CRIME DATA ANALYSIS with PYTHON and MySQL

Presented by Vinay Rajbhar

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

OBJECTIVES

01. SPATIAL ANALYSIS

- Utilize the geographical information (Latitude and Longitude) to perform spatial analysis.
- Visualize crime hotspots on a map.

02. VICTIM DEMOGRAPHICS

- Investigate the distribution of victim ages and genders.
- Identify common premises descriptions where crimes occur.

03. STATUS ANALYSIS

- Examine the status of reported crimes.
- Classify crimes based on their current status.

SPATIAL ANALYSIS

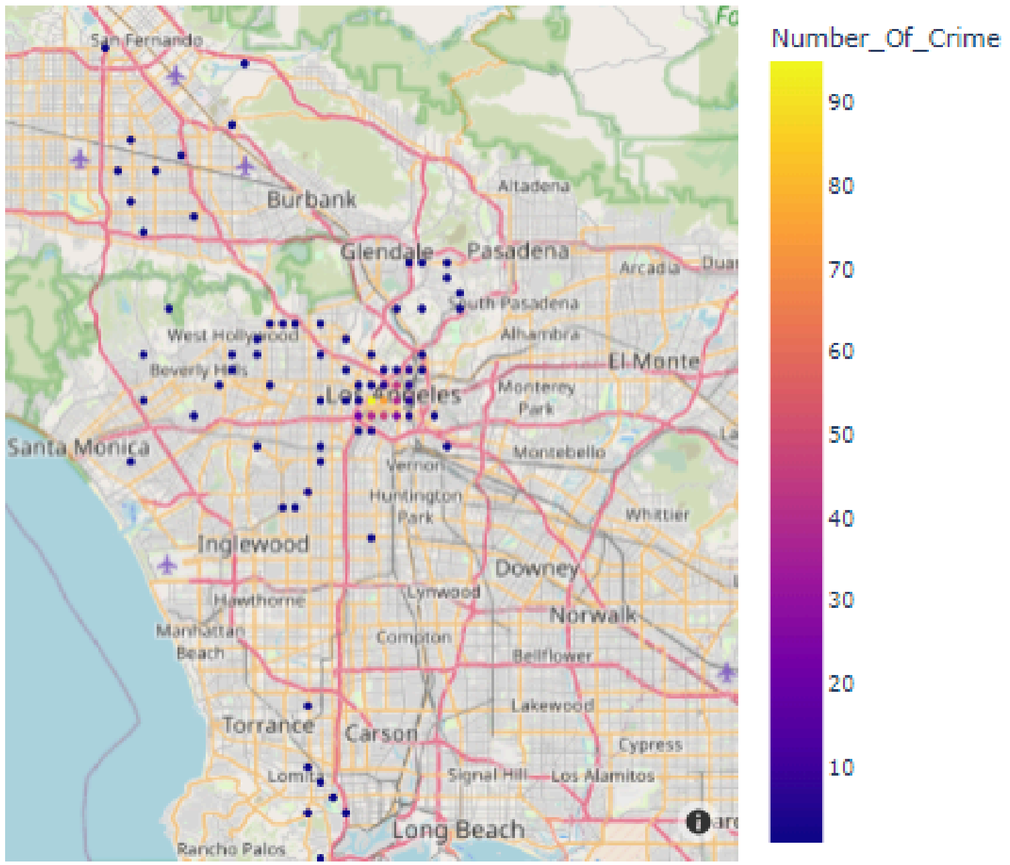
Q.1 Where are the geographical hotspots for reported crimes?

```
Q_1 = ('select LAT, LON ,count(*)as Number_Of_Crime from Crime_data GROUP BY LAT,LON
ORDER BY Number_Of_Crime desc ;')

cursor.execute(Q_1)
```

	LAT	LON	Number_Of_Crime
0	34.05	-118.26	95
1	34.05	-118.25	75
2	34.04	-118.25	47
3	34.06	-118.24	43
4	34.05	-118.24	39

```
In [33]: fig.update_layout(mapbox_style="open-street-map")
         fig.update_layout(margin={"r":10,"t":100,"l":10,"b":10})
         fig.show()
```



