PROJECT: LA CRIME ANALYTICS

INTRODUCTION

In the era of increasing urbanization and social complexity, understanding crime patterns through data has become crucial for improving public safety and supporting informed decision-making. This Python-based project, titled “Crime Analytics Dashboard”, aims to provide meaningful insights into crime trends using real-world crime data.

By leveraging libraries such as Pandas, Matplotlib, Seaborn, and NumPy, the project processes and visualizes data extracted from crime records, with the objective of uncovering significant patterns in criminal activities, victim demographics, and geographical distribution.

Key components of the analysis include:

- Identification of top crime types and premises where these crimes occur most frequently.
- Study of victim demographics, including age, gender, and descent distribution.
- Analysis of weapon usage, offering a deeper understanding of crime severity and risk levels.
- Visualization of crime locations using latitude and longitude data to identify hotspots across regions.
- Use of correlation heatmaps to discover interrelationships between numerical variables such as victim age and crime frequency.
- Advanced statistical breakdowns to highlight the most dangerous locations, top crimes by victim descent, and a cross-tab view of crime types vs. weapons.

The project offers multiple visual representations such as bar charts, pie charts, histograms, box plots, violin plots, scatter plots, and heatmaps—making the insights not only informative but also visually engaging and easy to interpret.

This interactive analytical framework serves as a foundational tool for law enforcement agencies, data scientists, and policy makers to evaluate trends, enhance safety strategies, and support data-driven crime prevention measures.

ANALYSIS ON DATASET

EDA PROCESS:

Exploratory Data Analysis (EDA) is a critical phase in any data science workflow. It helps uncover initial insights, detect anomalies, identify patterns, and prepare data for deeper analysis.

1. **Data Loading**:
   Using pandas.read_csv(), the dataset was read from the local path into a DataFrame for manipulation.

2. Data Cleaning:

- Removed duplicate records using drop_duplicates() to maintain unique entries.
- Stripped column name whitespaces with df.columns.str.strip() for consistency in referencing.
- Checked and printed null values per column using df.isna().sum() to identify missing data.
- Retained only rows with valid coordinates for location-based plotting (LAT, LON).

3. Visualization Setup:

- Set Seaborn’s theme to whitegrid for clean, consistent visuals.
- Used diverse plot types:
  - **Bar and Pie Charts**: Top crime types and premises.
  - **Histogram**: Victim age distribution.
  - **Box & Violin Plots**: Age variation by victim sex.
  - **Heatmap**: Correlation among numeric columns.
  - **Scatter Plot**: Crime hotspots via latitude and longitude.

4. Analysis Readiness:

- Ensured all essential columns (e.g., Crm Cd Desc, Vict Age, Vict Sex, LAT, LON) were free from critical missing values.
- Transformed data into appropriate formats for grouped aggregations and cross-tab analysis.
- Ready for insights generation on crime frequency, demographic impact, and spatial distribution.

Objective 1: Identifying the most common Crime Types

i. General Description

This objective focuses on identifying and analyzing the most frequently occurring crimes in Los Angeles. By understanding which crimes are most common, authorities can allocate resources efficiently, and the public can stay informed about prevalent threats.

ii. Specific Requirements

- Detailed crime type data (Crm Cd Desc column)
- Aggregation of occurrences by crime type
- Visualization to compare frequencies among the top reported crimes

