Mini Capstone Project: Crime Data Analysis with MySQL and Python

Click here to download dataset

Introduction:

In this capstone project, we will use Python, specifically the PyMySQL library, to interact with a MySQL database in order to analyze and gain insights from crime data. The dataset includes information such as DR NO, Date Reported, Date Occurred, Area Name, Crime Code, Crime Code Description, Victim Age, Victim Sex, Premises Description, Status, Location, Latitude, and Longitude.

Objectives:

1. Database Setup and Import :

 - Create a MySQL database.

  - Load the provided crime dataset into the MySQL database.

2. Database Connection :

 - Use PyMySQL to establish a connection to the database in Pycharm or VS code.

 - Verify the successful import of data in pycharm.

3. Data Exploration:

 - Retrieve basic statistics on the dataset, such as the total number of records and unique values in specific columns.

 - Identify the distinct crime codes and their descriptions.

4. Temporal Analysis:

 - Analyze the temporal aspects of the data.

 - Determine trends in crime occurrence over time.

5. Spatial Analysis:

 - Utilize the geographical information (Latitude and Longitude) to perform spatial analysis.

 - Visualize crime hotspots on a map.

6. Victim Demographics:

   - Investigate the distribution of victim ages and genders.

   - Identify common premises descriptions where crimes occur.

7. Status Analysis:

   - Examine the status of reported crimes.

   - Classify crimes based on their current status.

Questions:

Spatial Analysis:

Where are the geographical hotspots for reported crimes?

Victim Demographics:

What is the distribution of victim ages in reported crimes?

Is there a significant difference in crime rates between male and female victims?

Location Analysis:

Where do most crimes occur based on the "Location" column?

Crime Code Analysis:

What is the distribution of reported crimes based on Crime Code?

Tools and Libraries:

- PyCharm or Visual Studio Code for Python development.

- PyMySQL for interacting with MySQL database.

- Matplotlib and Seaborn for data visualization.

Deliverables:

- Python scripts for database setup, data import, and data analysis.

- Visualizations and insights derived from the analysis.

Note: Make sure to install the necessary libraries using `pip install pymysql matplotlib seaborn`. Additionally, ensure that you have a running MySQL server for this project.