DataTrained
Keep Skilling, Keep Growing

Capstone Project

Crimes In India

Submitted by: RANJAN PRATAP SINGH

ACKNOWLEDGMENT

I express my deepest gratitude to my Instructor, Mr. *Shankargouda Tegginmani*, and Data Trained Academy for providing me with the invaluable opportunity to undertake this project on ***CRIMES IN INDIA***. This journey has been enriching, allowing me to engage in extensive research and gain insights into various aspects, especially in the realm of data collection. In the section on data sources, I have acknowledged and detailed the external resources, including the mentioned contributors. This project has been a profound learning experience, shaping my analytical skills and expanding my knowledge base. All the external resources that were used in creating this project are listed below:

- https://www.prb.org/wp-content/uploads/2011/04/india-population-2001-2011.pdf
- https://ncert.nic.in/textbook/pdf/legy2a1.pdf
- https://www.educationforallinindia.com/page159.html
- https://www.census2011.co.in/states.php
- https://www.kaggle.com/datasets/rajanand/crime-in-india?resource=download

Introduction

This capstone project focuses on analyzing crimes in India, employing diverse datasets to derive insights. The analysis spans demographic indicators, state-wise crime patterns, SQL operations, and unsupervised machine learning clustering. The project aims to unveil correlations, trends, and potential recommendations for enhancing public safety.

Phase 1: Data Collection/Preparation

- Gathered population, literacy rate, and area data for each state.
- Collected additional relevant information for analysis.
- Created a new file containing the collected data.

Phase 2: State/UT Wise Analysis

- Explored the correlation between literacy rates and total crimes.
- Analyzed the relationship between crime types and literacy rates across states.
- Investigated year-on-year total crime rates.
- Explored the relationship between crime rates and state area or population.
- Examined crime reports for each state, allowing for subjective analysis.

Phase 3: SQL Operations

- Created separate tables for distinct crime datasets.
- Executed SQL queries to identify highest and lowest crime incidents.
- Inserted records into new tables for further analysis.
- Leveraged SQL for insightful queries on crime data.

Phase 4: Unsupervised ML (Clustering)

- Created three clusters categorizing areas as Sensitive, Moderate, and Peaceful.
- Formed DataFrames for each cluster, facilitating focused analysis.
- Conducted a detailed analysis of clusters, identifying factors impacting crime, suggesting measures for improvement, and highlighting safe and unsafe districts.

The project unveils a comprehensive exploration of crime in India, from demographic indicators to unsupervised machine learning clustering. Insights gained from each phase offer a nuanced understanding of crime patterns, enabling informed recommendations for public safety enhancement.

PHASE I & II

In the initial phase of the project, we diligently gathered census data for the years 2001 and 2011, leveraging ChatGPT for subsequent year predictions. Additionally, we meticulously sourced crime reports spanning 2001 to 2012, covering various offenses across all states. The conclusive report in Phase 4 includes a comprehensive list of links, providing transparency about the data sources.

Our analysis unveiled a noteworthy correlation between literacy rates and overall crime rates. Across all states, a consistent trend emerged: as literacy rates increased, crime rates, including murder, attempts to murder, culpable homicide, rape, and various other offenses, tended to decrease. This finding emphasizes the significance of education in contributing to the reduction of criminal activities.

Furthermore, state-wise examinations pinpointed regions with elevated crime rates. States such as Uttar Pradesh, Bihar, West Bengal, and Maharashtra exhibited higher incidences of specific crimes. This insight provides valuable information for targeted interventions and resource allocation.

Year-on-year total crime rate analysis disclosed intriguing patterns, such as a modest increase from 2001 to 2002, a substantial decline in 2003, relative stability from 2004 to 2010, a notable increase in 2011, and a subsequent significant decrease in 2012. These fluctuations provide a nuanced understanding of temporal variations in crime trends.

Exploring the relationship between crime rates and geographical area and population size uncovered valuable insights. Larger geographical extents tended to exhibit higher total IPC crime rates, and there was a clear positive correlation between population size and total IPC crimes. These findings inform policy decisions and resource planning for effective crime management.

The state-wise crime reports provided a detailed examination of individual states, highlighting variations and consistent patterns in crime rates over the years. This information is crucial for tailored interventions and state-specific crime management strategies.

