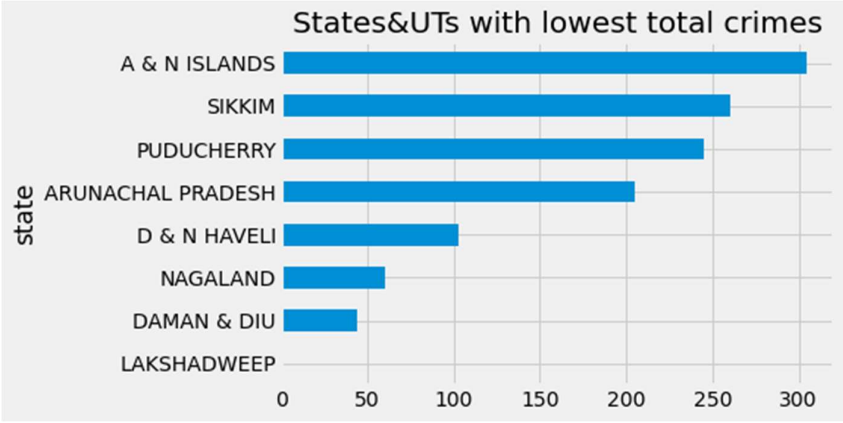
1. ***LEADING STATES AND UNION TERRITORIES IN CRIME RATES AGAINST CHILDREN IN INDIA***



The top states that registered most crime cases against children are Uttar Pradesh [58575], Madhya Pradesh [58329], Maharashtra [36351], Andhra Pradesh [18407], and Chattisgarh [15203]. Making sense out of this trend, Uttar Pradesh and Maharashtra are the most populated states of India.

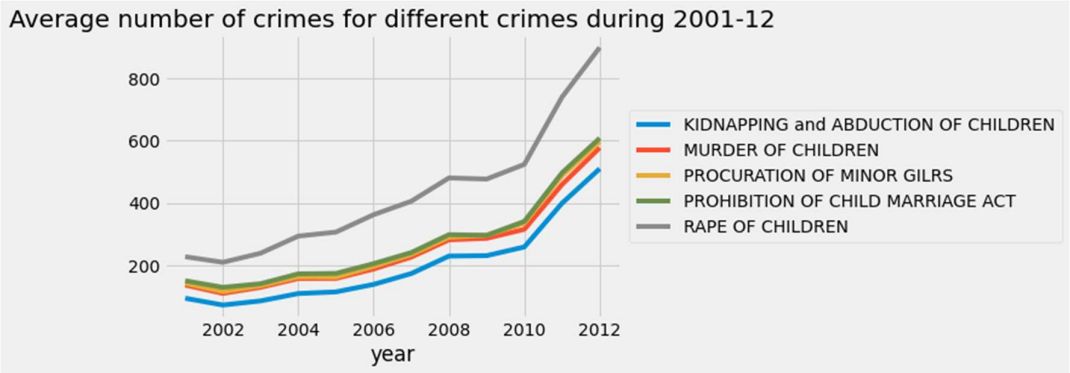
Also, Uttar Pradesh, Madhya Pradesh and Bihar have a major share of youngest population in the country, hence establishing to a reason to this pattern. Geographically too, the top three states take up a major chunk of the nation and hence account to larger representation of population.

