Capstone Project: Crimes in India

By Aishwary Shukla

ABSTRACT

The aim of this project was to analyse factors related to the crimes in India in different states and districts using Machine Learning. A data sets pertaining to different types of crimes occuring from 2001 to 2011 were provided. Additional task was to collect population, literacy rate and areas for each state in India to support the analysis. From this dataset intense data analysis was done using different visualization techniques using pandas library. univariate bivariate and multivariate analysis to analyse the relationship between the crimes and other demographic features of India. With the help of which we were able to present report and conclude the factors that affect different type of crimes in India. the whole capstone project consists of 4 phases we are in phase one contains data collection Phase 2 contains state wise crime analysis, phase 3 contains using SQLite to fetch records from data sets and phase v contains solution to clustering problem where we are clustering all the states into 3 groups which are critical sensitive where trying occurrences are very low.

Data References:

Criminal datasets were provided along with which demographic details of India work collected further analysis was done. Links from where data was collected is given below:

- https://en.wikipedia.org/wiki/List_of_states_in_India_by_past_population
- https://www.census2011.co.in/literacy.php
- https://en.wikipedia.org/wiki/List_of_states_and_union_territories_of_India_by_area

I would like to thank my trainer and my internship manager for giving me this opportunity to test my skills:

Mr. Shankar Gaud Tegimanni, DataTrained

Mr. Shwetank Mishra, FlipRobo

Mr. Prateeek Rajvanshi, Clevered Institute

And all colleagues, mentors, trainers from DataTrained and FlipRobo for training me and giving me the opportunity to be an intern and expand my knowledge in the field of Data Science and Artificial Intelligence.

INTRODUCTION

Crimes are a social nuisance, and it has a direct effect on a society. Governments spend lots of money through law enforcement agencies to try and stop crimes from taking place. Today, many law enforcement bodies have large volumes of data related to crimes, which need to be processed to turn into useful information [3]. Crime data are complex because they have many dimensions and in different formats, e.g., most of them contain string records and narrative records. Due to this diversity, it is difficult to mine them using off the shelf, statistical and machine learning data analytics tools.It is the primary reason for lack of general platform for crime data mining. While there are some propitiatory platforms to predict and analyze crime data, they are focused only on certain areas of crimes, not extensible, and do not provide an API to integrate with other tools [4]. Moreover, the same tool cannot be used for the analysis and well as planning such as patrol beads and district boundaries.

Problem Statement

The research problem that this project tries to address can be stated as follows:

To curb the criminal activities by identifying the hotspots, analysing the relationship between crime rate and population, area, and literacy rate of India.

Research Objectives

The project is divided into 4 phases which have certain set objectives outlined below:

Phase 1

Data Collection/preparation part.

1. *The population of each state.*
2. Literacy Rate in each state
3. *Area of each state*
4. *Collect any other data that helps with your analysis. There is no limitation for anything.*
5. *Create a new file and keep the above-collected data.*

Phase 2

State/UT wise analysis.

2.1 Analysis of Literacy rate vs Total crimes.

2.2 Analysis of the type of crime vs each state vs Literacy rate.

2.3 Analysis of year-on-year total crime rate.

2.4 Analysis of area vs overall crime

2.5 Analysis of population vs overall crime

2.6 Each state crime report. There is no fixed format to write a report, you can write a report inside the notebook itself based on what you have analyzed in the above points.

Phase 3

SQL Operations

3.1 Insert records from 42_District_wise_crimes_committed_against_women_2001_2012.csv into a table

3.2 Write SQL query to find the highest number of rapes & Kidnappings that happened in which state, District, and year

3.3 Write SQL query to find All the lowest number of rapes & Kidnappings that happened in which state, District, and year

3.4 Insert records from 02_District_wise_crimes_committed_against_ST_2001_2012.csv into a new table

3.5 Write SQL query to find the highest number of dacoity/robbery in which district.

3.6 Write SQL query to find in which districts(All) the lowest number of murders happened.

3.7 Write SQL query to find the number of murders in ascending order in district and yearwise.

3.8.1 Insert records of STATE/UT, DISTRICT, YEAR, MURDER, ATTEMPT TO MURDER, and RAPE columns only from 01_District_wise_crimes_committed_IPC_2001_2012.csv into a new table

3.8.2 Write SQL query to find which District in each state/ut has the highest number of murders yearwise. Your output should show STATE/UT, YEAR, DISTRICT, and MURDERS.

3.8.3Store the above data (the result of 3.2) in DataFrame and analyze districts that appear 3 or more than 3 years and print the corresponding state/ut, district, murders, and year in descending order.

3.8.4Use appropriate graphs to show your data (the result of 3.8.3)

Phase 4

Unsupervised ML (Clustering)

4.1 "Create 3 clusters as below.

1. Sensitive Area's

2. Moderate Area's

3. Peaceful Area's"

4.2 Create DataFrame for each cluster that shows data according to the areas.

4.3 "Analyse your clusters and prepare a report that explains all your observations.

Example -

1. What is impacting more crimes in sensitive areas.

2. What needs to be done to reduce crime.

3. Most safe and unsafe districts.

4. Etc.....Anything that you observe (If you want, you can plot various graphs to analyze)"

4.4 Capstone project overall story in your own words. Min 1000 words.

Outline

The rest of the report is organized as follows. Phase 1 presents from where the data was collected. State wise analysis of crimes and crime gates is done in Phase 2. in phase 3 we have fetched records by connecting with SQLite3 database. Phase 4 includes segregation of states into 3 reasons namely peaceful moderate and highly sensitive areas with respect to criminal activities happening in those regions. Summary and future work is presented 5 at the end.

LITERATURE REVIEW

A crime can be defined as any action or omission that violates a law, which results in a punishment. Usually what constitutes as a crime depends on the government bodies and laws that are in existence in those places. To understand the nature of crimes, one has to understand not only its spatio-temporal dimensions, but also the nature of the crime, the victim-offender relationship, role of guardians, and the history of similar incidents. Regardless of the reasons why crimes take place, they put a strain on the communities, towns, and cities. Usual monetary costs associated with them include cost of policing crime and prosecuting those who commit crimes. Non-monetary costs consist of social costs, where they affect the quality of life, mental health, and physical security of people living in those areas. Crimes are a social nuisance and being able to solve them faster is very important and will pay for itself.

Crime Analysis

Crime analysis is a difficult task, as it requires both collection and analysis of large volumes of data. Data mining techniques can be used in law and enforcement for crime data analysis, criminal career analysis, bank fraud analysis, and analysis of other critical problems. Some techniques are association analysis, classification and prediction, cluster analysis, and outlier analysis, which identify patterns in structured data. Data analysis and insights drone from the past criminal activities can be used to make conclusions which would help in simplifying the process while driving more accurate and insightful conclusions and predictions.

Algorithms used

K-means Clustering Algorithm This algorithm is mainly used to partition the clusters based on their mean. As a first step number of objects are grouped and specified as k clusters. K numbers of objects are initially selected as the cluster centers. Then again, these objects are assigned based on cluster center. Then cluster means are updated again. This algorithm is used as a base for most of the other clustering algorithms.