In summary, the initial phases of the project conducted a thorough exploration of demographic and crime data, revealing intricate relationships between literacy rates, geographic factors, population size, and crime rates. These insights form the foundation for subsequent phases, contributing to a holistic understanding of crime dynamics in India.

CRIME REPORT STATE OF EACH STATE

Andaman & Nicobar Is. UT

CRIME RATE: 2218.31

Percentage of CRIME RATE: 2.93%

Andaman & Nicobar Islands show a moderate crime rate of 2218.31, contributing 2.93% to the overall crime rate.

Andhra Pradesh

CRIME RATE: 2307.90

Percentage of CRIME RATE: 3.05%

Andhra Pradesh reports a CRIME RATE of 2307.90, making up 3.05% of the total crime rate.

Arunachal Pradesh

CRIME RATE: 1968.36

Percentage of CRIME RATE: 2.60%

Arunachal Pradesh exhibits a CRIME RATE of 1968.36, constituting 2.60% of the total crime rate.

Assam

CRIME RATE: 1983.11

Percentage of CRIME RATE: 2.62%

Assam reflects a CRIME RATE of 1983.11, accounting for 2.62% of the total crime rate.

Bihar

CRIME RATE: 1374.30

Percentage of CRIME RATE: 1.82%

Bihar records a CRIME RATE of 1374.30, contributing 1.82% to the overall crime rate.

Chandigarh UT

CRIME RATE: 3951.71

Percentage of CRIME RATE: 5.23%

Chandigarh UT has a notable crime rate of 3951.71, making up 5.23% of the total crime rate.

Chhattisgarh

CRIME RATE: 2277.85

Percentage of CRIME RATE: 3.01%

Chhattisgarh reports a CRIME RATE of 2277.85, contributing 3.01% to the overall crime rate.

Dadra & Nagar Haveli UT

CRIME RATE: 1817.50

Percentage of CRIME RATE: 2.40%

Dadra & Nagar Haveli UT displays a CRIME RATE of 1817.50, accounting for 2.40% of the total crime rate.

Daman & Diu UT

CRIME RATE: 1591.00

Percentage of CRIME RATE: 2.10%

Daman & Diu UT records a CRIME RATE of 1591.00, constituting 2.10% of the total crime rate.

Delhi

CRIME RATE: 3978.23

Percentage of CRIME RATE: 5.26%

Delhi exhibits a significant CRIME RATE of 3978.23, making up 5.26% of the total crime rate.

Goa

CRIME RATE: 2100.00

Percentage of CRIME RATE: 2.78%

Goa reports a CRIME RATE of 2100.00, contributing 2.78% to the overall crime rate.

Gujarat

CRIME RATE: 2221.21

Percentage of CRIME RATE: 2.94%

Gujarat displays a CRIME RATE of 2221.21, accounting for 2.94% of the total crime rate.

Haryana

CRIME RATE: 2421.36

Percentage of CRIME RATE: 3.20%

Haryana shows a CRIME RATE of 2421.36, constituting 3.20% of the total crime rate.

Himachal Pradesh

CRIME RATE: 2302.43

Percentage of CRIME RATE: 3.05%

In Himachal Pradesh, the CRIME RATE stands at 2302.43, contributing 3.05% to the total crime rate.

Jammu & Kashmir

CRIME RATE: 2232.59

Percentage of CRIME RATE: 2.95%

Jammu & Kashmir reports a CRIME RATE of 2232.59, making up 2.95% of the total crime rate.

Jharkhand

CRIME RATE: 1419.03

Percentage of CRIME RATE: 1.88%

Jharkhand records a CRIME RATE of 1419.03, contributing 1.88% to the overall crime rate.

Karnataka

CRIME RATE: 2253.56

Percentage of CRIME RATE: 2.98%

Karnataka exhibits a CRIME RATE of 2253.56, constituting 2.98% of the total crime rate.

Kerala

CRIME RATE: 4209.15

Percentage of CRIME RATE: 5.57%

Kerala has a high CRIME RATE of 4209.15, contributing 5.57% to the overall crime rate.

Lakshadweep UT

CRIME RATE: 965.22

Percentage of CRIME RATE: 1.28%

Lakshadweep UT records a CRIME RATE of 965.22, constituting 1.28% of the total crime rate.

Madhya Pradesh

CRIME RATE: 3474.27

Percentage of CRIME RATE: 4.60%

Madhya Pradesh exhibits a CRIME RATE of 3474.27, contributing 4.60% to the overall crime rate.