VICTIM DEMOGRAPHIC

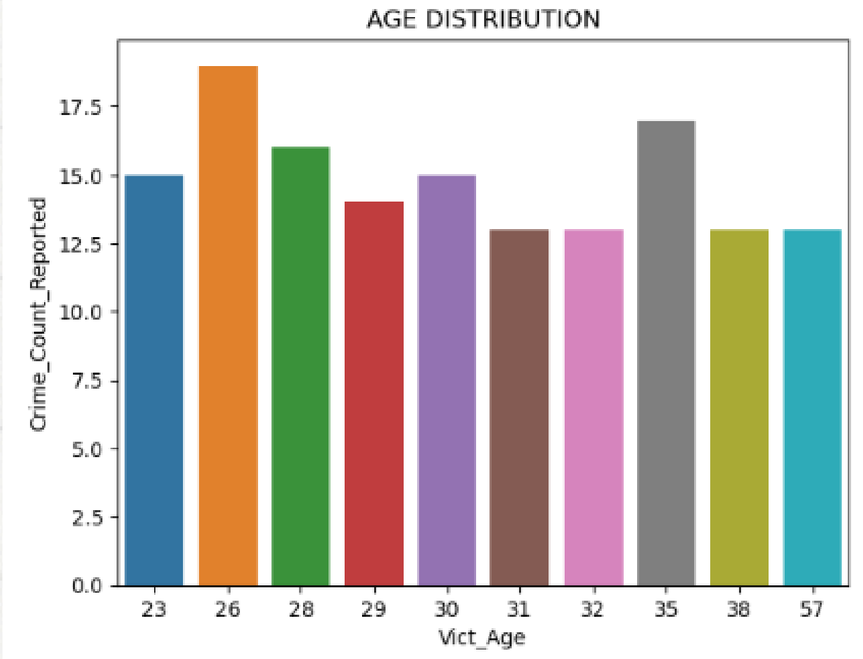
Q.2. What is the distribution of victim ages in reported crimes?

```
Q_2 = ('select Vict_Age,count(*) as Crime_Count_Reported from crime_data where Vict_age >10 group by
Vict_Age order by Crime_Count_Reported desc limit 10;')

cursor.execute(Q_2)
```

	Vict_Age	Crime_Count_Reported
0	26	19
1	35	17
2	28	16
3	23	15
4	30	15

```
sns.barplot(x="Vict_Age", y="Crime_Count_Reported" , data = df2)
plt.title('AGE DISTRIBUTION')
plt.show()
```



VICTIM DEMOGRAPHIC

Q.3. Is there a significant difference in crime rates between male and female victims?

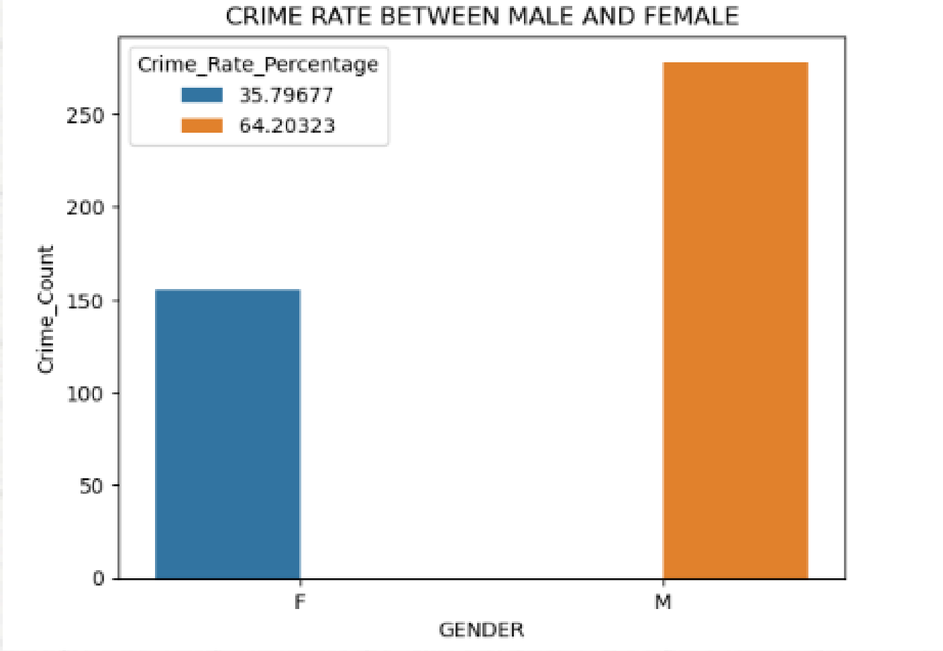
```
Q_3 = ("SELECT Vict_Sex AS GENDER, COUNT(*) AS Crime_Count,COUNT(*) * 100.0 / SUM(COUNT(*)) OVER () AS Crime_Rate_Percentage FROM crime_data WHERE Vict_Sex LIKE '%M%' OR Vict_Sex LIKE '%F%'GROUP BY Vict_Sex ORDER BY Vict_Sex;")

cursor.execute(Q_3)
```