Phase 1: Data Collection Phase

The data has been collected from following sites:

1. https://en.wikipedia.org/wiki/List_of_states_in_India_by_past_population

Statewise population in descending order of Population Census 2011 is mentioned below:

	State/UT	Population
0	Uttar Pradesh	199812341
1	Maharashtra	112374333
2	Bihar	104099452
3	West Bengal	91276115
4	Madhya Pradesh	72626809
5	Tamil Nadu	72147030
6	Rajasthan	68548437
7	Karnataka	61095297
8	Gujarat	60439692
9	Andhra Pradesh	49577103
10	Odisha	41974219
11	Telangana	35003674
12	Kerala	33406061
13	Jharkhand	32988134
14	Assam	31205576
15	Punjab	27743338
16	Chhattisgarh	25545198
17	Haryana	25351462
18	NCT of Delhi	16787941
19	Jammu and Kashmir	12267032
20	Uttarakhand	10086292
21	Himachal Pradesh	6864602
22	Tripura	3673917
23	Meghalaya	2966889
24	Manipur[c]	2570390
25	Nagaland	1978502
26	Goa	1458545
27	Arunachal Pradesh	1383727
28	Puducherry	1247953
29	Mizoram	1097206
30	Chandigarh	1055450
31	Sikkim	610577
32	Dadra and Nagar Haveli and Daman and Diu	585764
33	Andaman and Nicobar Islands	380581

34	Ladakh	274000
35	Lakshadweep	64473

2. https://en.wikipedia.org/wiki/List_of_states_and_union_territories_of_India_by_area

	State/UT	Area (km2)
0	RAJASTHAN	342239
1	MADHYA PRADESH	308245
2	MAHARASHTRA	307713
3	UTTAR PRADESH	240928
4	GUJARAT	196024
5	KARNATAKA	191792
6	ANDHRA PRADESH	162975
7	ODISHA	155707
8	CHHATTISGARH	135191
9	TAMIL NADU	130058
10	TELANGANA	112077
11	BIHAR	94163
12	WEST BENGAL	88752
13	ARUNACHAL PRADESH	83743
14	JHARKHAND	79716
15	ASSAM	78438
16	LADAKH	59146
17	HIMACHAL PRADESH	55673
18	UTTARAKHAND	53483
19	PUNJAB	50362
20	HARYANA	44212
21	JAMMU AND KASHMIR	42241
22	KERALA	38863
23	MEGHALAYA	22429
24	MANIPUR	22327
25	MIZORAM	21081
26	NAGALAND	16579
27	TRIPURA	10486
28	ANDAMAN AND NICOBAR ISLANDS	8249
29	SIKKIM	7096
30	GOA	3702
31	DELHI	1484
32	DADRA AND NAGAR HAVELI AND DAMAN AND DIU	603

3. https://www.census2011.co.in/literacy.php

	State/UT	2001	2011
0	A&N ISLANDS	81.30	86.27
1	ANDHRA PRADESH	60.47	67.66
2	ARUNACHAL PRADESH	54.34	66.95
3	ASSAM	63.25	73.45
4	BIHAR	47.00	61.82
5	CHANDIGARH	81.94	86.43
6	CHHATTISGARH	64.66	71.04
7	DADRA AND NAGAR HAVELI	57.63	77.65
8	DAMAN AND DIU	78.18	87.07
9	DELHI	81.67	86.34
10	GOA	82.01	87.40
11	GUJARAT	69.14	79.31
12	HARYANA	67.91	76.64
13	HIMACHAL PRADESH	76.48	83.78
14	JAMMU AND KASHMIR	55.52	68.74
15	JHARKHAND	53.56	67.63
16	KARNATAKA	66.06	75.60
17	KERALA	90.86	93.91
18	LAKSHADWEEP	86.66	92.28
19	MADHYA PRADESH	63.74	70.63
20	MAHARASHTRA	76.84	82.34
21	MANIPUR	70.50	79.85
22	MEGHALYA	62.56	75.48
23	MIZORAM	88.80	91.58
24	NAGALAND	66.59	80.11
25	ODISHA	63.08	73.45
26	PUDUCHERRY	81.24	86.55
27	PUNJAB	69.65	76.68
28	RAJASTHAN	60.41	67.06
29	SIKKIM	68.81	82.20
30	TAMIL NADU	73.45	80.33
31	TRIPURA	73.19	87.75
32	UTTAR PRADESH	56.27	69.72
33	UTTARAKHAND	71.62	79.63
34	WEST BENGAL	68.64	77.08

ASSUMPTIONS: Since population census happens after 10 years thus data was available for years 2001 and 2011. For the rest of the years, we have assumed that

population is increasing at a constant rate so I have interpolated the population and literacy rate for rest of the years.

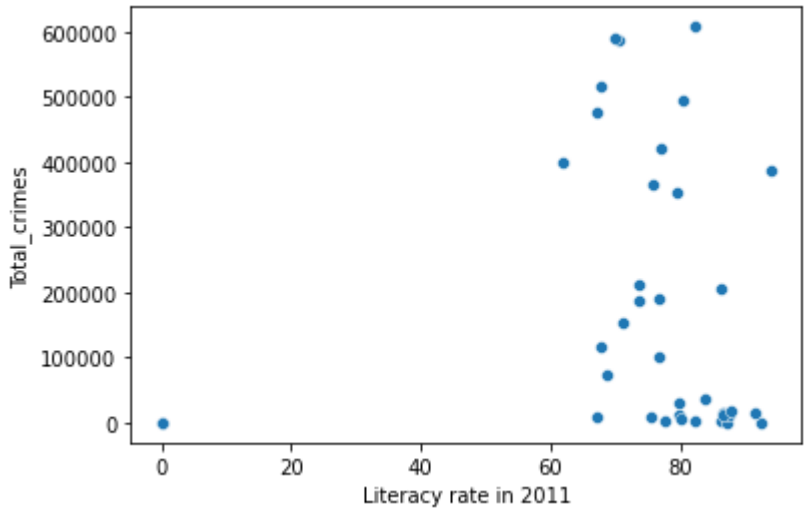
Phase 2

2.1 Analysis of Literacy rate vs Total crimes.

	State/UT	Literacy rate in 2011	Total crimes
0	A&N ISLANDS	86.27	2982.0
1	ANDHRA PRADESH	67.66	515853.0
2	ARUNACHAL PRADESH	66.95	8339.0
3	ASSAM	73.45	210809.0
4	BIHAR	61.82	399281.0
5	CHANDIGARH	86.43	14785.0
6	CHHATTISGARH	71.04	153886.0
7	DADRA AND NAGAR HAVELI	77.65	3134.0
8	DAMAN AND DIU	87.07	1023.0
9	DELHI	86.34	205547.0
10	GOA	87.40	11191.0
11	GUJARAT	79.31	353387.0
12	HARYANA	76.64	189336.0
13	HIMACHAL PRADESH	83.78	36007.0
14	JAMMU AND KASHMIR	68.74	73821.0
15	JHARKHAND	67.63	117426.0
16	KARNATAKA	75.60	364521.0
17	KERALA	93.91	388068.0
18	LADAKH	0.00	0.0
19	LAKSHADWEEP	92.28	398.0
20	MADHYA PRADESH	70.63	587540.0
21	MAHARASHTRA	82.34	608358.0
22	MANIPUR	79.85	11391.0
23	MEGHALAYA	75.48	9686.0
24	MIZORAM	91.58	13746.0
25	NAGALAND	80.11	5203.0
26	ODISHA	73.45	186593.0
27	PUDUCHERRY	86.55	13136.0
28	PUNJAB	76.68	100452.0
29	RAJASTHAN	67.06	477129.0
30	SIKKIM	82.20	2705.0

