**Crimes in India - Capstone Project**

Evaluation

New

Due Date: 17 Mar 2024

**Phase 1**

**Data Collection/preparation part.**

**Note:**

1. You can collect data from anywhere (wiki, google, etc) but mention the link from where data is being collected.
2. You can collect the data manually. It's NOT a web scraping task.
3. Data should be collected for the years mentioned in the files. (If you don't find proper data, make your best judgment and fill it, It's absolutely OK. Do not worry about data accuracy)
4. The population of each state.
5. Literacy Rate in each state
6. Area of each state
7. Collect any other data that helps with your analysis. There is no limitation for anything.
8. Create a new file and keep the above-collected data.

**Phase 2**

**State/UT wise analysis.**

Note :

1. You need to be careful as this phase is whole state/ut wise analysis (not district-wise).
2. It is up to you which columns you want to keep or delete to make proper 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**

Note: Create a separate table for each file and give a meaningful name.

When you are inserting make sure you are NOT inserting the column name in the table, else you don’t get a result.

3.1 Insert records from 42\_District\_wise\_crimes\_committed\_against\_women\_2001\_2012.csv into a table

CREATE TABLE crimes (

STATE\_UT TEXT,

DISTRICT TEXT,

Year INTEGER,

Rape INTEGER,

Kidnapping\_and\_Abduction INTEGER,

Dowry\_Deaths INTEGER,

Assault\_on\_women\_with\_intent\_to\_outrage\_her\_modesty INTEGER,

Insult\_to\_modesty\_of\_Women INTEGER,

Cruelty\_by\_Husband\_or\_his\_Relatives INTEGER,

Importation\_of\_Girls INTEGER

);

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

SELECT

STATE\_UT,

DISTRICT,

Year,

MAX(Rape) AS Max\_Rapes,

MAX(Kidnapping\_and\_Abduction) AS Max\_Kidnappings

FROM

crimes

GROUP BY

STATE\_UT,

DISTRICT,

Year

ORDER BY

Max\_Rapes DESC,

Max\_Kidnappings DESC

LIMIT 1;

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

SELECT

STATE\_UT,

DISTRICT,

Year,

Min(Rape) AS Min\_Rapes,

Min(Kidnapping\_and\_Abduction) AS Min\_Kidnappings

FROM

crimes

GROUP BY

STATE\_UT,

DISTRICT,

Year

ORDER BY

Min\_Rapes ASC,

Min\_Kidnappings ASC

LIMIT 1;

3.4 Insert records from 02\_District\_wise\_crimes\_committed\_against\_ST\_2001\_2012.csv into a new table

CREATE TABLE crimes\_against\_st (

STATE\_UT TEXT,

DISTRICT TEXT,

Year INTEGER,

Murder INTEGER,

Rape INTEGER,

Kidnapping\_Abduction INTEGER,

Dacoity INTEGER,

Robbery INTEGER,

Arson INTEGER,

Hurt INTEGER,

Protection\_of\_Civil\_Rights\_PCR\_Act INTEGER,

Prevention\_of\_atrocities\_POA\_Act INTEGER,

Other\_Crimes\_Against\_STs INTEGER

);

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

SELECT

DISTRICT,

STATE\_UT,

SUM(Dacoity + Robbery) AS Total\_Dacoity\_Robbery

FROM

crimes\_against\_st

WHERE

DISTRICT != 'Total'

GROUP BY

DISTRICT, STATE\_UT

ORDER BY

Total\_Dacoity\_Robbery DESC

LIMIT 1;

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

SELECT

STATE\_UT,

DISTRICT,

Murder

FROM

crimes\_against\_st

WHERE

Murder = (

SELECT

MIN(Murder)

FROM

crimes\_against\_st

)

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

SELECT

DISTRICT,

Year,

Murder

FROM

crimes\_against\_st

ORDER BY

DISTRICT ASC,

Year ASC;

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

CREATE TABLE CrimeData (

State\_UT TEXT,

District TEXT,

Year INT,

Murder INT,

Attempt\_To\_Murder INT,

Rape INT

);

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.

WITH RankedMurders AS (

SELECT

State\_UT,

Year,

District,

Murder,

ROW\_NUMBER() OVER (PARTITION BY State\_UT, Year ORDER BY Murder DESC) AS RowNum

FROM

crimedata\_ipc\_2001\_2012

WHERE

District != 'TOTAL'

)

SELECT

State\_UT,

Year,

District,

Murder

FROM

RankedMurders

WHERE

RowNum = 1;

3.8.3 Store 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.

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

**Phase 4**

**Unsupervised ML (Clustering)**

**Note:** This phase requires your general knowledge and thought process to merge the right datasets and the right columns. Collecting quality data is a key factor to create meaningful clusters and Analysis.

