The data is collected and analysed for the years 2001- 2012 on population, literacy rate and area of each state in India. The data is taken manually from reliable sources like government websites and the statistical databases. The SQL Queries are written one after another. The Data collected is Literacy Rate and Population which have been collected in the previous collections. The data has been taken from NRCB reports accessible in the website. In order to execute the tasks outlined the SQL query syntax is used to handle the data and perform each of the remaining queries.

In a simple way it can be done by a stepwise approach-

The records are inserted from CSV files into SQL tables. SQL queries are run for required analysis. Store and Analyse Results Using Python. The results are then visualised using the matplotlib visualisation functions.

The code is based on the assumption that the user has used the correct pathways and saved the data into CSV files named 'population\_ data.csv' , 'literach\_rate\_data.csv', 'area\_data.csv' and 'crime\_ data.csv.'Adjust the file names accordingly if they differ. The code performs the requested analyses and visualizations based on the collected data.

File Paths Used-

"C:\Users\Rishav Gupta\Desktop\CAPSTONE\India\_Crime - Copy"

"C:\Users\Rishav Gupta\Desktop\CAPSTONE\India\_Crime - Copy\01\_District\_wise\_crimes\_committed\_IPC\_2001\_2012.csv"

"C:\Users\Rishav Gupta\Desktop\CAPSTONE\India\_Crime - Copy\02\_01\_District\_wise\_crimes\_committed\_against\_SC\_2001\_2012.csv"

"C:\Users\Rishav Gupta\Desktop\CAPSTONE\India\_Crime - Copy\03\_District\_wise\_crimes\_committed\_against\_children\_2001\_2012.csv"

"C:\Users\Rishav Gupta\Desktop\CAPSTONE\India\_Crime - Copy\02\_District\_wise\_crimes\_committed\_against\_ST\_2001\_2012.csv"

For the final clustering part this is the description-

To accomplish the tasks outlined, you will need to perform several steps in a Jupyter notebook, including loading the data, creating clusters, analyzing the data, and generating a report with observations. Below, I'll guide you through each of these steps with code snippets.

1. Create Clusters:

• Sensitive Areas

• Moderate Areas

• Peaceful Areas

2. Create DataFrames for Each Cluster:

• Extract relevant data for each cluster.

3. Analyze Clusters and Prepare a Report:

• Identify factors impacting crime.

• Suggest measures to reduce crime.

• Highlight the most safe and unsafe districts.

• Additional observations and visualizations.

First, let's set up the environment and load the data.

Step 1: Setting up the environment and loading the data

Step 2: Creating Clusters

For clustering, assume we have columns such as 'Crime Rate', 'Police Presence', 'Socioeconomic Status', etc. We'll categorize areas into Sensitive, Moderate, and Peaceful based on these factors.

Summary of Changes:

1. Renamed the problematic column "STATE/UT" to "State\_UT" in all DataFrames.

2. Updated the SQL table creation statements to match the new column names.

3. Updated the SQL queries to reflect the new column names.

This should resolve the OperationalError related to the column name syntax issue.

To perform the tasks outlined in your query, you need to follow several steps involving data import, table creation, and SQL queries. This requires setting up a SQL database, importing data from CSV files, executing queries, and analyzing the results. Below, I'll guide you through each step, including SQL commands and Python code for Jupyter Notebook.