31	TAMIL NADU	80.33	494098.0
32	TELANGANA	0.00	0.0
33	TRIPURA	87.75	18586.0
34	UTTAR PRADESH	69.72	589289.0
35	UTTARAKHAND	79.63	30264.0
36	WEST BENGAL	77.08	421566.0

Scatterplot between total crimes till 2011 and literacy rate at 2011:



2.2 Analysis of the type of crime vs each state vs Literacy rate.

```
RIVER and SEA - Dacoity                          0.148196
BANKS - Burglary                                 0.141823
Value_of_Property_Stolen                         0.115470
Police_Injured_In_Accidents                      0.110396
Girls 7-12 Years                                 0.103398
Total Crime                                      0.097172
COUNTERFIETING                                   0.088059
BANKS - Theft                                    0.085877
PREPARATION AND ASSEMBLY FOR DACOITY             0.084512
Police_Injured_By_Criminals                      0.084175
Girls 12-16 Years                                0.082708
OTHER IPC CRIMES                                 0.077816
Cases_Property_Recovered                         0.073105
COMMERCIAL ESTABLISHMENTS - Theft                0.072019
COMMERCIAL ESTABLISHMENTS - Robbery              0.070953
COMMERCIAL ESTABLISHMENTS - Burglary             0.070155
```

OTHER PLACES - Burglary	0.068956
OTHER PLACES - Theft	0.065375
OTHER PLACES - Robbery	0.063667
Police_Killed_In_Accidents	0.058928
Dacoity_y	0.058658
Total for girls all Age Groups	0.057475
Exposure and abandonment	0.057444
Robbery_y	0.056093
COMMERCIAL ESTABLISHMENTS - Dacoity	0.055288
Cases_Property_Stolen	0.053591
Police_Injured_In_Dacoity_OperationsOther_raids	0.053132
Buying of girls for prostitution	0.052988
BANKS - Robbery	0.051642
Boys 7-12 Years	0.048071
Girls 16-18 Years	0.045790
OTHER PLACES - Dacoity	0.044143
TOTAL IPC CRIMES	0.038677
THEFT	0.037706
TOTAL - Theft	0.037706
RAILWAYS - Robbery	0.037460
Prohibition of child marriage act	0.037445
TOTAL - Burglary	0.036235
BURGLARY	0.036235
OTHER THEFT	0.035481
Rape_x.1	0.031567
AUTO THEFT	0.031399
RESIDENTIAL PREMISES - Theft	0.030786
Grand total	0.029830
Police_Killed_Total_Policemen	0.029781
RIVER and SEA - Theft	0.029442
Total for boys all Age Groups	0.028950
Dacoity_x	0.028923
Other Crimes	0.028121
RIOTS	0.026928
TOTAL - Robbery	0.026408
ROBBERY	0.026408
Boys 16-18 Years	0.026199
Total	0.023381
ASSAULT ON WOMEN WITH INTENT TO OUTRAGE HER MODESTY	0.023052
Assault on women with intent to outrage her modesty	0.023052
Boys 12-16 Years	0.021952
HIGHWAYS - Burglary	0.017142
RESIDENTIAL PREMISES - Burglary	0.016716
ARSON	0.016077
Police_Injured_Total_Policemen	0.016066
Rape_y	0.012616
RESIDENTIAL PREMISES - Robbery	0.012315
Cruelty by Husband or his Relatives	0.005901
CRUELTY BY HUSBAND OR HIS RELATIVES	0.005901

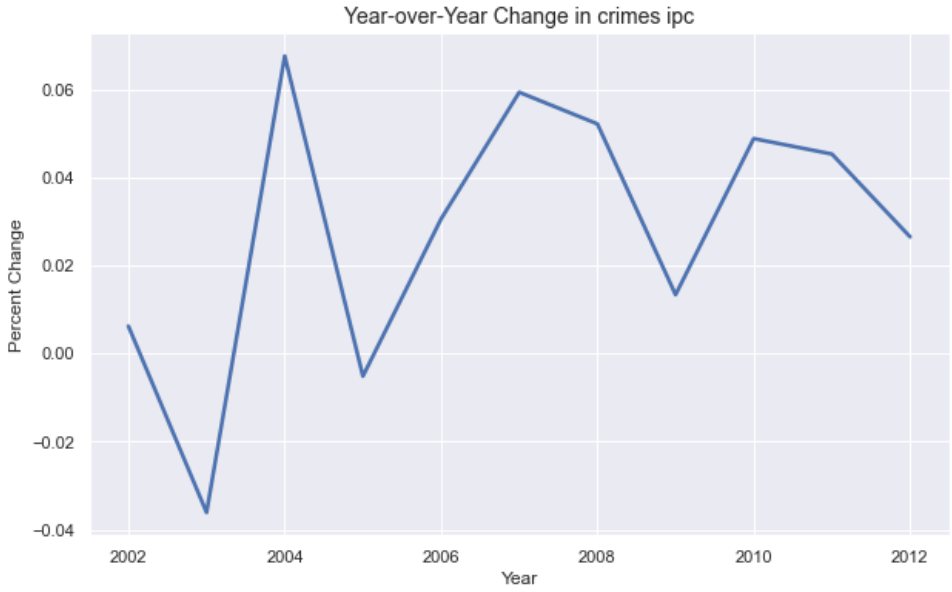
Kidnapping and Abduction_y	0.005405
CULPABLE HOMICIDE NOT AMOUNTING TO MURDER	0.002939
RAILWAYS - Theft	0.002403
Police_Injured_By_Riotous_Mobs	0.001793
Police_Injured_On_Border_Duties	0.000634
Foeticide	-0.000015
Selling of girls for prostitution	-0.000947
Police_Killed_By_Riotous_Mobs	-0.001305
Value_of_Property_Recovered	-0.001540
Murder	-0.008130
Procuration of minor girls	-0.009038
OTHER RAPE	-0.012353
RAPE	-0.012354
Rape_y.1	-0.012354
CHEATING	-0.015938
CAUSING DEATH BY NEGLIGENCE	-0.017373
RAILWAYS - Burglary	-0.018158
Abetment of suicide	-0.018957
Murder_y	-0.019760
CRIMINAL BREACH OF TRUST	-0.020772
INSULT TO MODESTY OF WOMEN	-0.021343
Insult to modesty of Women	-0.021343
Police_Killed_In_TerroristsExtremists_Operations	-0.022509
Police_Injured_In_TerroristsExtremists_Operations	-0.028141
KIDNAPPING AND ABDUCTION OF OTHERS	-0.029580
Police_Killed_By_Criminals	-0.030304
KIDNAPPING & ABDUCTION	-0.033678
HURT/GREVIOUS HURT	-0.033685
Police_Killed_In_Dacoity_OperationsOther_raids	-0.034106
CUSTODIAL RAPE	-0.034135
Robbery_x	-0.035263
Cases_Registered_under_Human_Rights_Violations	-0.035331
HIGHWAYS - Theft	-0.035962
Policemen_Convicted	-0.040084
Rape_x	-0.041201
BANKS - Dacoity	-0.042580
Police_Killed_On_Border_Duties	-0.045879
KIDNAPPING AND ABDUCTION OF WOMEN AND GIRLS	-0.046410
Kidnapping and Abduction	-0.046410
Protection of Civil Rights (PCR) Act_x	-0.046689
Policemen_Chargesheeted	-0.047465
DACOITY	-0.048988
TOTAL - Dacoity	-0.048988
Kidnapping and Abduction_x	-0.051125
Kidnapping Abduction	-0.055584
Murder_x	-0.058719
RIVER and SEA - Burglary	-0.059360
Arson_y	-0.060198
MURDER	-0.060524

```
Hurt_y                                                -0.064325
Other Crimes Against STs                              -0.065175
Hurt_x                                                -0.069709
Protection of Civil Rights (PCR) Act_y                -0.071678
RIVER and SEA - Robbery                               -0.071725
ATTEMPT TO MURDER                                     -0.072680
HIGHWAYS - Robbery                                    -0.078195
Other Crimes Against SCs                              -0.079292
HIGHWAYS - Dacoity                                    -0.080259
Arson_x                                               -0.082478
IMPORTATION OF GIRLS FROM FOREIGN COUNTRIES           -0.088260
Prevention of atrocities (POA) Act_y                  -0.088559
RESIDENTIAL PREMISES - Dacoity                        -0.090891
DOWRY DEATHS                                          -0.093159
Dowry Deaths                                          -0.093159
RAILWAYS - Dacoity                                    -0.111205
Prevention of atrocities (POA) Act_x                  -0.114355
```