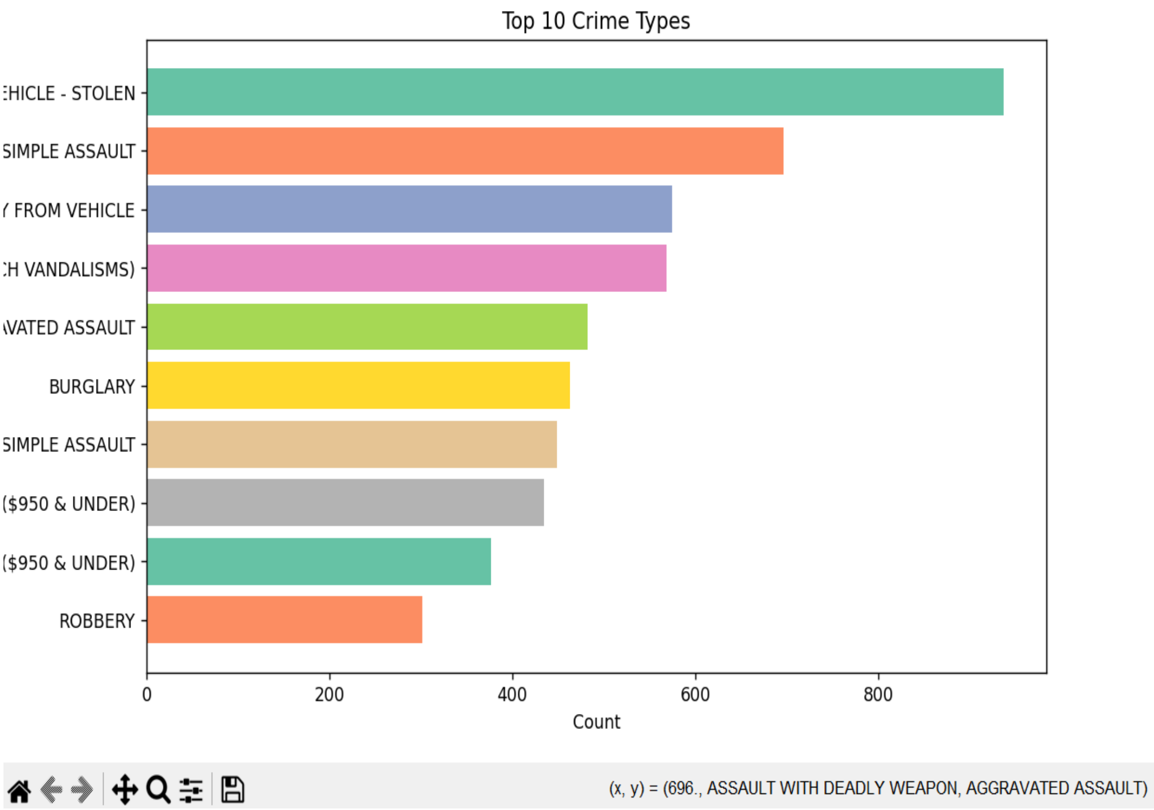
iii. Analysis Results

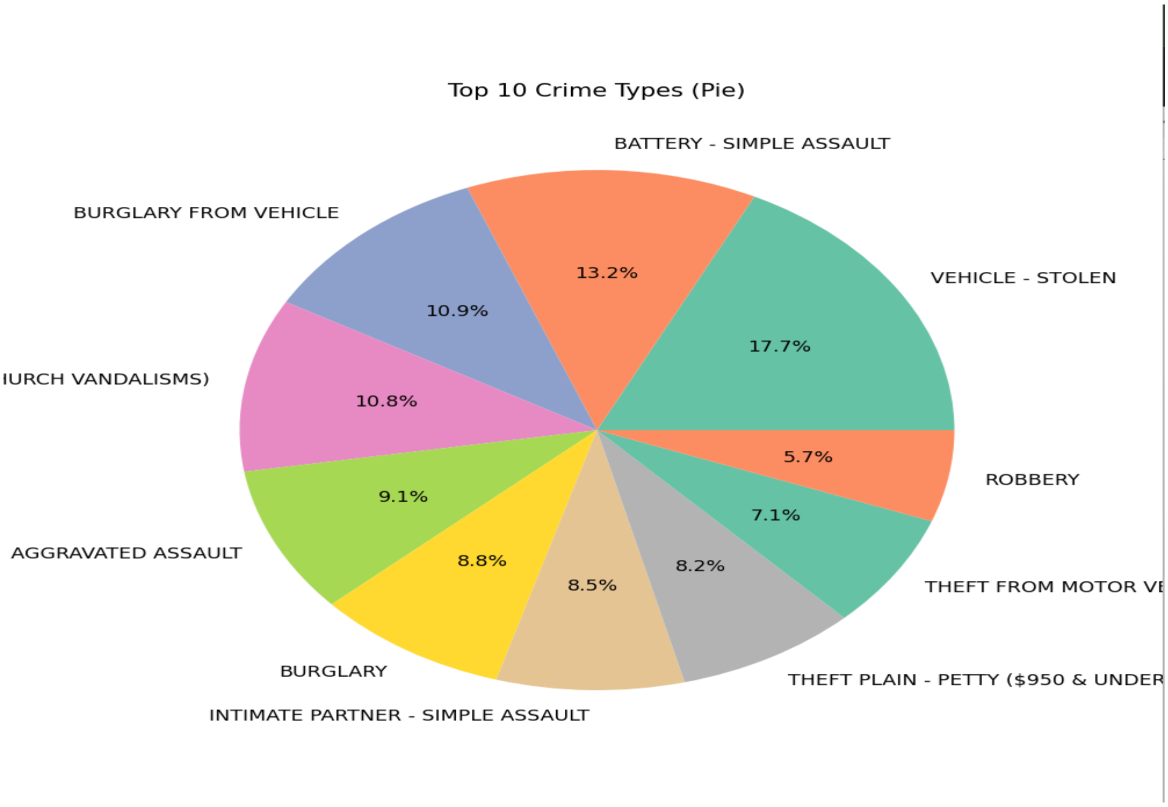
The analysis revealed that crimes like *Battery - Simple Assault*, *Vehicle Theft*, and *Burglary* are among the most common in Los Angeles. This frequency insight provides a basis for developing crime-prevention strategies and raising public awareness in affected areas.

iv. Visualization

Bar and pie charts were used to represent the **top 10 crime types** by frequency.