**Assumption:** You were given various crime datasets that contains all the DISTRICTS in each state and you were asked to provide the below data to the higher authorities for further action.

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.***

Capstone Project

**Problem Statement:**

Crime datasets that contain all the DISTRICTS in each state are provided from 2001 to 2012.

Dataset contain Indian State/ UT, Year, District, MURDER, ATTEMPT TO MURDER, CULPABLE HOMICIDE NOT AMOUNTING TO MURDER, RAPE, CUSTODIAL RAPE, OTHER RAPE, KIDNAPPING & ABDUCTION, KIDNAPPING AND ABDUCTION OF WOMEN AND GIRLS, KIDNAPPING AND ABDUCTION OF OTHERS, DACOITY, PREPARATION AND ASSEMBLY FOR DACOITY, ROBBERY, BURGLARY, THEFT, AUTO THEFT, OTHER THEFT

RIOTS, CRIMINAL BREACH OF TRUST, CHEATING, COUNTERFIETING, ARSON, HURT /GREVIOUS HURT, 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 FROM FOREIGN COUNTRIES, CAUSING DEATH BY NEGLIGENCE, OTHER IPC CRIMES and Total Crimes.

**In Phase 1 and 2 :** We have to create a table having Indian States/ UT adding the Literacy rate, Area ad Population Year on Year from 2001 to 2012 from internet

Also, We have to find the relation of the crimes with Literacy at State/ UT, relation of types of Crimes Vs State Vs Literacy rate, Year on Year total crime rate, Area vs overall crime, population Vs overall Crime.

**Below are the observations:**

- *Minimum Literacy Rate is in Bihar*

- Most of the Crimes are in Literacy Rate between 45- 82

**- In Kerala, where** is the literacy rate is **high > 90%, crimes are also high**

- Madhya Pradesh, Maharashtra, UP and Tamil Nadu has the highest Crimes

- Crime MURDER is highest in Uttar Pradesh and Bihar with Literacy Rate (<65%). It is Lowest in LAKSHADWEEP and DAMAN & DIU with Literacy Rate (>80%)

- Crime ATTEMPT TO MURDER is highest in Uttar Pradesh and Bihar with Literacy Rate (<65%). It is Lowest in LAKSHADWEEP and DAMAN & DIU with Literacy Rate (>80%)

- Crime CULPABLE HOMICIDE NOT AMOUNTING TO MURDER is highest in Uttar Pradesh and West Bengal with Literacy Rate is between 56% and 75%. It is Lowest in LAKSHADWEEP and DAMAN & DIU with Literacy Rate (>80%)

- Crime RAPE is highest in Madhya Pradesh and West Bengal with Literacy Rate is between 63% - 75%. It is Lowest in LAKSHADWEEP and DAMAN & DIU with Literacy Rate (>80%)

- Crime OTHER RAPE is highest in Madhya Pradesh and West Bengal with Literacy Rate is between 63% - 75%. It is Lowest in LAKSHADWEEP and DAMAN & DIU with Literacy Rate (>80%)

- Crime KIDNAPPING & ABDUCTION is highest in Uttar Pradesh with Literacy Rate is between 56% - 65%. It is Lowest in LAKSHADWEEP and DAMAN & DIU with Literacy Rate (>80%)

- Crime KIDNAPPING AND ABDUCTION OF WOMEN AND GIRLS is highest in Uttar Pradesh and West Bengal with Literacy Rate is between 56% - 75%. It is Lowest in LAKSHADWEEP and DAMAN & DIU with Literacy Rate (>80%)

- Crime KIDNAPPING AND ABDUCTION OF OTHERS is highest in Uttar Pradesh and Bihar with Literacy Rate (<65%). It is Lowest in LAKSHADWEEP and DAMAN & DIU with Literacy Rate (>80%)

- Crime DACOITY is highest in Uttar Pradesh and Bihar with Literacy Rate (<65%). It is Lowest in LAKSHADWEEP and DAMAN & DIU with Literacy Rate (>80%)

- Crime PREPARATION AND ASSEMBLY FOR DACOITY is highest in West Bengal with Literacy Rate is between 63% - 75%. It is Lowest in LAKSHADWEEP and DAMAN & DIU with Literacy Rate (>80%)