2. ***LOWEST CRIME RATES IN STATES AND UNION TERRITORIES.***



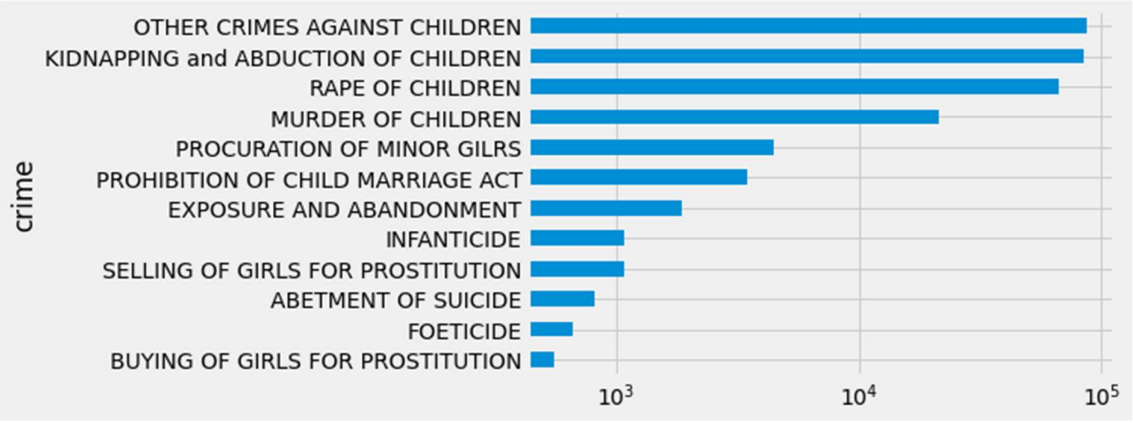
The states that registered least crime cases against children are Lakshadweep [1], Daman and Diu [44], Nagaland [60], D & N Haveli [103], and Arunachal Pradesh [205]. These all constitute to Union Territories. All the north eastern states, due to its significantly small geographic boundaries, hold accountable to lesser child crime rates as well. Andaman and Nicobar Islands is considered to be a sensitive area, which also accounts for occurrence of criminal activities.

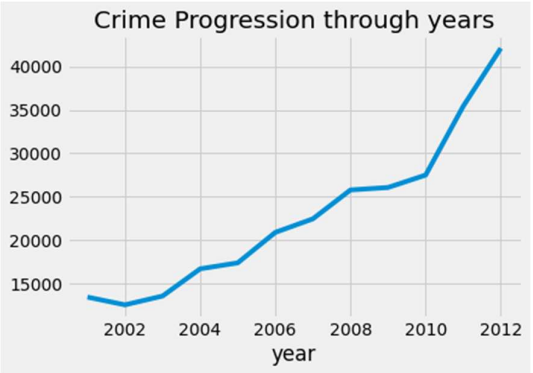
3. *YEAR WISE TREND OF CRIMES FOR YEAR 2001-2012*



To look up closer at fewer crimes, these are the top 5 crime, (excluding other crime) . This graph shows an upward trend in correspondence to the progression trend line that we saw. Rape of Children has dominated this graph, followed by Prohibition of Child Marriage Act, Procuration of Minor Girls and Murder of Children and lastly Kidnapping and abduction of children, all of which shows very similar trend.

4. THE NUMBER OF CRIMES AGAINST CHILDREN IN NATION





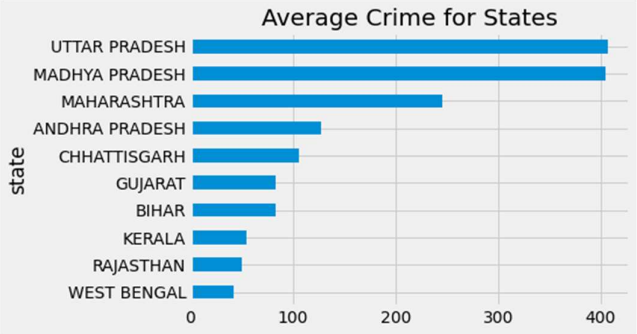
As per National Crime Records Bureau the major filed cases under crime against children in India (2019) were **Kidnapping and Abduction of Children (70%), Rape (5.07%), Assault on Women (4.35%), Other IPC Crimes (10.95%)** respectively.

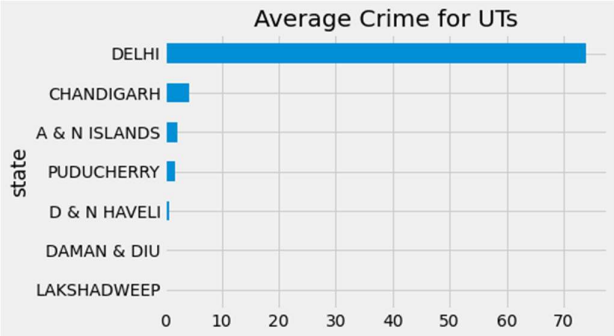
Due to Population rise, inflation, widening of economic gap and other contributory factors, a positive trend between years and crime rates is observed.