	GENDER	Crime_Count	Crime_Rate_Percentage
0	F	155	35.79677
1	M	278	64.20323

```
sns.barplot(x="GENDER" , y="Crime_Count", data = df3,hue='Crime_Rate_Percentage')
plt.title('CRIME RATE BETWEEN MALE AND FEMALE')

plt.show()
```



STATUS ANALYSIS

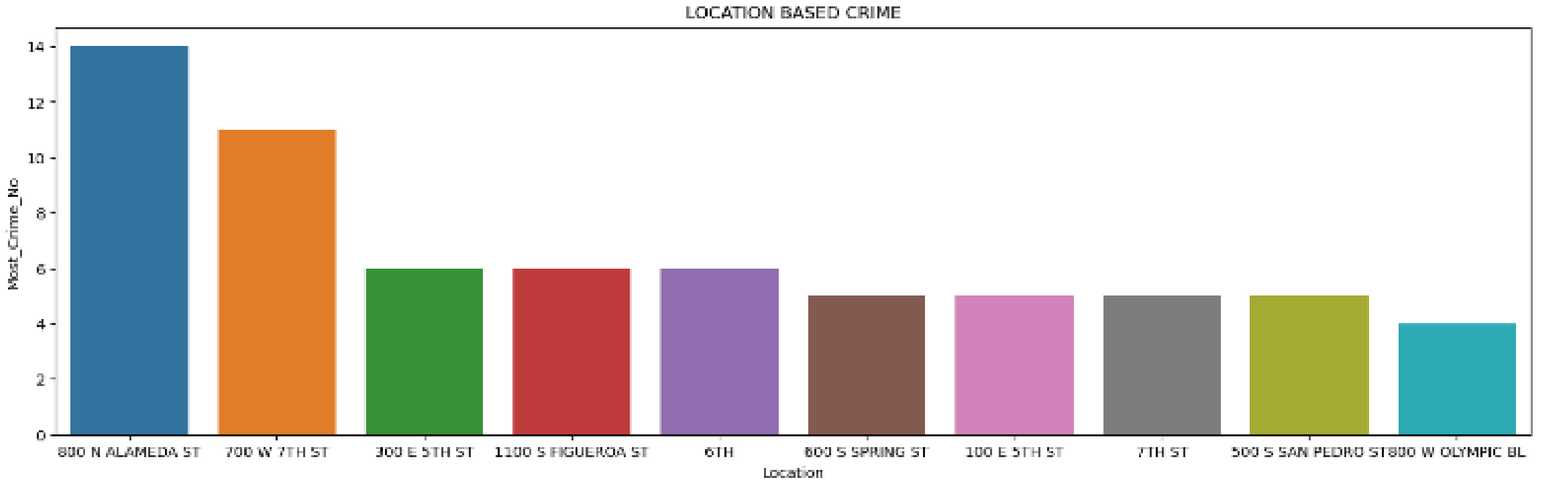
Q.4.Where do most crimes occur based on the "Location" column?

```
Q_4 = ("select Location , Count(*) as Most_Crime_No from crime_data group by Location
order by Most_Crime_No desc limit 10;")

cursor.execute(Q_4)
```