- **Bar Chart** allowed for direct comparison of absolute numbers of crimes.
- **Pie Chart** offered a proportional view to understand how each crime contributes to the overall distribution.





Objective 2: Understanding Age Distribution of Victims

i. General Description

This objective aims to analyze the age distribution of crime victims in Los Angeles. By identifying which age groups are most frequently targeted, this analysis supports the development of focused safety campaigns and community outreach initiatives.

ii. Specific Requirements

- Cleaned and validated data from the Vict Age column
- Removal of missing or unrealistic age values
- Aggregated age data using appropriate binning techniques

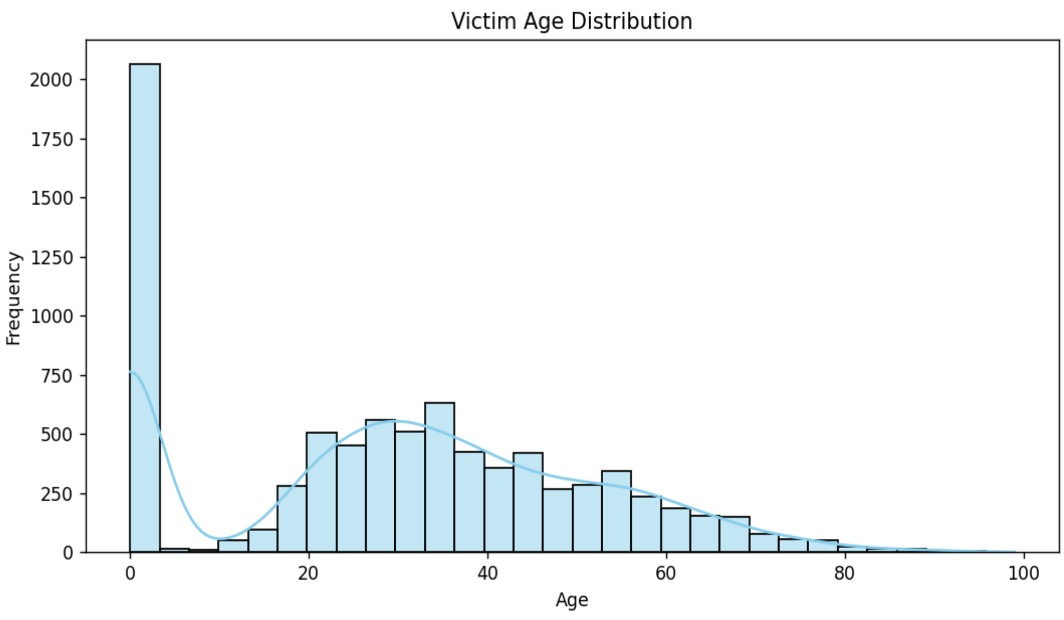
iii. Analysis Results

The results indicated that individuals aged between **20 and 40 years** are most commonly victimized, followed by teenagers and middle-aged adults. This insight emphasizes the need for protective measures tailored to these vulnerable groups.

iv. Visualization

A **Histogram with KDE (Kernel Density Estimation) plot** was used to visualize the distribution:

- The **Histogram** shows the frequency of victims in various age brackets.
- The **KDE Plot** overlays a smoothed curve, helping to identify age peaks and overall trends in the data.



-

Objective 3: Analyzing Victim Age Based on Gender

i. General Description

This objective aims to explore how victim age distribution varies across different genders. Understanding this pattern can help authorities design gender-sensitive crime prevention strategies and support systems.

ii. Specific Requirements

- Clean and complete data in Vict Sex and Vict Age columns
- Grouping data based on gender
- Removing invalid or missing age entries for accurate distribution analysis