5. AVERAGE CRIMES STATE-WISE AND UNION TERRITORY WISE.

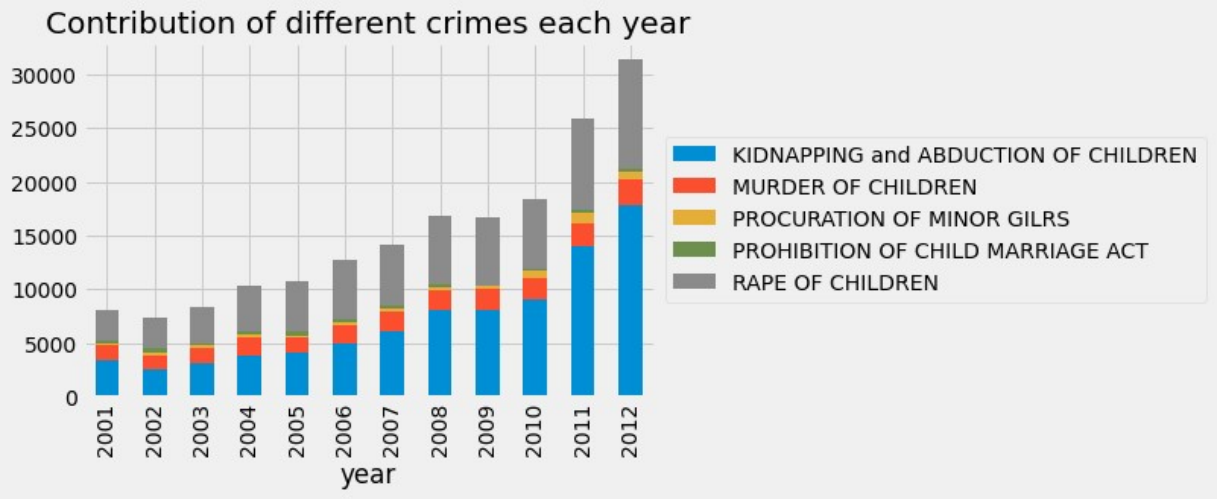
The report, released by The New Indian Express, claims that a case of rape is registered with the state police every two hours while crime against a child is reported every 90 minutes in Uttar Pradesh. The National Crime Record Bureau's (NCRB) recent data suggests that Uttar Pradesh tops the list in crime against minors, with around 15% share of total crime against children in the country. This report next highlighted Madhya Pradesh and Maharashtra in its contribution to crimes.

Delhi accounted for 35% of such crimes in metropolitan city by a report released by The New Indian Express. The data trend we observe for year 2001-2012 supports these patterns.





6. CONTRIBUTION OF DIFFERENT CRIMES IN EACH YEAR.



Out of the top 5 crimes, Kidnapping and Abduction of Children holds a major chunk in the total number of crimes committed against children. Also, its share increases progressively yearwise. Next ew have Rape of children, with a steady rise in contribution to toal crimes, yearwise. Procuration of Minor Girls also witnesses a rise, with a constant rate of Murder of Children.

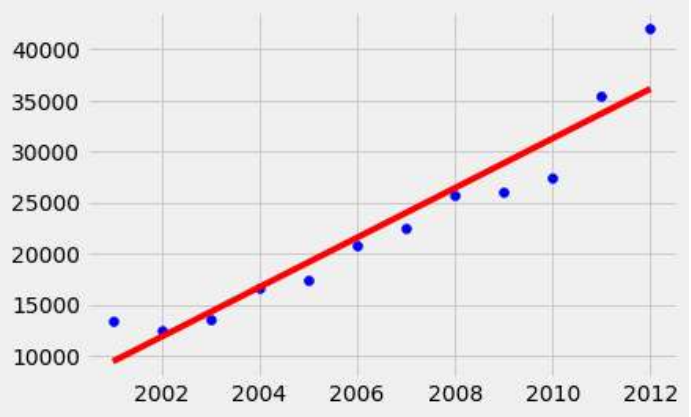
- *DATA MINING TECHNIQUES*

By performing data mining, we extract useful and pivotal information from the data set. This, is also known as knowledge discovery in data. This helps us understand some interesting phenomenon of the relationship between societal trends. We aim at eliminating redundant information and keeping relevant relationships under study.

