# CRIME DATA ANALYSIS

Ingoude Company

# INTRODUCTION

In this  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.

# OBJECTIVE

01 Database Setup and Import

02 Database Connection

03 Data Exploration

04 Temporal Analysis

05 Spatial Analysis

06 Victim Demographics

07 Analysis Feature

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?

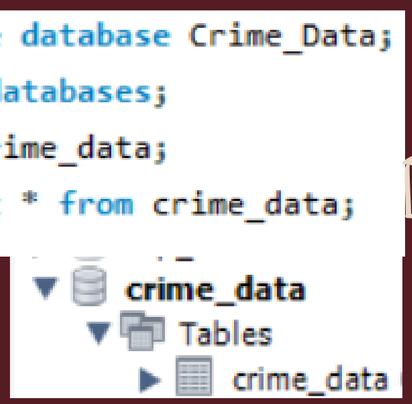
**Location Analysis:**
Where do most crimes occur based on the "Location" column?

**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.

# MAKING CONNECTION WITH THE DB

```
create database Crime_Data;
show databases;
use crime_data;
select * from crime_data;
```




Result Grid | Filter Rows: | Export: | Wrap Cell Content:

| DR_NO | Date_Rptd | DATE_OCC | AREA_NAME | Crm_Cd | Crm_Cd_Desc | Vict_Age | Vict_Sex | Premis_Desc | Status | Location | LAT | LON |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 10304468 | 01-08-2020 | 01-08-2020 | Southwest | 624 | BATTERY - SIMPLE ASSAULT | 36 | F | SINGLE FAMILY DWELLING | AO | 1100 W 39TH PL | 34.01 | -118.3 |
| 190101086 | 01-02-2020 | 01-01-2020 | Central | 624 | BATTERY - SIMPLE ASSAULT | 25 | M | SIDEWALK | IC | 700 S HILL ST | 34.05 | -118.25 |
| 191501505 | 01-01-2020 | 01-01-2020 | N Hollywood | 745 | VANDALISM - MISDEAMEANOR ($399 OR UNDER) | 76 | F | MULTI-UNIT DWELLING (APARTMENT, DUPLEX, ... | IC | 5400 CORTEEN PL | 34.17 | -118.4 |
| 191921269 | 01-01-2020 | 01-01-2020 | Mission | 740 | VANDALISM - FELONY ($400 & OVER, ALL CHU... | 31 | X | BEAUTY SUPPLY STORE | IC | 14400 TITUS ST | 34.22 | -118.45 |
| 200100502 | 01-02-2020 | 01-02-2020 | Central | 442 | SHOPLIFTING - PETTY THEFT ($950 & UNDER) | 23 | M | DEPARTMENT STORE | IC | 700 S FIGUEROA ST | 34.05 | -118.26 |
| 200100504 | 01-04-2020 | 01-04-2020 | Central | 946 | OTHER MISCELLANEOUS CRIME | 0 | X | POLICE FACILITY | IC | 200 E 6TH ST | 34.04 | -118.25 |
| 200100507 | 01-04-2020 | 01-04-2020 | Central | 341 | THEFT-GRAND ($950.01 & OVER)EXCPT,GUNS,... | 23 | M | MULTI-UNIT DWELLING (APARTMENT, DUPLEX, ... | IC | 700 BERNARD ST | 34.07 | -118.24 |
| 200100509 | 01-04-2020 | 01-04-2020 | Central | 330 | BURGLARY FROM VEHICLE | 29 | M | STREET | IC | 800 N ALAMEDA ST | 34.04 | -118.26 |
| 200100510 | 01-05-2020 | 01-05-2020 | Central | 930 | CRIMINAL THREATS - NO WEAPON DISPLAYED | 35 | M | PARKING LOT | IC | 800 S OLIVE ST | 34.06 | -118.24 |
| 200100514 | 01-05-2020 | 01-05-2020 | Central | 341 | THEFT-GRAND ($950.01 & OVER)EXCPT,GUNS,... | 41 | M | HOTEL | AA | 700 W 7TH ST | 34.05 | -118.26 |
| 200100515 | 01-07-2020 | 01-07-2020 | Central | 648 | ARSON | 0 | X | DEPARTMENT STORE | IC | 100 S LOS ANGELE... | 34.05 | -118.26 |
| 200100520 | 01-08-2020 | 01-08-2020 | Central | 442 | SHOPLIFTING - PETTY THEFT ($950 & UNDER) | 24 | F | COFFEE SHOP (STARBUCKS, COFFEE BEAN, PE... | IC | 13600 LEADWELL ST | 34.05 | -118.24 |
| 200914517 | 09-10-2020 | 09-09-2020 | Van Nuys | 354 | THEFT OF IDENTITY | 40 | M | CONDOMINIUM/TOWNHOUSE | IC | 8TH | 34.2 | -118.43 |
| 200716724 | 12-04-2020 | 12-03-2020 | Wilshire | 442 | SHOPLIFTING - PETTY THEFT ($950 & UNDER) | 19 | M | OTHER BUSINESS | IC | WALL | 34.08 | -118.37 |
| 200100583 | 02-04-2020 | 02-04-2020 | Central | 230 | ASSAULT WITH DEADLY WEAPON, AGGRAVATE... | 38 | F | OTHER BUSINESS | AA | 800 N BROADWAY | 34.05 | -118.25 |
| 200100584 | 02-04-2020 | 02-04-2020 | Central | 442 | SHOPLIFTING - PETTY THEFT ($950 & UNDER) | 55 | M | DEPARTMENT STORE | IC | 400 S SPRING ST | 34.05 | -118.26 |
| 200100587 | 02-06-2020 | 02-06-2020 | Central | 740 | VANDALISM - FELONY ($400 & OVER, ALL CHU... | 66 | F | VEHICLE, PASSENGER/TRUCK | IC | 300 W 5TH ST | 34.04 | -118.25 |
| 200218458 | 12-11-2020 | 12-11-2020 | Rampart | 761 | BRANDISH WEAPON | 34 | M | PARKING LOT | AO | 1800 S MAIN ST | 34.07 | -118.28 |
| 200100596 | 02-08-2020 | 02-08-2020 | Central | 442 | SHOPLIFTING - PETTY THEFT ($950 & UNDER) | 35 | M | OTHER STORE | IC | 600 W 7TH ST | 34.05 | -118.26 |

```
import pymysql
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

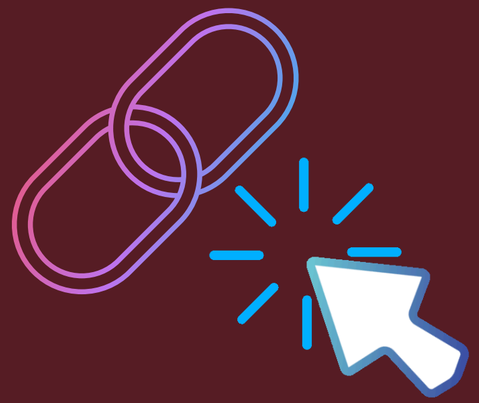
```
connection = pymysql.connect(host='localhost',
        user='root',
        password='12345678',
        database='crime_data')
```

