Project: US Crime Data Analysis and Exploration using SQL

Problem Scenario: An organized quantitative and qualitative investigation is done to find trends in crime and disorder. Information on these patterns helps law enforcement agencies deploy resources more effectively. Crime analysis plays an important role in devising solutions to crime problems and formulating crime prevention strategies.

Objective: It is required to delve deeper into data on different types of crimes and figure out the types of crimes which are more frequent and how they are trending over time.

Data Description: The Crime_us.csv is a huge dataset with 220550 rows and 23 columns. It contains information of various crime in a state of the US for the year 2022.

Data Cleaning: Preliminary inspection of the data revealed that the date column was not in the MySQL format. Hence, before uploading to MySQL workbench the dataset had to be formatted. It was done using Excel.

- The columns, "Date" and "Updated On" were formatted as "y-mm-dd h:mm:ss".
- The duplicates were removed.

The cleaned data now consisted of 220510 rows and 23 columns.

Tasks:

1. Created a SQL database containing data related to the case number, primary crime category, crime description, crime location, and arrest status using the dataset.
2. Made a database in SQL where theft costs more than $500. There were 18770 cases in this category.
3. Determined the overall number of cases for each major category of crime. The highest number of cases comprised of Theft which consisted of 102072 cases.
4. Applied 1NF normalization to the dataset provided.