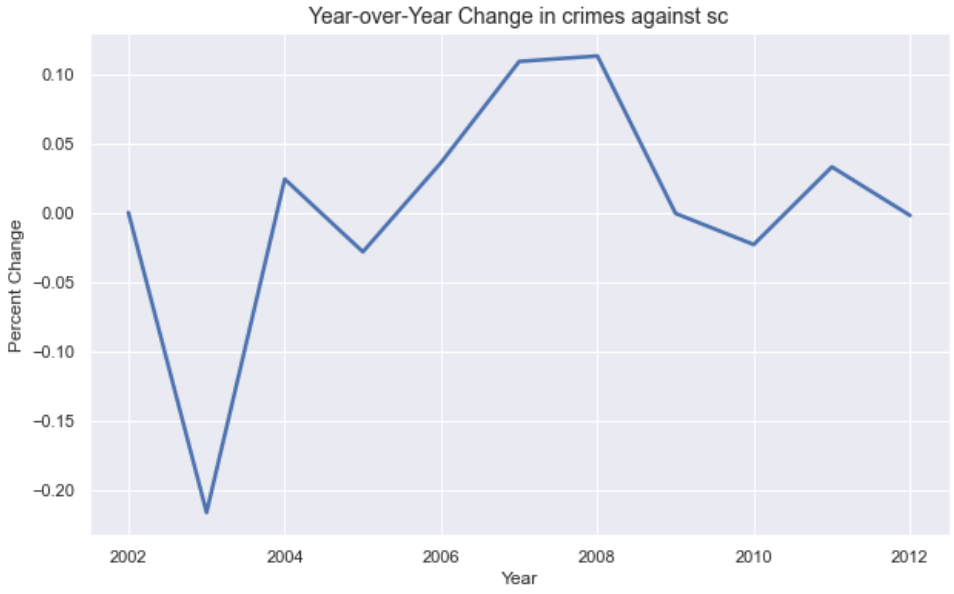
Above, I calculate the correlation between each crime type and literacy rate using the corr() method on the crime_literacy_by_state dataframe and the ['Literacy Rate'] column of the resulting correlation matrix. This will give us a measure of the linear relationship between each crime type and literacy rate, with values closer to -1 indicating a negative correlation, values closer to 1 indicating a positive correlation, and values close to 0 indicating no correlation.

2.3 Analysis of year-on-year total crime rate

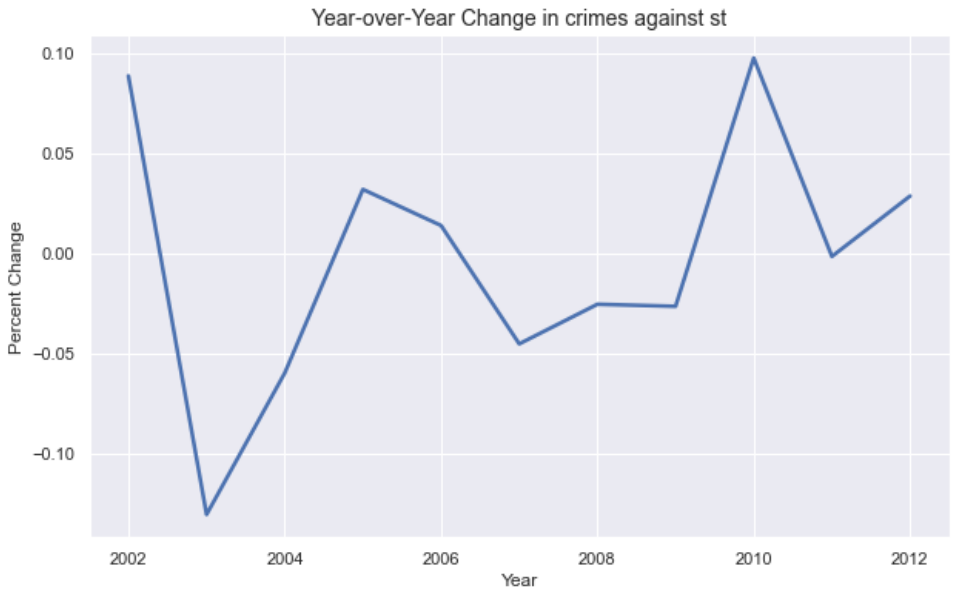
	yearly_total_crimes ipc	yoy_change ipc
YEAR		
2001	3538616	NaN
2002	3560660	0.006230
2003	3432240	-0.036066
2004	3664020	0.067530
2005	3645204	-0.005135
2006	3756586	0.030556
2007	3979346	0.059299
2008	4186758	0.052122
2009	4242690	0.013359
2010	4449662	0.048783
2011	4651150	0.045282
2012	4774376	0.026494