Maharashtra

CRIME RATE: 2089.52

Percentage of CRIME RATE: 2.76%

Maharashtra shows a CRIME RATE of 2089.52, making up 2.76% of the total crime rate.

Manipur

CRIME RATE: 1364.32

Percentage of CRIME RATE: 1.81%

Manipur reports a CRIME RATE of 1364.32, contributing 1.81% to the overall crime rate.

Meghalaya

CRIME RATE: 946.50

Percentage of CRIME RATE: 1.25%

Meghalaya displays a CRIME RATE of 946.50, accounting for 1.25% of the total crime rate.

Mizoram

CRIME RATE: 2517.95

Percentage of CRIME RATE: 3.33%

Mizoram exhibits a CRIME RATE of 2517.95, making up 3.33% of the total crime rate.

Nagaland

CRIME RATE: 612.33

Percentage of CRIME RATE: 0.81%

Nagaland records a CRIME RATE of 612.33, constituting 0.81% of the total crime rate.

Orissa

CRIME RATE: 1608.43

Percentage of CRIME RATE: 2.13%

Orissa shows a CRIME RATE of 1608.43, contributing 2.13% to the overall crime rate.

Puducherry UT

CRIME RATE: 4600.59

Percentage of CRIME RATE: 6.09%

Puducherry UT reports a substantial CRIME RATE of 4600.59, making up 6.09% of the total crime rate.

Punjab

CRIME RATE: 1488.21

Percentage of CRIME RATE: 1.97%

Punjab exhibits a CRIME RATE of 1488.21, accounting for 1.97% of the total crime rate.

Rajasthan

CRIME RATE: 2767.12

Percentage of CRIME RATE: 3.66%

Rajasthan shows a CRIME RATE of 2767.12, contributing 3.66% to the overall crime rate.

Sikkim

CRIME RATE: 1067.83

Percentage of CRIME RATE: 1.41%

Sikkim reports a CRIME RATE of 1067.83, constituting 1.41% of the total crime rate.

Tamil Nadu

CRIME RATE: 2881.33

Percentage of CRIME RATE: 3.81%

Tamil Nadu displays a CRIME RATE of 2881.33, making up 3.81% of the total crime rate.

Tripura

CRIME RATE: 1444.63

Percentage of CRIME RATE: 1.91%

Tripura records a CRIME RATE of 1444.63, contributing 1.91% to the overall crime rate.

Uttar Pradesh

CRIME RATE: 2891.50

Percentage of CRIME RATE: 3.83%

Uttar Pradesh reports a CRIME RATE of 2891.50, accounting for 3.83% of the total crime rate.

Uttarakhand

CRIME RATE: 993.36

Percentage of CRIME RATE: 1.31%

Uttarakhand shows a CRIME RATE of 993.36, constituting 1.31% of the total crime rate.

West Bengal

CRIME RATE: 1244.54

Percentage of CRIME RATE: 1.65%

West Bengal exhibits a CRIME RATE of 1244.54, making up 1.65% of the total crime rate

PHASE III

In Phase 3, a meticulous database operation was executed, beginning with the insertion of records from the "District_wise_crimes_committed _against_women_2001_2012.csv" file into a designated table. Subsequent SQL queries were formulated to extract insightful information:

1. A query was devised to identify the state, district, and year with the highest occurrences of rapes and kidnappings.
2. Another query was crafted to uncover the state, district, and year with the lowest incidents of rapes and kidnappings.
3. Records from the "District_wise_crimes_committed_against_ST_ 2001_2012.csv" file were successfully inserted into a new table.
4. An additional query pinpointed the district with the highest number of dacoity or robbery incidents.

The subsequent set of queries focused on murder-related data:

1. The districts with the lowest occurrences of murders across all states were discerned.
2. A query was constructed to display murders in ascending order by district and year.
3. Records specifically related to murder, attempt to murder, and rape from the "District_wise_crimes_committed_IPC_2001_2012.csv" file were strategically inserted into a new table.

4. Another query identified the districts with the highest number of murders yearwise, presenting results for state/UT, year, district, and murders.

The extracted data, specifically the districts with three or more years of appearance, was then stored in a DataFrame for further analysis. Patna emerged as the district with the highest number of murders in this specific analysis, signifying Bihar as the state with the highest murder count. This comprehensive approach to data management and analysis in Phase 3 lays the groundwork for insightful interpretations and subsequent visualization through appropriate graphs.