	Location	Most_Crime_No
0	800 N ALAMEDA ST	14
1	700 W 7TH ST	11
2	300 E 5TH ST	6
3	1100 S FIGUEROA ST	6
4	6TH	6
5	600 S SPRING ST	5
6	100 E 5TH ST	5
7	7TH ST	5
8	500 S SAN PEDRO ST	5
9	800 W OLYMPIC BL	4

```
plt.figure(figsize=(18, 5))
sns.barplot(x="Location" , y="Most_Crime_No", data = df4)
plt.title('LOCATION BASED CRIME')
plt.show()
```



STATUS ANALYSIS

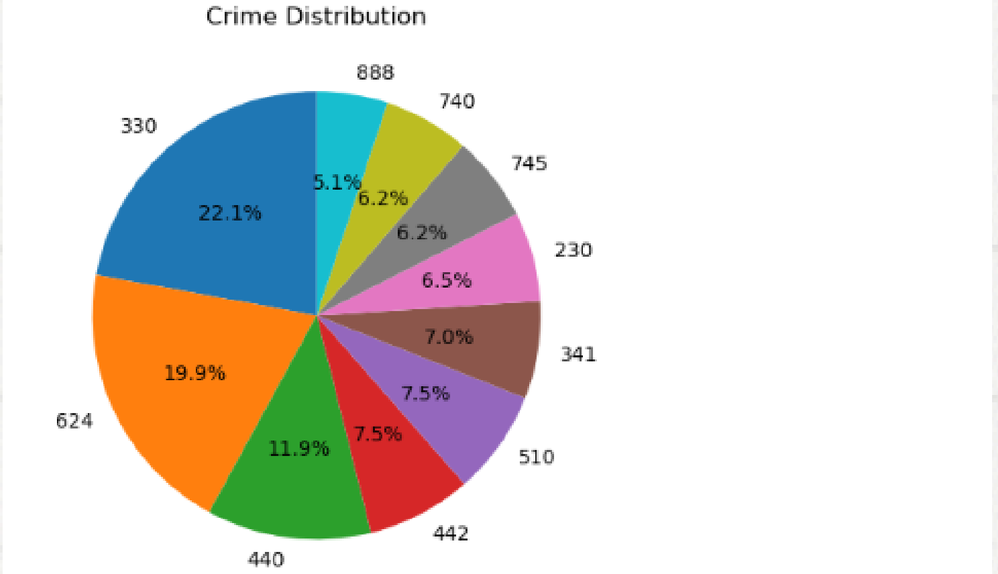
Q.5. What is the distribution of reported crimes based on Crime Code?

```
Q_5 = ('select Crm_Cd, count(Crm_Cd) as Reported_Crime from crime_data group by
Crm_Cd order by Reported_Crime desc limit 10;')

cursor.execute(Q_5)
```

	Crm_Cd	Reported_Crime
0	330	82
1	624	74
2	440	44
3	442	28
4	510	28
5	341	26
6	230	24
7	745	23
8	740	23
9	888	19

```
plt.figure(figsize=(5, 15))
plt.pie(df5['Reported_Crime'], labels=df5['Crm_Cd'], autopct='%1.1f%%', startangle=90)
plt.title('Crime Distribution')
plt.show()
```



TOOLS AND LIBRARIES

01

JUPYTER NOTEBOOK for Python Code development and MySQL Work

02

PyMySQL for interacting with MySQL database.

03

PANDAS Library for data frame, manipulation and analysis.

04

Matplotlib, Seaborn and Plotly Express for data visualization.

ANALYSIS

Based on Location, Crime Code 330, in Los Angeles has the HIGHEST Crime Number(95) Located at Latitude 34.04 and Longitude -118.26 happened and Male Gender has the HIGHEST Crime Count.

SUGGESTIONS

Increasing police patrols and community watch programs in the area around Latitude 34.04 and Longitude -118.26 in Los Angeles, as it has the highest reported crime number (95). Additionally, enhancing street lighting and installing surveillance cameras could help deter criminal activities in this hotspot.

Thank you very much!

Gmail Github