iii. Analysis Results

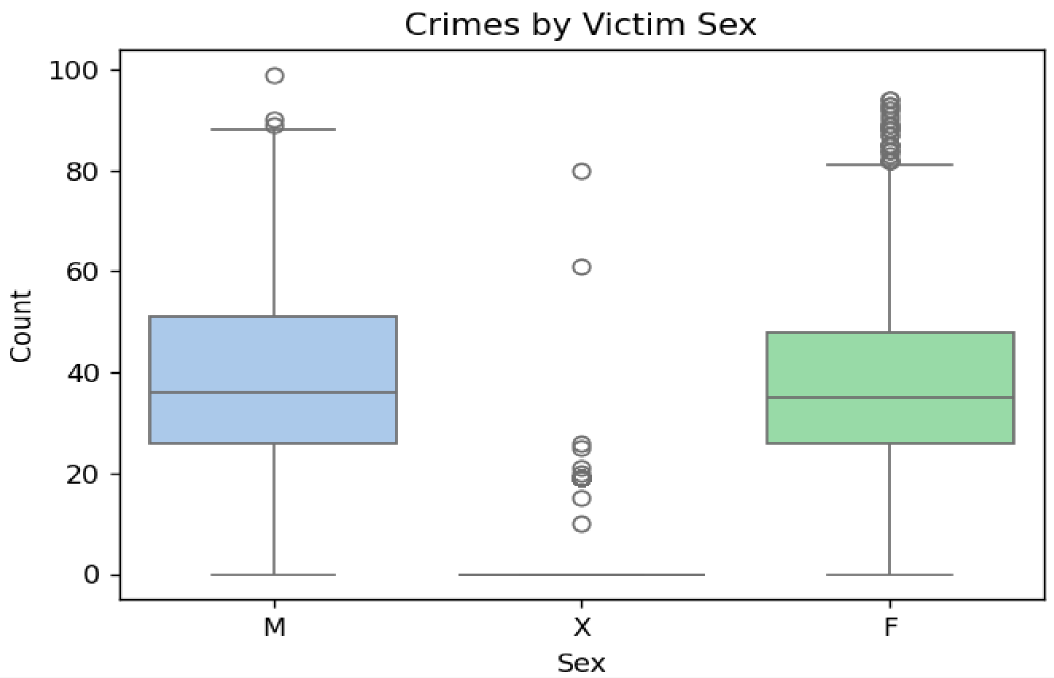
The box plot highlighted that male and female victims show different age distribution patterns. Males had a wider range of ages affected by crimes, whereas females had more concentration in certain age groups. Outliers were also visible, indicating occasional crimes affecting extreme age ranges.