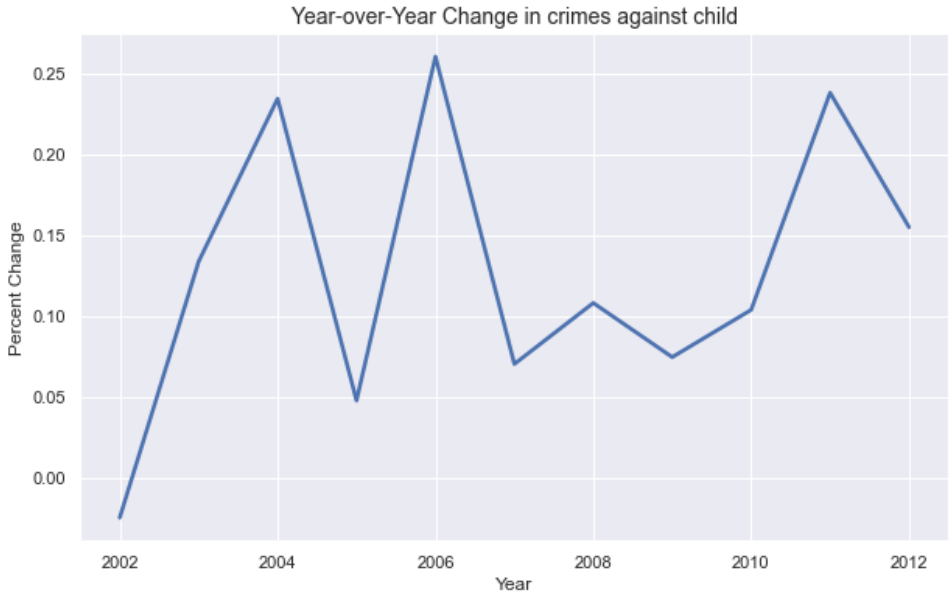
	yearly_total_crimes against sc	yoy_change against sc
Year		
2001	67003	NaN
2002	67014	0.000164
2003	52502	-0.216552
2004	53775	0.024247
2005	52254	-0.028285
2006	54146	0.036208
2007	60062	0.109260
2008	66860	0.113183
2009	66824	-0.000538
2010	65286	-0.023016
2011	67446	0.033085
2012	67310	-0.002016



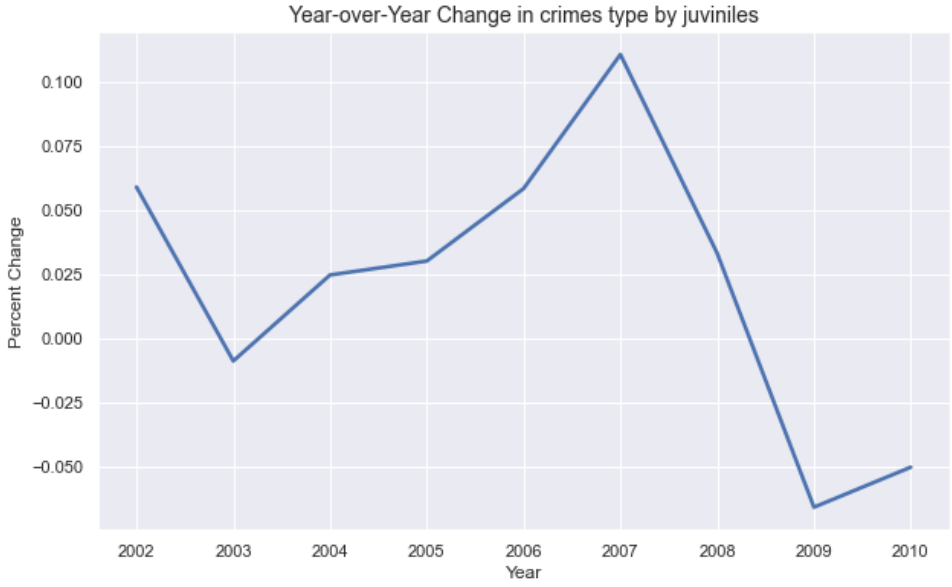
	yearly_total_crimes against st	yoy_change against st
Year		
2001	12434	NaN
2002	13540	0.088950
2003	11773	-0.130502
2004	11070	-0.059713
2005	11426	0.032159
2006	11587	0.014091
2007	11064	-0.045137
2008	10784	-0.025307
2009	10500	-0.026335
2010	11528	0.097905
2011	11512	-0.001388
2012	11844	0.028839



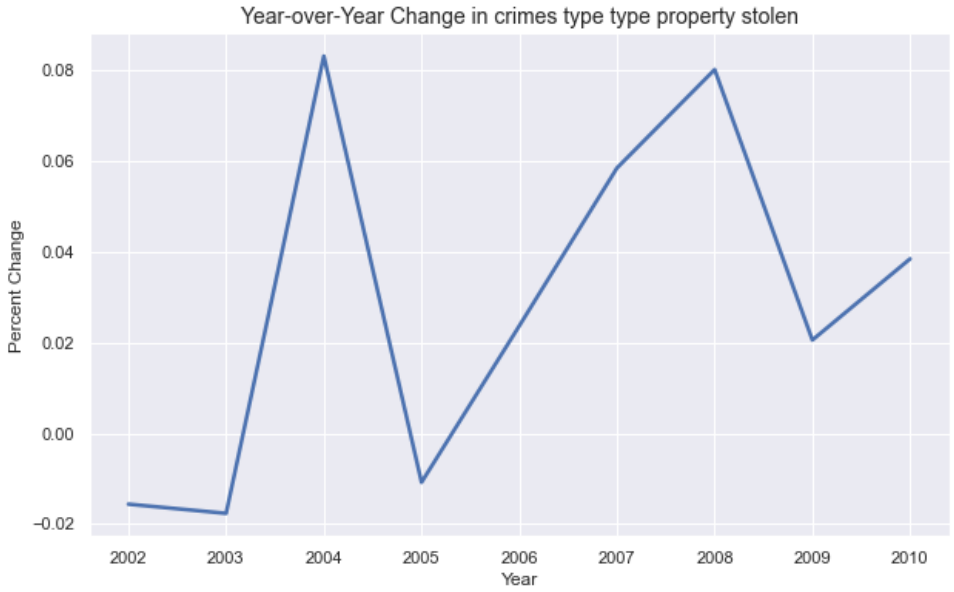
	yearly_total_crimes against children	yoy_change against child
Year		
2001	21035	NaN
2002	20520	-0.024483
2003	23266	0.133821
2004	28721	0.234462
2005	30096	0.047874
2006	37934	0.260433
2007	40607	0.070464
2008	45000	0.108183
2009	48362	0.074711
2010	53388	0.103925
2011	66098	0.238068
2012	76344	0.155012



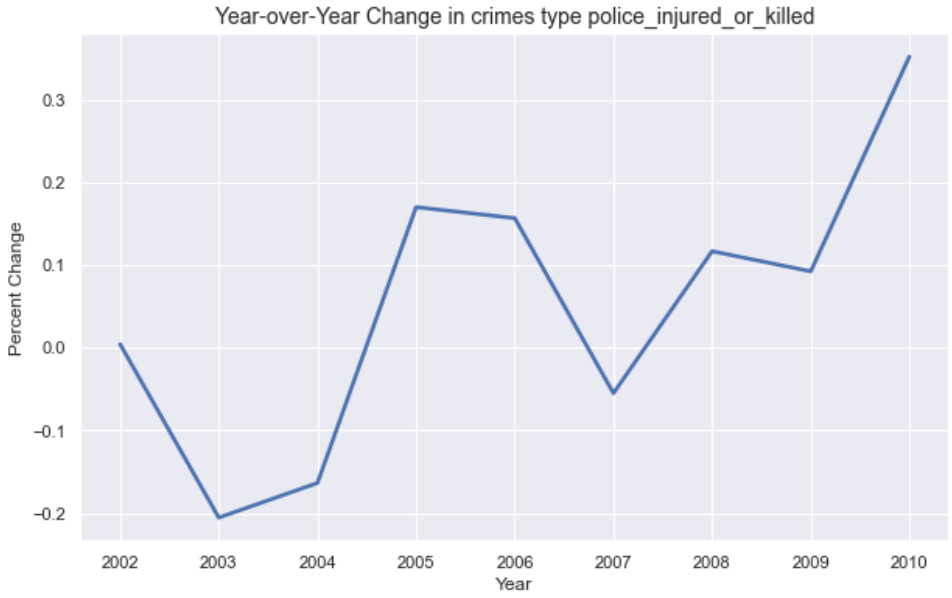
	yearly_total_crimes by juviniles	yoy_change in crimes by juviniles
Year		
2001	52314	NaN
2002	55411	0.059200
2003	54926	-0.008753
2004	56291	0.024852
2005	57997	0.030307
2006	61397	0.058624
2007	68212	0.110999
2008	70486	0.033337
2009	65841	-0.065900
2010	62532	-0.050257