- Crime ROBBERY is highest in Maharashtra and Uttar Pradesh with Maharashtra has high Literacy Rate >80% and UP has lower than 65%. It is Lowest in LAKSHADWEEP and DAMAN & DIU with Literacy Rate (>80%)

- Crime BURGLARY is highest in Maharashtra and Madhya Pradesh with Maharashtra has high Literacy Rate >80% and UP has lower than 70%. It is Lowest in LAKSHADWEEP and DAMAN & DIU with Literacy Rate (>80%)

- Crime THEFT is highest in Maharashtra and Uttar Pradesh with Maharashtra has high Literacy Rate >80% and UP has lower than 65%. It is Lowest in LAKSHADWEEP and DAMAN & DIU with Literacy Rate (>80%)

- Crime AUTO THEFT is highest in Maharashtra and Uttar Pradesh with Maharashtra has high Literacy Rate >80% and UP has lower than 65%. It is Lowest in LAKSHADWEEP and DAMAN & DIU with Literacy Rate (>80%)

- Crime OTHER THEFT is highest in Maharashtra with Maharashtra has high Literacy Rate >80%. It is Lowest in LAKSHADWEEP and DAMAN & DIU with Literacy Rate (>80%)

- Crime RIOTS is highest in Rajasthan, Kerala Bihar. It is Lowest in Mizoram with Literacy Rate (>80%)

- Crime CRIMINAL BREACH OF TRUST is highest in Maharashtra and Uttar Pradesh with Maharashtra has high Literacy Rate >80% and UP has lower than 65%. It is Lowest in with

- Crime CHEATING is highest in Rajasthan with Literacy Rate (<65%). It is Lowest in LAKSHADWEEP and DAMAN & DIU with Literacy Rate (>80%)

- Crime COUNTERFIETING is highest in Tamil Nadu and Maharashtra. It is Lowest in LAKSHADWEEP and DAMAN & DIU with Literacy Rate (>80%)

- Crime ARSON is highest in Maharashtra. It is Lowest in LAKSHADWEEP and DAMAN & DIU with Literacy Rate (>80%)

- Crime HURT/GREVIOUS HURT is highest in Andhra Pradesh. It is Lowest in LAKSHADWEEP and DAMAN & DIU with Literacy Rate (>80%)

- Crime DOWRY DEATHS is highest in Uttar Pradesh. It is Lowest in LAKSHADWEEP and DAMAN & DIU with Literacy Rate (>80%)

- Crime ASSAULT ON WOMEN WITH INTENT TO OUTRAGE HER MODESTY is highest in Madhya Pradesh. It is Lowest in LAKSHADWEEP and DAMAN & DIU with Literacy Rate (>80%)

- Crime INSULT TO MODESTY OF WOMEN is highest in Andhra Pradesh and Uttar Pradesh. It is Lowest in LAKSHADWEEP and DAMAN & DIU with Literacy Rate (>80%)

- Crime CRUELTY BY HUSBAND OR HIS RELATIVES is highest in West Bengal and Andhra Pradesh. It is Lowest in LAKSHADWEEP and DAMAN & DIU with Literacy Rate (>80%)

- Not much case, but most of them in Bihar for "IMPORTATION OF GIRLS FROM FOREIGN COUNTRIES"

- Crime CAUSING DEATH BY NEGLIGENCE is highest in Tamil Nadu and Uttar Pradesh. It is Lowest in LAKSHADWEEP and DAMAN & DIU with Literacy Rate (>80%)

- YoY Crime Rate is increasing with a dip in 2003

- Highest Crime rate is observed in Madhya Pradesh and Maharashtra. with Area > 300K sq km.

- States with low area like Lakshadweep, Daman Diu, D Nagar Haveli, Sikkim, Chandigarh and Pondicherry has low crime count.

- There is a linear relation between the population and the crime numbers. With increase in population number of crimes are higher

**In Phase 3:** It was a SQL exercise to create tables, apply rules with conditions. Then the database to be imported in python for analysis and for supporting the analysis by plot creation.

**In Phase 4:** To create 3 clusters using Unsupervised Clustering ML model. These Create 3 clusters are as Sensitive Area's, Moderate Area's and Peaceful Area's. And to add the observations.

Observations:

* In Sensitive area’s – crimes most impacting are Theft (33.5%), OtherTheft(18%) and Auto Theft (16%).
* Districts with highest crime rate are: Delhi UT, Banglore Comm, Mumbai, Mumbai Comm
* Districts with least crime are Manipur, Katra, CID (HP) and GRP(Tripura)