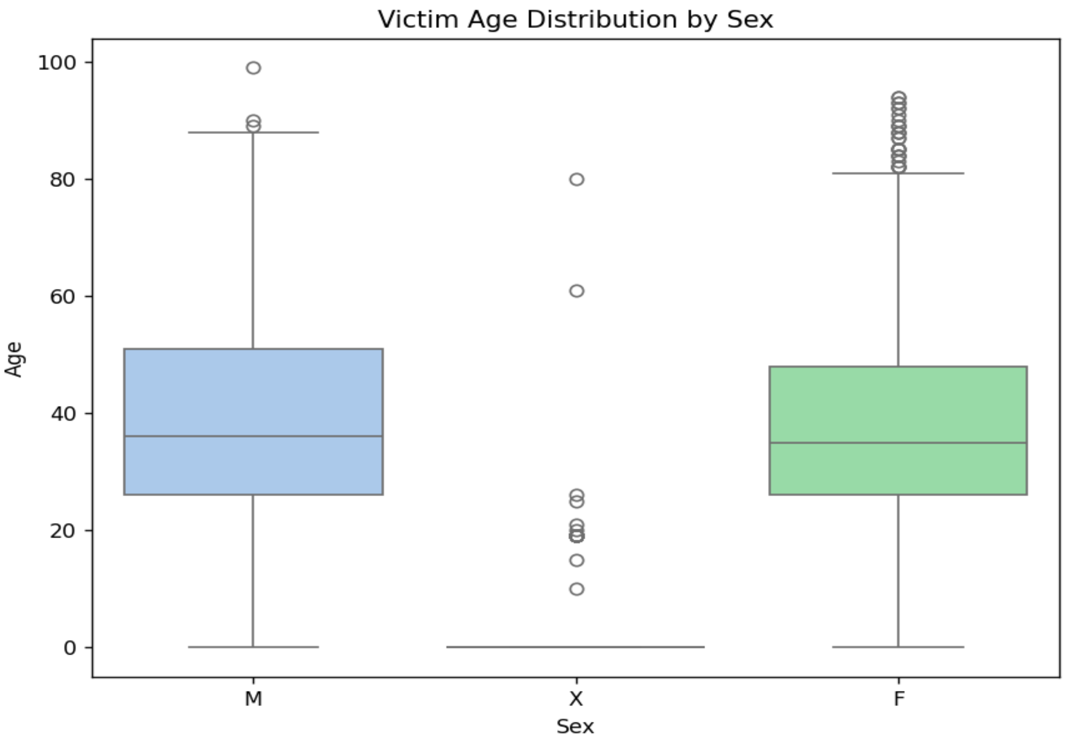
iv. Visualization

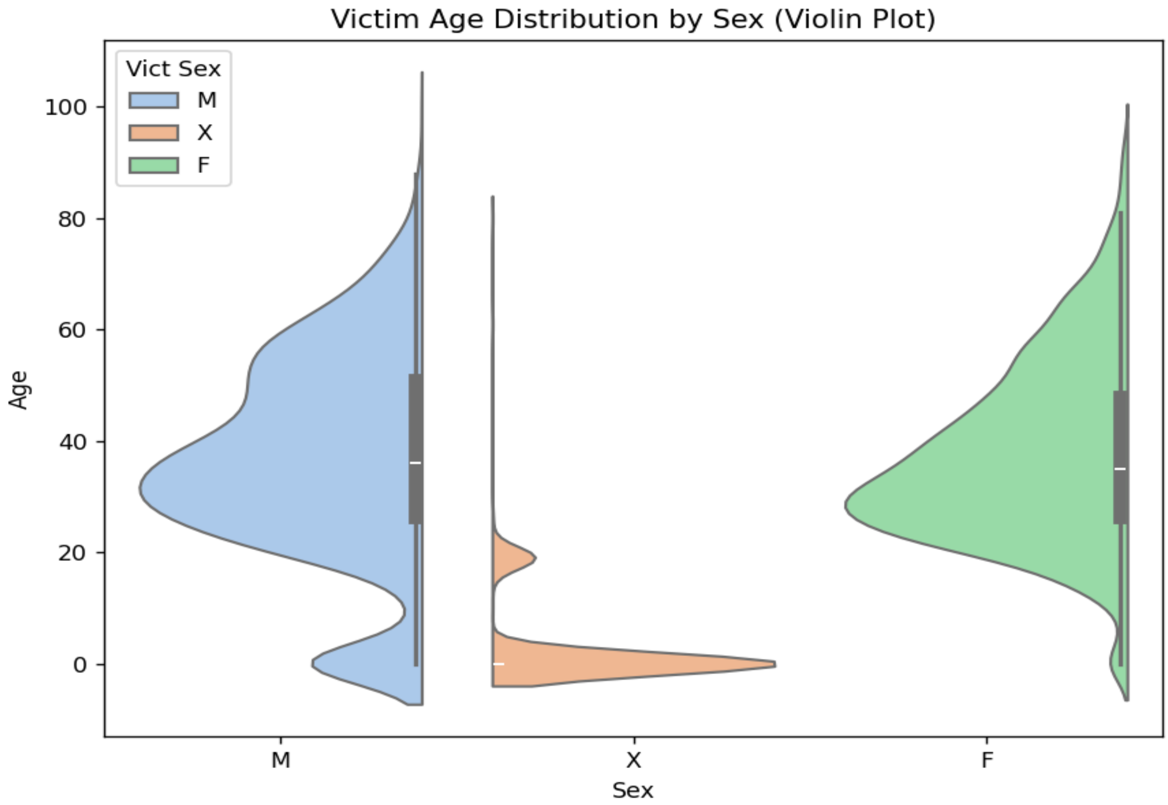
A Box Plot and Violin Plot was used to compare victim age across genders:

- It shows the median, interquartile range, and outliers for each gender.
- Helps quickly identify which age groups are most affected and how spread out the data is.

It provides a clear view of crime concentration and aids in comparative risk assessment across locations


Crimes by Victim Sex
Count
100
80
60
40
20
0
M
X
F
Sex


Victim Age Distribution by Sex
Age
100
80
60
40
20
0
M
X
F
Sex



Objective 4: Exploring Correlations in Crime Data

i. General Description

This objective seeks to uncover potential relationships between numerical features in the dataset, such as victim age, time of occurrence, and crime-related variables. These correlations may reveal underlying behavioural or situational patterns.

ii. Specific Requirements

- Selection of only numeric columns from the dataset
- Removal of missing or non-numeric values
- Application of correlation matrix techniques to measure linear relationships

iii. Analysis Results

The correlation heatmap identified weak to moderate correlations among numeric fields. While most variables showed low correlation, certain fields like victim age and crime resolution time showed interesting trends that can be investigated further.

iv. Visualization

A **Correlation Heatmap** was used:

- Displays Pearson correlation values between numeric features
- Annotated and color-coded for easier interpretation
- Helps in feature selection for predictive modelling or further statistical analysis

Objective 5: Mapping Geographical Distribution of Crimes

i. General Description

This objective focuses on identifying where crimes occur most frequently within Los Angeles. By analyzing spatial patterns, law enforcement can allocate resources more effectively and proactively monitor high-risk areas.

ii. Specific Requirements

- Latitude (LAT) and Longitude (LON) data points without missing values
- Removal of invalid or zeroed coordinates
- Visualization with spatial scatter plots to display crime density