	yearly_total_crimes type property stolen	yoy_change in crimes type property stolen
Year		
2001	870858	NaN
2002	857257	-0.015618
2003	842126	-0.017650
2004	912123	0.083119
2005	902267	-0.010806
2006	923607	0.023652
2007	977597	0.058456
2008	1055947	0.080145
2009	1077643	0.020546
2010	1119066	0.038439

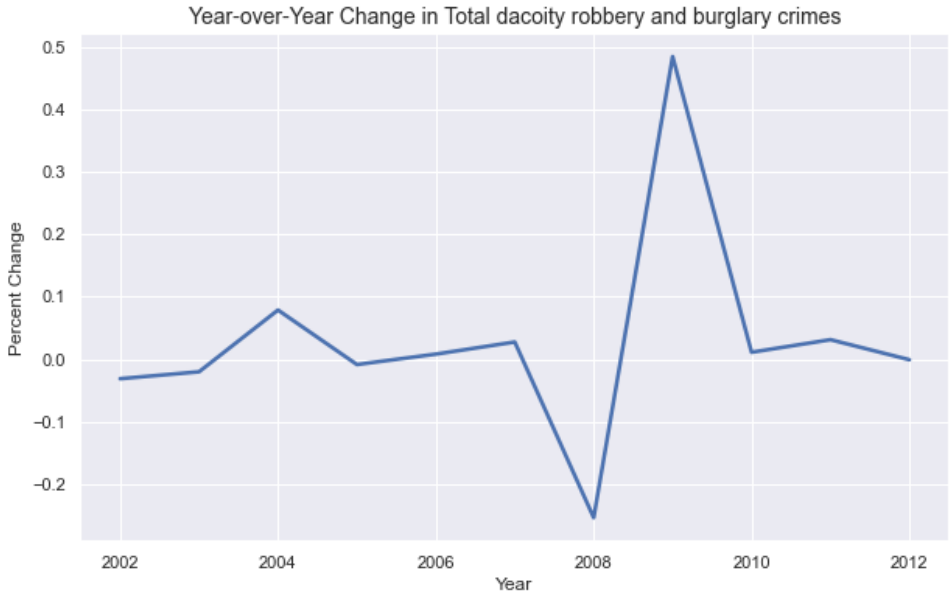


	yearly_total_crimes type police_injured_or_killed	yoy_change in crimes type police_injured_or_killed
Year		
2001	19056	NaN
2002	19144	0.004618
2003	15208	-0.205600
2004	12720	-0.163598
2005	14892	0.170755
2006	17236	0.157400
2007	16292	-0.054769
2008	18204	0.117358
2009	19896	0.092947
2010	26924	0.353237



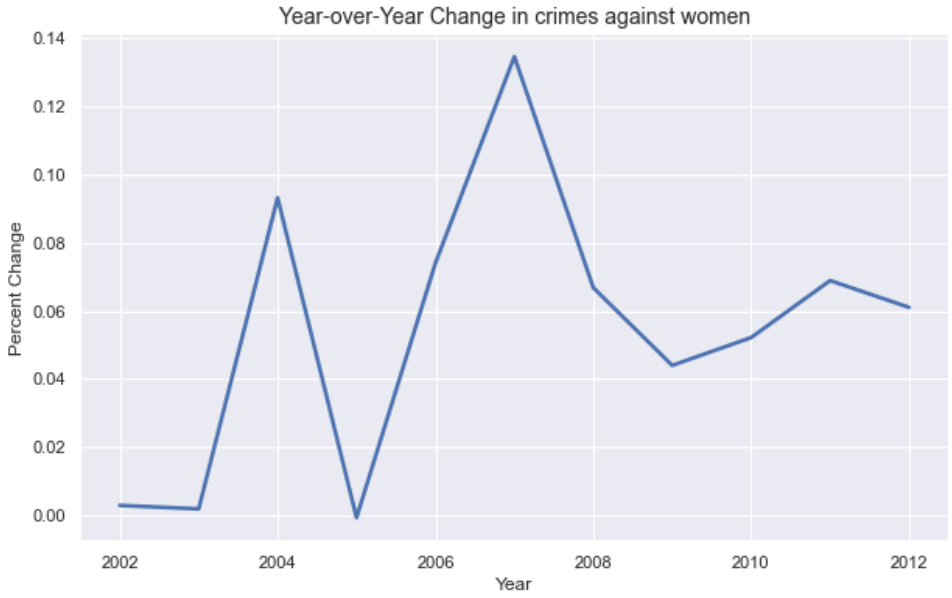
	yearly_total_crimes type property taken away	yoy_change in crimes type property taken away
YEAR		
2001	30529027487	NaN
2002	31728702995	0.039296
2003	27867903578	-0.121682
2004	24360185139	-0.125869
2005	37599757603	0.543492
2006	30175234903	-0.197462
2007	34402079395	0.140077
2008	48329106062	0.404831
2009	45449399545	-0.059585
2010	68579275272	0.508915
2011	60956785873	-0.111149
2012	85182301702	0.397421

	yearly_total_crimes type Total dacoity robbery and burglary	yoy_change in crimes type Total dacoity robbery and burglary
YEAR		
2001	1140120	NaN
2002	1104723	-0.031047
2003	1082637	-0.019992
2004	1167912	0.078766
2005	1158099	-0.008402
2006	1167669	0.008264
2007	1199928	0.027627
2008	896053	-0.253244
2009	1329780	0.484042
2010	1344726	0.011239
2011	1386867	0.031338
2012	1385868	-0.000720



	yearly_total_crimes type Total human rights violation crimes	yoy_change in crimes type Total human rights violation crimes
Year		
2001	640.0	NaN
2002	308.0	-0.518750
2003	236.0	-0.233766
2004	132.0	-0.440678
2005	110.0	-0.166667
2006	114.0	0.036364
2007	202.0	0.771930
2008	550.0	1.722772
2009	664.0	0.207273
2010	110.0	-0.834337