Data mining is bifurcated into two sub branches as predictive and descriptive. In our case, we are using Regression Analysis to predict a future observation by studying the historical data. We also use elements of Clustering Analysis where each group of objects are defined accordingly.

This helps us in the direction of using the best choice of using algorithms that cater to our meets.

1] To understand the inherit patterns of this data set, a simple linear regression model was fit on total crimes against children in each year to predict number of crimes for the upcoming year. This is an elementary process of estimating the relationship between one dependent and other independent variable by employing a best fit straight line that minimises the square difference between predicted and actual values.



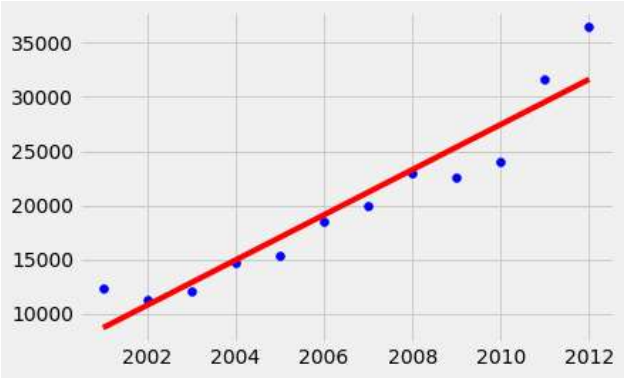
In our case, X are the number of years (independent) and Y are the total crimes against children (dependant). A scatter plot of the combination of these points shows an upward sloping trend. Hence, a simple regression is a good way to initialise the data mining process. Also, intuitively it does make sense that with progress in years the crime rates are also on rise. The model predicts a value of **38563.65** for the **year 2013**.

2] Now, we make another simple linear regression, by forming three clusters.

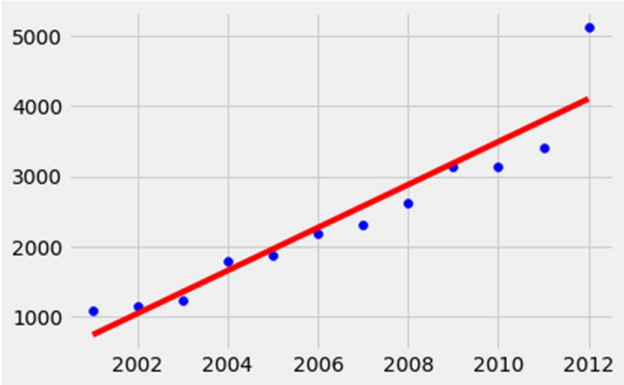
Clusters are formed on basis of number of crimes. We calculate the total number of crimes in each state overall, and then group them into high crime rate, medium crime rate and low crime rate.

Then we create a regression equation that is a combination of a cluster and a year of crime that we noticed earlier. By doing so we are now going a step further and exploring the geographical distribution of how the crime rates against children is dispersed in the nation.

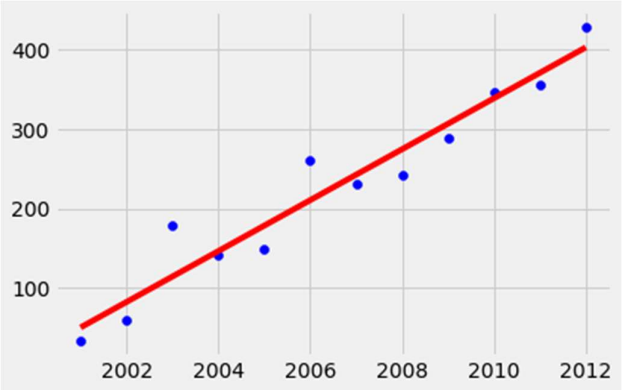
1. **Cluster 1** corresponds to **High total rates**. For **2013**, the regression line predicts a value of **33724.06**