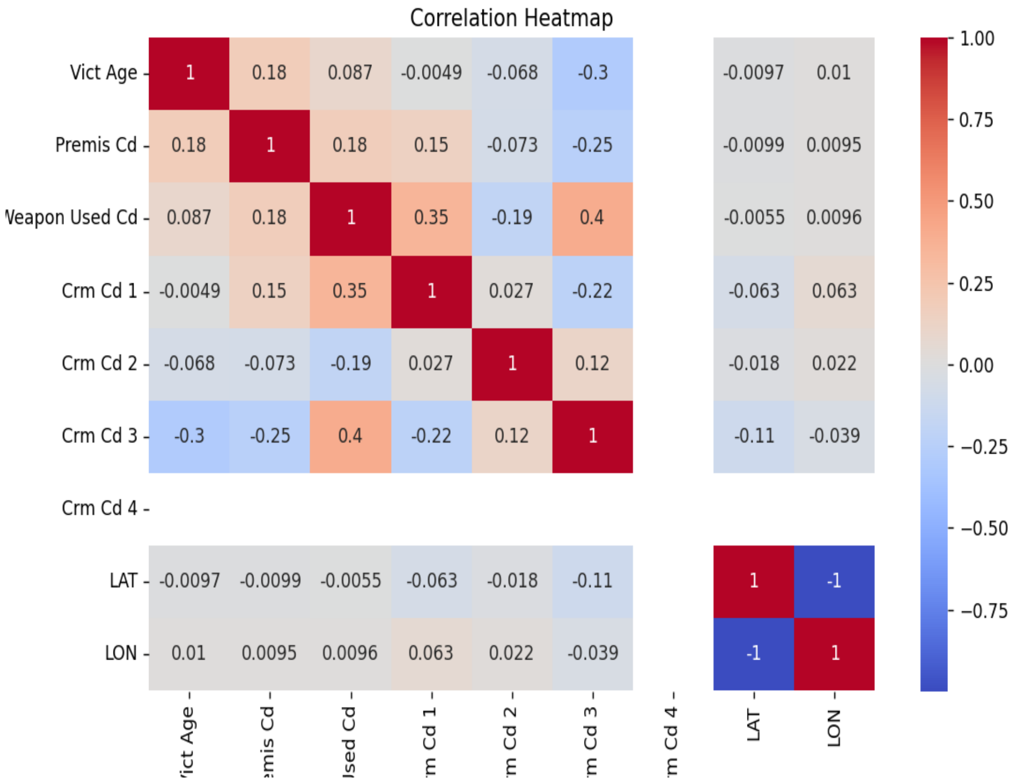
iii. Analysis Results

The scatter plot revealed crime hotspots, particularly in central and downtown Los Angeles. Densely clustered points indicate areas with high crime volume, suggesting a need for increased surveillance and community outreach in those regions.

iv. Visualization

A **Scatter Plot using Latitude and Longitude** was implemented:

- Plots crime locations as points on a coordinate plane
- Makes spatial clusters of crime immediately visible
- Supports geographic targeting for patrol planning and community safety initiatives