```
df=pd.read_sql_query(""" Select * from crime_data """,connection)
df
```

| | DR_NO | Date_Rptd | DATE_OCC | AREA_NAME | Crm_Cd | Crm_Cd_Desc | Vict_Age | Vict_Sex | Premis_Desc | Status | Location | LAT | LON |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 10304468 | 01-08-2020 | 01-08-2020 | Southwest | 624 | BATTERY - SIMPLE ASSAULT | 36 | F | SINGLE FAMILY DWELLING | AO | 1100 W 39TH PL | 34.01 | -118.30 |
| 1 | 190101086 | 01-02-2020 | 01-01-2020 | Central | 624 | BATTERY - SIMPLE ASSAULT | 25 | M | SIDEWALK | IC | 700 S HILL ST | 34.05 | -118.25 |
| 2 | 191501505 | 01-01-2020 | 01-01-2020 | N Hollywood | 745 | VANDALISM - MISDEAMEANOR ($399 OR UNDER) | 76 | F | MULTI-UNIT DWELLING (APARTMENT, DUPLEX, ETC) | IC | 5400 CORTEEN PL | 34.17 | -118.40 |
| 3 | 191921269 | 01-01-2020 | 01-01-2020 | Mission | 740 | VANDALISM - FELONY ($400 & OVER, ALL CHURCH VA... | 31 | X | BEAUTY SUPPLY STORE | IC | 14400 TITUS ST | 34.22 | -118.45 |
| 4 | 200100502 | 01-02-2020 | 01-02-2020 | Central | 442 | SHOPLIFTING - PETTY THEFT ($950 & UNDER) | 23 | M | DEPARTMENT STORE | IC | 700 S FIGUEROA ST | 34.05 | -118.26 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |

SUCCESSFULLY CONNECTED

# ”Project Demo Video Overview: Addressing Key Questions and Insights”



# INSIGHTS

1. Spatial Analysis:
    a. Crime rates are particularly high in Los Angeles, concentrated within the latitude range of 34.0 to 34.1 and longitude range of -118.30 to -118.25.
    b. Action: Increase police presence and surveillance in high-crime areas to deter criminal activity and improve public safety.
2. Victim Demographic:
    a. Most criminals do not mention their age, indicating a lack of demographic data in crime reports.
    b. Action: Implement measures to ensure accurate reporting of demographic information in crime reports to better understand and address the root causes of criminal behavior.
3. Location Analysis:
    a. Certain locations such as 800 N Alamenda St, 700 W 7th St, 300 E 5th St, and 1100 S Figueroa St are hotspots for criminal activity.
    b. Action: Implement targeted policing strategies and community outreach programs in high-crime locations to address underlying issues contributing to crime.
4. Crime Code Analysis:
    a. A significant portion of crimes are attributed to the top 4 crime codes (330, 624, 440, 442, 510).
    b. Action: Increase law enforcement focus on addressing crimes associated with these specific crime codes through targeted investigations and enforcement efforts.
5. Temporal Analysis:
    a. The majority of crimes are reported in the first three months of the dataset, indicating temporal patterns in crime occurrence over time.
    b. Action: Implement proactive policing measures and public awareness campaigns during peak crime periods to prevent criminal activity and promote community safety.

# SUGGESTION

1. Spatial Analysis:
    - Increase police presence in high-crime areas.
2. Victim Demographic:
    - Improve reporting accuracy for age demographics.
3. Location Analysis:
    - Implement targeted policing strategies in crime hotspots.
4. Crime Code Analysis:
    - Focus law enforcement efforts on top crime codes.
5. Temporal Analysis:
    - Implement proactive policing during peak crime periods.

# TOOLS AND LIBRARIES USED


## JUPYTER NOTEBOOK

Jupyter Notebook facilitates interactive data analysis and visualization in Python.


## PYMYSQL LIBRARY

- PyMySQL for interacting with MySQL database.



## MYTPLOTLIB & SEABORN

- Matplotlib and Seaborn for data visualization.

# Thank You

## Contact


9820673268

gauravslnkh@gmail.com

portfolio