PHASE IV

In Phase 4, a comprehensive approach to unsupervised machine learning using the KMeans clustering algorithm was employed to analyze crime datasets encompassing districts within each state. The primary objectives were achieved as follows:

1. **Cluster Creation:** Three distinct clusters, namely 'Sensitive Areas,' 'Moderate Areas,' and 'Peaceful Areas,' were generated through the application of the KMeans clustering algorithm. The cluster creation process involved thoughtful merging of pertinent datasets and columns.

2. **DataFrame Creation:** Subsequent to cluster creation, dedicated DataFrames were formulated for each cluster, presenting data categorized according to the specific nature or level of criminal activity within each area.

3. **Cluster Analysis and Report Preparation:** The clusters were meticulously analyzed, leading to the preparation of a comprehensive report. This report delves into various aspects, including the factors influencing crime in sensitive areas, recommended measures for crime reduction, and an identification of the most safe and unsafe districts. Key observations are detailed in the report, providing valuable insights into the inherent patterns within the dataset.

Upon the application of the KMeans clustering algorithm, a notably favorable silhouette score of 0.6122685391695629 was achieved. This metric serves as a pivotal indicator, assessing the clarity and distinctiveness of the clusters. The obtained score strongly suggests the presence of well-defined groupings, demonstrating robust internal cohesion among data points within each cluster and conspicuous separation from other clusters.

The ensuing clusters were judiciously categorized into 'Peaceful_area,' 'Moderate_area,' and 'Sensitive_area.' This classification provides a nuanced lens through which crime patterns within the dataset can be comprehensively understood. The delineation of these clusters enhances the interpretability of underlying trends, offering valuable insights into the varying levels of criminal activity across different areas. This structured approach ensures a systematic and clear representation of the results, fostering a deeper understanding of the dynamics inherent in the crime data analysis.

In the assessment of areas categorized as peaceful, there is a discernible prevalence of heightened occurrences in crimes related to theft, other theft, hurt/grievous hurt, and total crimes against women. Similarly, within moderate areas, an observable prevalence exists with higher occurrences in crimes related to theft, other theft, hurt/grievous hurt, and total crimes against women. Conversely, in areas classified as sensitive, there is an evident prevalence of increased occurrences in crimes related to theft, other theft, and hurt/grievous hurt.

Furthermore, noteworthy observations highlight that Andhra Pradesh and Kerala consistently exhibit the highest overall crime rates. In terms of crimes against Scheduled Castes (SCs), Rajasthan, Uttar Pradesh, Puducherry, Andhra Pradesh, and Sikkim are prominent with the highest incidence. The states of Delhi, Madhya Pradesh, Maharashtra, and Chhattisgarh prominently feature in having the highest occurrences of crimes against children. For crimes against women, Andhra Pradesh, West Bengal, Kerala, and Rajasthan stand out with elevated rates. Lastly, Rajasthan, Chhattisgarh, Puducherry, Sikkim, Odisha, and Madhya Pradesh are noteworthy for having the highest occurrences of crimes against Scheduled Tribes (STs).

These findings contribute valuable insights into the spatial distribution of crime across different regions, aiding in the identification of areas requiring targeted intervention and strategic law enforcement measures.

TOP SAFE DISTRICTS

Sr.no	STATE/UT	DISTRICT	TOTAL IPC CRIMES	Total crimes against SCs	Total crimes against children	Total crimes against women	Total crimes against STs	Overall Safety Score
1	JAMMU & KASHMIR	RAILWAYS KATRA	2	0	10	0	0	12
2	JAMMU & KASHMIR	RAILWAYS KMR	10	0	0	2	0	12
3	MANIPUR	CID	9	0	12	0	0	21
4	A & N ISLANDS	CAR	24	0	2	2	0	28
5	JAMMU & KASHMIR	RAILWAYS JAMMU	29	0	0	1	0	30
6	ASSAM	BIEO	16	0	14	0	0	30
7	JAMMU & KASHMIR	CRIME KASHMIR	32	0	0	0	0	32
8	TRIPURA	GRP	21	10	15	0	1	47
9	JAMMU & KASHMIR	RAILWAYS KASHMIR	43	0	6	2	0	51
10	TRIPURA	G.R.P.	6	0	65	1	0	72
11	NAGALAND	LONGLENG	84	0	2	2	0	88
12	ARUNACHAL PRADESH	ANJAW	85	0	0	4	0	89
13	A & N ISLANDS	NORTH	115	0	7	9	0	131
14	ARUNACHAL PRADESH	UPPER DIBANG VALLEY	148	0	1	2	1	152
15	ASSAM	HAMREN	150	0	0	20	0	170
16	GUJARAT	CID CRIME	54	5	97	2	16	174
17	HIMACHAL PRADESH	CID	14	0	171	0	0	185
18	NAGALAND	KIPHIRE	195	0	0	3	0	198
19	RAJASTHAN	G.R.P. JODHPUR	149	2	44	5	0	200
20	A & N ISLANDS	NICOBAR	235	0	1	11	3	250