Correlation Heatmap
Vict Age
Premis Cd
Weapon Used Cd
Crm Cd 1
Crm Cd 2
Crm Cd 3
Crm Cd 4
LAT
LON
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75

```
Data Shape: (8249, 19)
Data Info:
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 8249 entries, 0 to 8248
Data columns (total 19 columns):
 #   Column           Non-Null Count  Dtype
---  ------           --------------  -----
 0   Crm Cd Desc      8249 non-null   object
 1   Mocodes          7097 non-null   object
 2   Vict Age         8249 non-null   int64
 3   Vict Sex         7133 non-null   object
 4   Vict Descent     7132 non-null   object
 5   Premis Cd        8249 non-null   int64
 6   Premis Desc      8244 non-null   object
 7   Weapon Used Cd   3136 non-null   float64
 8   Weapon Desc      3136 non-null   object
 9   Status           8249 non-null   object
 10  Status Desc      8249 non-null   object
 11  Crm Cd 1         8248 non-null   float64
 12  Crm Cd 2         704 non-null    float64
 13  Crm Cd 3         17 non-null     float64
 14  Crm Cd 4         1 non-null      float64
 15  LOCATION         8249 non-null   object
 16  Cross Street     1564 non-null   object
 17  LAT              8249 non-null   float64
 18  LON              8249 non-null   float64
dtypes: float64(7), int64(2), object(10)
memory usage: 1.2+ MB
Summary Statistics:
             Crm Cd Desc Mocodes  ...          LAT          LON
count               8249    7097  ...  8249.000000  8249.000000
unique                96    4900  ...          NaN          NaN
top      VEHICLE - STOLEN     344  ...          NaN          NaN
freq                 948     276  ...          NaN          NaN
mean                 NaN     NaN  ...    33.942420  -117.965834
std                  NaN     NaN  ...     1.874142     6.505045
min                  NaN     NaN  ...     0.000000  -118.565100
25%                  NaN     NaN  ...    34.017300  -118.361900
50%                  NaN     NaN  ...    34.050900  -118.300300
75%                  NaN     NaN  ...    34.096000  -118.264300
max                  NaN     NaN  ...    34.221200     0.000000

[11 rows x 19 columns]
```

```
count                    8249   7097  ...  8249.000000  8249.000000
unique                     96   4900  ...          NaN          NaN
top      VEHICLE - STOLEN   344  ...          NaN          NaN
freq                      948    276  ...          NaN          NaN
mean                      NaN    NaN  ...    33.942420  -117.965834
std                       NaN    NaN  ...     1.874142     6.505045
min                       NaN    NaN  ...     0.000000  -118.565100
25%                       NaN    NaN  ...    34.017300  -118.361900
50%                       NaN    NaN  ...    34.050900  -118.300300
75%                       NaN    NaN  ...    34.096000  -118.264300
max                       NaN    NaN  ...    34.221200     0.000000

[11 rows x 19 columns]
NaN Count per Column:
Crm Cd Desc            0
Mocodes             1141
Vict Age               0
Vict Sex            1105
Vict Descent        1106
Premis Cd              0
Premis Desc            5
Weapon Used Cd      5097
Weapon Desc         5097
Status                 0
Status Desc            0
Crm Cd 1               1
Crm Cd 2            7529
Crm Cd 3            8216
Crm Cd 4            8232
LOCATION               0
Cross Street        6670
LAT                    0
LON                    0
dtype: int64
Top 10 Crime Types:
Crm Cd Desc
VEHICLE - STOLEN                                           937
BATTERY - SIMPLE ASSAULT                                   696
BURGLARY FROM VEHICLE                                      575
VANDALISM - FELONY ($400 & OVER, ALL CHURCH VANDALISMS)    568
ASSAULT WITH DEADLY WEAPON, AGGRAVATED ASSAULT             482
BURGLARY                                                   463
INTIMATE PARTNER - SIMPLE ASSAULT                          449
THEFT PLAIN - PETTY ($950 & UNDER)                         434
THEFT FROM MOTOR VEHICLE - PETTY ($950 & UNDER)            377
ROBBERY                                                    301
Name: count, dtype: int64
```

```
UNKNOWN FIREARM                                          60
OTHER KNIFE                                              52
BOTTLE                                                   35
ROCK/THROWN OBJECT                                       34
Name: count, dtype: int64
Crime Status Summary:
Status Desc
Invest Cont     6230
Adult Other     1108
Adult Arrest     854
Juv Arrest        34
Juv Other          7
Name: count, dtype: int64

============== RESTART: C:\Users\hp\OneDrive\Desktop\project\r.py ==
Data Shape: (8249, 19)
Data Info:
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 8249 entries, 0 to 8248
Data columns (total 19 columns):
 #   Column          Non-Null Count  Dtype
---  ------          --------------  -----
 0   Crm Cd Desc     8249 non-null   object
 1   Mocodes         7097 non-null   object
 2   Vict Age        8249 non-null   int64
 3   Vict Sex        7133 non-null   object
 4   Vict Descent    7132 non-null   object
 5   Premis Cd       8249 non-null   int64
 6   Premis Desc     8244 non-null   object
 7   Weapon Used Cd  3136 non-null   float64
 8   Weapon Desc     3136 non-null   object
 9   Status          8249 non-null   object
 10  Status Desc     8249 non-null   object
 11  Crm Cd 1        8248 non-null   float64
 12  Crm Cd 2        704 non-null    float64
 13  Crm Cd 3        17 non-null     float64
 14  Crm Cd 4        1 non-null      float64
 15  LOCATION        8249 non-null   object
 16  Cross Street    1564 non-null   object
 17  LAT             8249 non-null   float64
 18  LON             8249 non-null   float64
dtypes: float64(7), int64(2), object(10)
memory usage: 1.2+ MB
Summary Statistics:
             Crm Cd Desc Mocodes  ...          LAT          LON
count               8249    7097  ...  8249.000000  8249.000000
unique                96    4900  ...          NaN          NaN
```

```
DRIVEWAY                                                  145
