

**US Crime Data Exploration and Analysis**

Phase-End Project Problem Statement





**Phase-End Project: US Crime Data Exploration and Analysis**

### 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.

**Problem 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 file Crime\_us.csv contains the details of crimes, that have occurred in a state of the US in the year 2022.

**Variable Description:**

|  |  |
| --- | --- |
| **Variable** | **Description** |
| ID | Shows case id |
| Case Number | Shows case number |
| Date | Explicit the date and time of occurrence of the crime |
| Block | Explains the address where the crime had taken place |
| IUCR | Is a four-digit code that law enforcement agencies use to classify criminal incidents when taking individual reports |

|  |  |
| --- | --- |
| Primary Type | Classify the type of crime |
| Description | Demonstrates the crime event |
| Location Description | Explains the type of location where the crime has taken place |
| Arrest | Shows whether an arrest had been made or not |
| Domestic | Shows whether the crime was of domestic nature |
| Beat | Is the territory that a police officer is assigned to patrol |
| District | Shows the district of the victim |
| Ward | Shows the ward of the victim |
| Community Area | Shows the community of the victim |
| FBI Code | Is the investigation code used to find the criminals |
| X Coordinate | Shows various information about the location |
| Y Coordinate | Shows various information about the location |
| Year | Shows the year |
| Update On | Signifies the last updating date of the data |
| Latitude | Shows the latitude of the place |
| Longitude | Shows the longitude of the place |
| Location | Shows the location |

**Steps to Perform:**

**Excel:**

1. Check the dataset for any missing values, and if any are present, delete the row corresponding to that particular cell. Verify once more that 2022 is the only year for which data has been collected.
2. Sort the data in the provided dataset based on the number of primary types of crime instances, then create a Pareto chart to visualize the results. Next, sort the data according to the crime's location and description, and then create a histogram to represent the pertinent facts.
3. Use conditional formatting in excel to generate data bars for the top occurring crime based on its primary type. Use red color to identify the crime where theft is over $500 in residence and arrest has not been done.
4. Generate a frequency distribution chart of the data for the community area and note any skewness.
5. To find trends and patterns for the most common sort of crime over the different months of 2022, create a heatmap and sparklines.
6. To predict the future number of total crimes across all categories, create a trendline for the frequency of each key crime type occurring in 2022.
7. Create a cross-pivot table for the community area and crime description, then analyze the results.
8. Create an excel dashboard that displays trend charts over the last three months for the several key categories of crimes based on the overall number of crimes that have occurred. (To make a dynamic chart, use form controls.)

**SQL:**

1. Create a SQL database containing data related to the case number, primary crime category, crime description, crime location, and arrest status using the dataset.
2. Make a database in SQL where theft costs more than $500.
3. Determine the overall number of cases for each major category of crime.
4. Apply 1NF normalization to the dataset provided.