2. **Cluster 2** corresponds to **Medium total rates**. For **2013,** the regression line predicts a value of **4404.81.**

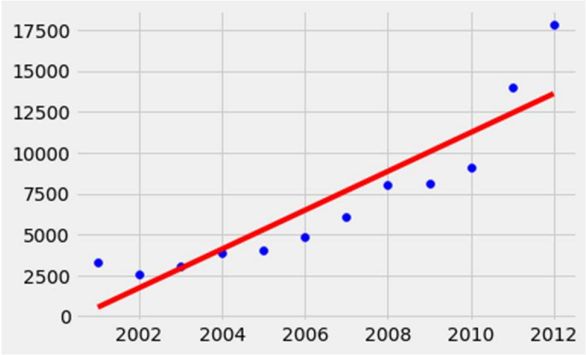


3. **Cluster 3** corresponds to **Low total rates**. **For 2013**, the regression line predicts a value of **434.77.**

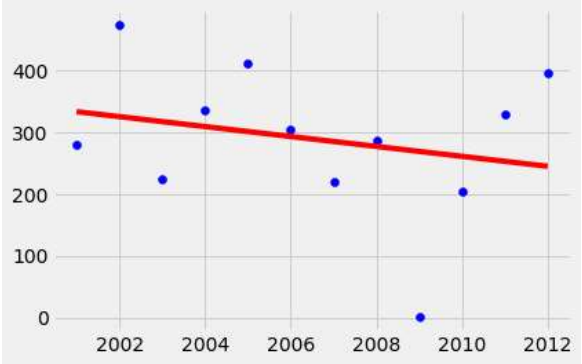


3]We repeat the first pattern of simple linear regression by manipulating the above equation. We take a particular of total of crime (for example foeticide) instead of total crimes that we used earlier in equation. This step helps us explore each crime into more detail. This helps us also correlate with research papers that have been conducted on these particular crimes. It also highlights how laws are not being strictly adhered to and where the loop holes in governance arise.

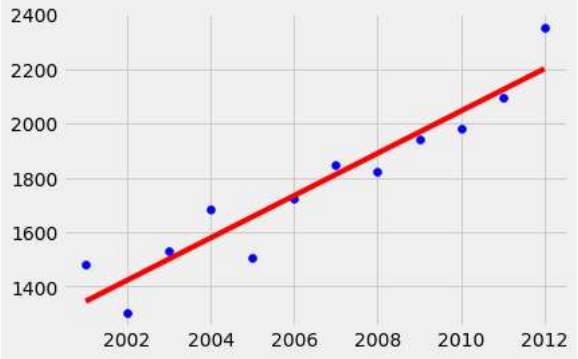
1. **KIDNAPPING and ABDUCTION OF CHILDREN: For 2013,** the regression line predicts a value of **14795.31.**



2. **PROHIBITION OF CHILD MARRIAGE ACT For 2013**, the regression line predicts a value of **237**.



3. **MURDER OF CHILDREN ACT For 2013**, the regression line predicts a value of **2277.69**



RESULTS AND CONCLUSION

- Union Territories in general account to lesser number of crimes as against states, naturally due to their geographic characteristics.
- Uttar Pradesh, Madhya Pradesh and Maharashtra [States], Delhi, Chandigarh, Andaman and Nicobar Islands [Union Territories] account to large crime rates.
- There has been a positive progression in crime rates year wise which can be attributed to variuos factors like: increasing population, , inflation, widening of economic gap etc.
- The strongest correlation is between Kidnapping and Abduction of Children and Rape of Children: 0.965063. And the weakest correlation is between Kidnapping and Abduction of Children and Prohibition of child marriage act: -0.0092
- Highest Crime Rates are due to Kidnapping and abduction of Children, Rape of Children and Murder of Children. Their share to contribution of total crimes has also risen significantly yearwise.
- The data mining techniques employed simple linear regression in three variations and predicted the values for the year 2013 respectively. These have been elaboratively discussed above. Overall, the predictions seem a value slighltly lower than previous year indicating a steady progression.