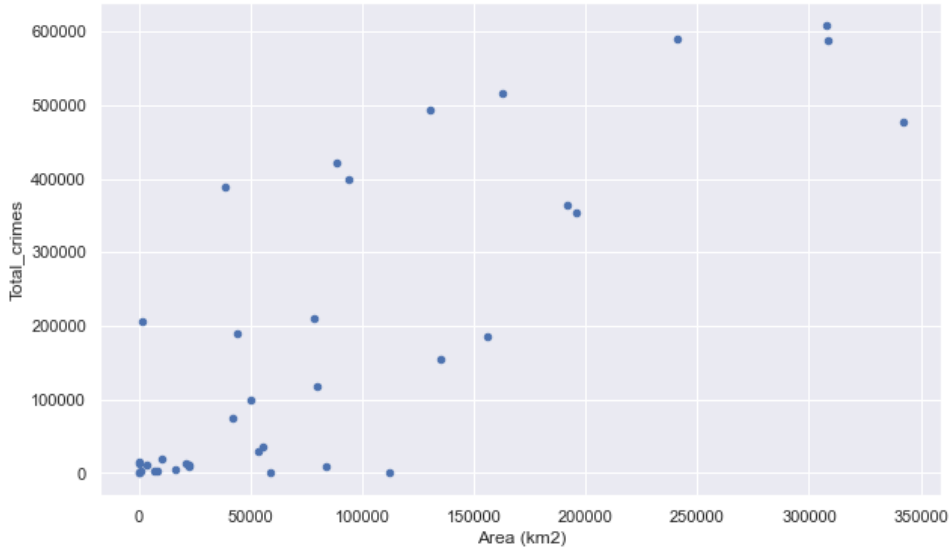
	yearly_total_crimes type Total human rights violation crimes	yoy_change in crimes type Total human rights violation crimes
Year		
2001	261450	NaN
2002	262224	0.002960
2003	262728	0.001922
2004	287230	0.093260
2005	287046	-0.000641
2006	308316	0.074100
2007	349842	0.134686
2008	373232	0.066859
2009	389670	0.044042
2010	410018	0.052219
2011	438284	0.068938
2012	465056	0.061084



2.4 Analysis of area vs overall crime

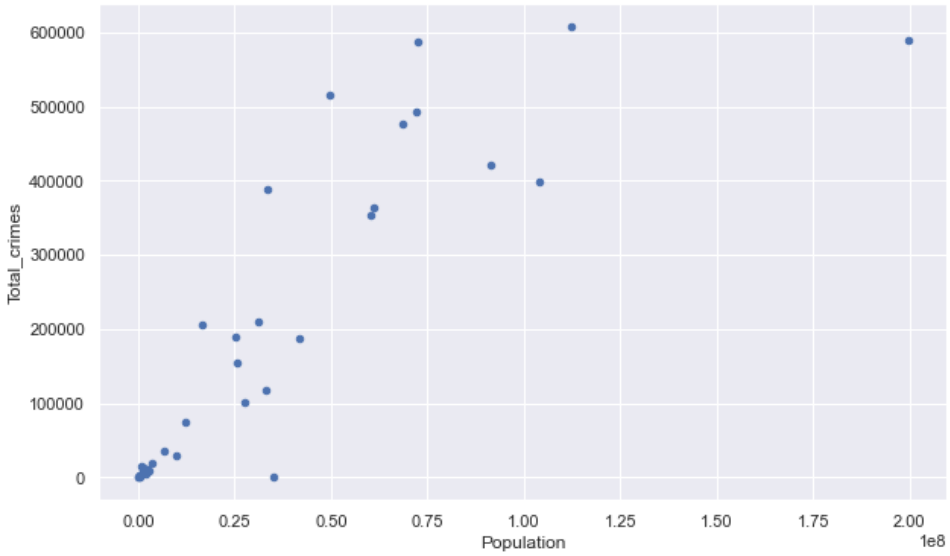
	State/UT	Population	Area (km2)	Total_crimes	Crime Rate
0	A&N ISLANDS	380581.0	8249.0	2982.0	0.361498
1	ANDHRA PRADESH	49577103.0	162975.0	515853.0	3.165228
2	ARUNACHAL PRADESH	1383727.0	83743.0	8339.0	0.099578
3	ASSAM	31205576.0	78438.0	210809.0	2.687588
4	BIHAR	104099452.0	94163.0	399281.0	4.240317
5	CHANDIGARH	1055450.0	114.0	14785.0	129.692982
6	CHHATTISGARH	25545198.0	135191.0	153886.0	1.138286
7	DADRA AND NAGAR HAVELI	585764.0	603.0	3134.0	5.197347
8	DAMAN AND DIU	243000.0	112.0	1023.0	9.133929
9	DELHI	16787941.0	1484.0	205547.0	138.508760
10	GOA	1458545.0	3702.0	11191.0	3.022961
11	GUJARAT	60439692.0	196024.0	353387.0	1.802774
12	HARYANA	25351462.0	44212.0	189336.0	4.282457
13	HIMACHAL PRADESH	6864602.0	55673.0	36007.0	0.646759
14	JAMMU AND KASHMIR	12267032.0	42241.0	73821.0	1.747615
15	JHARKHAND	32988134.0	79716.0	117426.0	1.473054
16	KARNATAKA	61095297.0	191792.0	364521.0	1.900606
17	KERALA	33406061.0	38863.0	388068.0	9.985539

18	LADAKH	274000.0	59146.0	0.0	0.000000
19	LAKSHADWEEP	64473.0	32.0	398.0	12.437500
20	MADHYA PRADESH	72626809.0	308245.0	587540.0	1.906081
21	MAHARASHTRA	112374333.0	307713.0	608358.0	1.977031
22	MANIPUR	2570390.0	22327.0	11391.0	0.510189
23	MEGHALAYA	2966889.0	22429.0	9686.0	0.431852
24	MIZORAM	1097206.0	21081.0	13746.0	0.652056
25	NAGALAND	1978502.0	16579.0	5203.0	0.313831
26	ODISHA	41974219.0	155707.0	186593.0	1.198360
27	PUDUCHERRY	1247953.0	479.0	13136.0	27.423800
28	PUNJAB	27743338.0	50362.0	100452.0	1.994599
29	RAJASTHAN	68548437.0	342239.0	477129.0	1.394140
30	SIKKIM	610577.0	7096.0	2705.0	0.381201
31	TAMIL NADU	72147030.0	130058.0	494098.0	3.799059
32	TELANGANA	35003674.0	112077.0	0.0	0.000000
33	TRIPURA	3673917.0	10486.0	18586.0	1.772459
34	UTTAR PRADESH	199812341.0	240928.0	589289.0	2.445913
35	UTTARAKHAND	10086292.0	53483.0	30264.0	0.565862
36	WEST BENGAL	91276115.0	88752.0	421566.0	4.749932



Area has positive relationship with total crimes.

2.5 Analysis of population vs overall crime



Population also has direct relationship with total crimes.

2.6 Each state crime report

(Tabulated in ipynb file)

Phase 3

Analysis Summary

Highest number of rapes & Kidnappings that happened in ('WEST BENGAL', 'MURSHIDABAD', 2009, 568, 342)

The lowest number of rapes & Kidnappings that happened in ('A & N ISLANDS', 'NICOBAR', 2001, 0, 0)

The highest number of dacoity/robbery in which district ('GUJARAT', 'DAHOD', 29 and 32).

