Description:

Capstone Project
(Crimes in India)

Important notes.

1. By now you have worked on many use cases. As this is a capstone project and absolutely No support will be given from the technical team so avoid asking questions on the capstone project.
2. When you are analyzing any situation if you don't get proper data/information or are in doubt about anything. You are free to make assumptions and proceed. But make sure whenever you assume something you MUST write what is your assumption w.r.t that situation so that we understand your work accordingly.
3. You can merge all the files or work on them individually it's up to you how you think and want to deal with it.
4. You can make use of any column data for your analysis there is no fixed procedure to follow as long as your analysis makes sense.
5. A simple point to keep in mind, the more the analysis is the better.
6. For every analysis, document your analysis within the jupyter notebook (in the same cell where your graph or code exists)
7. Write Phase number and Req. Numb for every task to identify later. You can combine requirement numbers where it is necessary.
8. You need to post your complete work in LMS
9. These marks will be considered for your placement-related activities.
10. You can use ANY library/module.

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)

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.

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

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.

   3.8.3   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.

3. 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.