TOP UNSAFE DISTRICTS

Sr.no	STATE/UT	DISTRICT	TOTAL IPC CRIMES	Total crimes against SCs	Total crimes against children	Total crimes against women	Total crimes against STs	Overall Safety Score
1	KARNATAKA	BANGALORE COMMR.	350347	1423	26	9010	16	360822
2	GUJARAT	AHMEDABAD COMMR.	218005	823	418	14973	35	234254
3	MAHARASHTRA	MUMBAI COMMR.	222670	186	99	9304	11	232270
4	ANDHRA PRADESH	HYDERABAD CITY	202931	1293	872	19685	76	224857
5	MADHYA PRADESH	INDORE	204398	1139	3	10392	421	216353
6	MADHYA PRADESH	BHOPAL	158725	1459	1125	6618	208	168135
7	MAHARASHTRA	MUMBAI	141815	67	99	4251	2	146234
8	ANDHRA PRADESH	CYBERABAD	121527	1495	294	13715	274	137305
9	TAMIL NADU	CHENNAI	118295	5156	65	3759	81	127356
10	BIHAR	PATNA	109340	1359	164	4609	31	115503
11	RAJASTHAN	JAIPUR	99527	874	22	5007	7	105437
12	WEST BENGAL	KOLKATA	96622	0	867	6278	0	103767
13	MAHARASHTRA	PUNE COMMR.	91482	124	149	4100	3	95858
14	WEST BENGAL	24 PARGANAS SOUTH	72465	20	51	16067	18	88621
15	WEST BENGAL	24 PARGANAS NORTH	57374	162	8	12371	121	70036
16	WEST BENGAL	MURSHIDABAD	52269	4	27	17661	11	69972
17	KERALA	ERNAKULAM RURAL	59213	104	250	1112	6	60685
18	KERALA	ERNAKULAM COMMR.	56251	98	142	1032	9	57532
19	UTTAR PRADESH	LUCKNOW	38585	179	41	4693	9	43507
20	KERALA	ERNAKULAM	30516	97	18	766	2	31399

TOP SAFE STATES

Sr.no	STATE/UT	TOTAL IPC CRIMES	Total crimes against SCs	Total crimes against children	Total crimes against women	Total crimes against STs	Overall Safety Score
18	LAKSHADWEEP	743	1	16	20	1	781
8	DAMAN & DIU	2948	5	988	90	5	4036
7	D & N HAVELI	4651	7	103	243	73	5077
26	PUDUCHERRY	8464	164	158	236	0	9022
0	A & N ISLANDS	9102	0	245	570	41	9958

TOP UNSAFE STATES

Sr.no	STATE/UT	TOTAL IPC CRIMES	Total crimes against SCs	Total crimes against children	Total crimes against women	Total crimes against STs	Overall Safety Score
6	MAHARASHTRA	455967	377	347	17655	16	474362
5	MADHYA PRADESH	363123	2598	1128	17010	629	384488
10	WEST BENGAL	315912	192	1085	59683	156	377028
0	ANDHRA PRADESH	324458	2788	1166	33400	350	362162
3	KARNATAKA	350347	1423	26	9010	16	360822

Expressing Gratitude for the Capstone Project Experience

I extend my heartfelt gratitude for entrusting me with this insightful capstone project. Working on the analysis of crime data in India has been a transformative journey, enabling me to apply and enhance my analytical and technical skills. The project not only deepened my understanding of data collection, SQL operations, and machine learning but also provided a practical platform for real-world application. I appreciate the opportunity to independently navigate through the intricacies of each phase, contributing significantly to my learning experience. Thank you for this enriching project that has undoubtedly expanded my knowledge and skill set.