Districts (All) in which lowest number of murders happened is ('CHHATTISGARH', 'KONDAGAON', 1)

Critical districts among all the states are: CYBERABAD, LOHIT, DIBRUGARH, TINSUKIA, PATNA, NORTH GOA, AHMEDABAD COMMR, SURAT COMMR,

SONIPAT nad GURGAON, KANGRA, BARAMULLA, RANCHI, BANGALORE COMMR, JHABUA in MP, MUMBAI, MUMBAI COMMR, IMPHAL(WEST), JAINTIA HILLS, AIZAWL, DIMAPUR, LUDHIANA, JALANDHAR, BHARATPUR, SIKKIM EAST, CHENNAI, TRIPURA WEST, MUZAFFARNAGAR, HARIDWAR, 24 PARGANAS NORTH, ANDAMAN, CHANDIGARH AND DELHI- NORTH WEST.

Amongst all the IPC crimes occurred between 2001-2012, Delhi tops the list with 565 IPC crimes in 2010.

State	District	Year	Value
DELHI UT	DELHI UT TOTAL	2010	565
BIHAR	PATNA	2004	542
UTTAR PRADESH	MUZAFFARNAGAR	2001	324
MAHARASHTRA	MUMBAI	2001	295
JHARKHAND	RANCHI	2002	270
WEST BENGAL	PASCHIM MIDNAPUR	2010	268
KARNATAKA	BANGALORE COMMR.	2010	266
GUJARAT	AHMEDABAD COMMR.	2002	222
ANDHRA PRADESH	RANGA REDDY	2001	214
TAMIL NADU	CHENNAI	2012	180
JAMMU & KASHMIR	BARAMULLA	2001	144
ASSAM	TINSUKIA	2007	143
MADHYA PRADESH	JHABUA	2004	139
CHHATTISGARH	BIZAPUR	2006	139
TRIPURA	WEST	2003	136
ODISHA	KEONJHAR	2009	109
RAJASTHAN	ALWAR	2011	104
UTTARAKHAND	HARIDWAR	2001	94
HARYANA	GURGAON	2006	93
PUNJAB	CP LUDHIANA	2010	90
NAGALAND	DIMAPUR	2008	64
KERALA	TRIVANDRUM	2002	58
MEGHALAYA	JAINTIA HILLS	2003	52
HIMACHAL PRADESH	KANGRA	2003	38
GOA	NORTH GOA	2009	33
PUDUCHERRY	PUDUCHERRY	2011	32
CHANDIGARH	CHANDIGARH	2002	31
MIZORAM	AIZAWL	2010	23
ARUNACHAL PRADESH	LOHIT	2007	17
A & N ISLANDS	ANDAMAN	2003	16
D & N HAVELI	D and N HAVELI	2011	14

SIKKIM	EAST	2001	13
DAMAN & DIU	DAMAN	2007	12
LAKSHADWEEP	LAKSHADWEEP	2001	1

Phase 4

K Means clustering algorithm was used to cluster states into Sensitive, Moderate and Peaceful areas which are shared below:

Below is a first cluster of Sensitive areas

State/UT	Population	Area (km2)	Literacy rate in 2011	Total_crimes	cluster_name
TAMIL NADU	72147030.0	130058.0	80.33	494098.0	Sensitive Area
RAJASTHAN	68548437.0	342239.0	67.06	477129.0	Sensitive Area
WEST BENGAL	91276115.0	88752.0	77.08	421566.0	Sensitive Area
BIHAR	104099452.0	94163.0	61.82	399281.0	Sensitive Area
KERALA	33406061.0	38863.0	93.91	388068.0	Sensitive Area
KARNATAKA	61095297.0	191792.0	75.60	364521.0	Sensitive Area
GUJARAT	60439692.0	196024.0	79.31	353387.0	Sensitive Area
ASSAM	31205576.0	78438.0	73.45	210809.0	Sensitive Area
DELHI	16787941.0	1484.0	86.34	205547.0	Sensitive Area
HARYANA	25351462.0	44212.0	76.64	189336.0	Sensitive Area
ODISHA	41974219.0	155707.0	73.45	186593.0	Sensitive Area
CHHATTISGARH	25545198.0	135191.0	71.04	153886.0	Sensitive Area
JHARKHAND	32988134.0	79716.0	67.63	117426.0	Sensitive Area

Backward areas having large population and large area with less literacy rate are grouped into sensitive cluster.

Below is a 2nd cluster of Moderate areas:

State/UT	Population	Area (km2)	LR in 2011	Total_crimes	cluster_name
MAHARASHTRA	112374333.0	307713.0	82.34	608358.0	Moderate Area
UTTAR PRADESH	199812341.0	240928.0	69.72	589289.0	Moderate Area
MADHYA PRADESH	72626809.0	308245.0	70.63	587540.0	Moderate Area
ANDHRA PRADESH	49577103.0	162975.0	67.66	515853.0	Moderate Area

States grouped into moderate cluster are highly prone to more crimes and becoming sensitive areas. These can be considered as outliers to as all states included in moderate cluster have large areas, densely populated with less literacy rate,

Below is a 3rd cluster of Peaceful areas:

State/UT	Population	Area (km2)	2011	Total_crimes	cluster_name

PUNJAB	27743338.0	50362.0	76.68	100452.0	Peaceful Area
JAMMU AND KASHMIR	12267032.0	42241.0	68.74	73821.0	Peaceful Area
HIMACHAL PRADESH	6864602.0	55673.0	83.78	36007.0	Peaceful Area
UTTARAKHAND	10086292.0	53483.0	79.63	30264.0	Peaceful Area
TRIPURA	3673917.0	10486.0	87.75	18586.0	Peaceful Area
CHANDIGARH	1055450.0	0.0	86.43	14785.0	Peaceful Area
MIZORAM	1097206.0	21081.0	91.58	13746.0	Peaceful Area
PUDUCHERRY	1247953.0	479.0	86.55	13136.0	Peaceful Area
MANIPUR	2570390.0	22327.0	79.85	11391.0	Peaceful Area
GOA	1458545.0	3702.0	87.40	11191.0	Peaceful Area
MEGHALAYA	2966889.0	22429.0	75.48	9686.0	Peaceful Area
ARUNACHAL PRADESH	1383727.0	83743.0	66.95	8339.0	Peaceful Area
NAGALAND	1978502.0	16579.0	80.11	5203.0	Peaceful Area
DADRA AND NAGAR HAVELI	585764.0	603.0	77.65	3134.0	Peaceful Area
A&N ISLANDS	380581.0	8249.0	86.27	2982.0	Peaceful Area
SIKKIM	610577.0	7096.0	82.20	2705.0	Peaceful Area
DAMAN AND DIU	-	0.0	87.07	1023.0	Peaceful Area
LAKSHADWEEP	64473.0	0.0	92.28	398.0	Peaceful Area

North and north eastern states with Indian islands are grouped into peaceful cluster contributing to the fact that these have quite less areas.

Concluding Remarks

As the population is increasing demand and supply gap, lack of education, poverty, and many more reasons are the main causes of increasing crime rates and casualties. For sensitive and moderate clusters, immediate steps need to be taken by the government to curb population growth, improve education faccilaties and literacy rate.

Furthur analysis can be used to target highly prone districts throught the country.