PARKING UNDERGROUND/BUILDING                              131
Name: count, dtype: int64
Victim Sex Distribution:
Vict Sex
M    3422
F    2983
X     723
Name: count, dtype: int64
Victim Descent Distribution:
Vict Descent
H    2445
W    1767
B    1286
X     764
O     597
A     188
K      20
C      19
F      18
V       9
I       8
J       3
P       2
Z       1
Name: count, dtype: int64
Victim Age Stats:
count    8233.00000
mean       29.36876
std        21.67559
min         0.00000
25%         2.00000
50%        30.00000
75%        45.00000
max        99.00000
Name: Vict Age, dtype: float64
Top Weapon Types:
Weapon Desc
STRONG-ARM (HANDS, FIST, FEET OR BODILY FORCE)    1699
UNKNOWN WEAPON/OTHER WEAPON                        347
VERBAL THREAT                                      198
HAND GUN                                           179
KNIFE WITH BLADE 6INCHES OR LESS                    76
SEMI-AUTOMATIC PISTOL                               63
UNKNOWN FIREARM                                     60
OTHER KNIFE                                         52
```

```
[11 rows x 19 columns]
NaN Count per Column:
Crm Cd Desc          0
Mocodes           1141
Vict Age             0
Vict Sex          1105
Vict Descent      1106
Premis Cd            0
Premis Desc          5
Weapon Used Cd    5097
Weapon Desc       5097
Status               0
Status Desc          0
Crm Cd 1             1
Crm Cd 2          7529
Crm Cd 3          8216
Crm Cd 4          8232
LOCATION             0
Cross Street      6670
LAT                  0
LON                  0
dtype: int64
Top 10 Crime Types:
Crm Cd Desc
VEHICLE - STOLEN                                            937
BATTERY - SIMPLE ASSAULT                                    696
BURGLARY FROM VEHICLE                                       575
VANDALISM - FELONY ($400 & OVER, ALL CHURCH VANDALISMS)     568
ASSAULT WITH DEADLY WEAPON, AGGRAVATED ASSAULT              482
BURGLARY                                                    463
INTIMATE PARTNER - SIMPLE ASSAULT                           449
THEFT PLAIN - PETTY ($950 & UNDER)                          434
THEFT FROM MOTOR VEHICLE - PETTY ($950 & UNDER)             377
ROBBERY                                                     301
Name: count, dtype: int64
Top 10 Crime Premises:
Premis Desc
STREET                                          2200
SINGLE FAMILY DWELLING                          1234
MULTI-UNIT DWELLING (APARTMENT, DUPLEX, ETC)    1059
PARKING LOT                                      557
SIDEWALK                                         449
OTHER BUSINESS                                   362
VEHICLE, PASSENGER/TRUCK                         276
GARAGE/CARPORT                                   161
DRIVEWAY                                         145
```

```
ROBBERY
Name: count, dtype: int64
Top 10 Crime Premises:
Premis Desc
STREET                                          2200
SINGLE FAMILY DWELLING                          1234
MULTI-UNIT DWELLING (APARTMENT, DUPLEX, ETC)    1059
PARKING LOT                                      557
SIDEWALK                                         449
OTHER BUSINESS                                   362
VEHICLE, PASSENGER/TRUCK                         276
GARAGE/CARPORT                                   161
DRIVEWAY                                         145
PARKING UNDERGROUND/BUILDING                     131
Name: count, dtype: int64
Victim Sex Distribution:
Vict Sex
M    3422
F    2983
X     723
Name: count, dtype: int64
Victim Descent Distribution:
Vict Descent
H    2445
W    1767
B    1286
X     764
O     597
A     188
K      20
C      19
F      18
V       9
I       8
J       3
P       2
Z       1
Name: count, dtype: int64
Victim Age Stats:
count    8233.00000
mean       29.36876
std        21.67559
min         0.00000
25%         2.00000
50%        30.00000
75%        45.00000
```

```
75%          45.00000
max          99.00000
Name: Vict Age, dtype: float64
Top Weapon Types:
Weapon Desc
STRONG-ARM (HANDS, FIST, FEET OR BODILY FORCE)    1699
UNKNOWN WEAPON/OTHER WEAPON                        347
VERBAL THREAT                                      198
HAND GUN                                           179
KNIFE WITH BLADE 6INCHES OR LESS                    76
SEMI-AUTOMATIC PISTOL                               63
UNKNOWN FIREARM                                     60
OTHER KNIFE                                         52
BOTTLE                                              35
ROCK/THROWN OBJECT                                  34
Name: count, dtype: int64
Crime Status Summary:
Status Desc
Invest Cont     6230
Adult Other     1108
Adult Arrest     854
Juv Arrest        34
Juv Other          7
Name: count, dtype: int64

============== RESTART: C:\Users\hp\OneDrive\Desktop\project\r.py =
Data Shape: (8249, 19)
Data Info:
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 8249 entries, 0 to 8248
Data columns (total 19 columns):
 #   Column          Non-Null Count  Dtype
---  ------          --------------  -----
 0   Crm Cd Desc     8249 non-null   object
 1   Mocodes         7097 non-null   object
 2   Vict Age        8249 non-null   int64
 3   Vict Sex        7133 non-null   object
 4   Vict Descent    7132 non-null   object
 5   Premis Cd       8249 non-null   int64
 6   Premis Desc     8244 non-null   object
 7   Weapon Used Cd  3136 non-null   float64
 8   Weapon Desc     3136 non-null   object
 9   Status          8249 non-null   object
 10  Status Desc     8249 non-null   object
 11  Crm Cd 1        8248 non-null   float64
 12  Crm Cd 2        704 non-null    float64
 13  Crm Cd 3        17 non-null     float64
```

```
ROBBERY                                                                301
Name: count, dtype: int64
Top 10 Crime Premises:
Premis Desc
STREET                                              2200
SINGLE FAMILY DWELLING                              1234
MULTI-UNIT DWELLING (APARTMENT, DUPLEX, ETC)        1059
PARKING LOT                                          557
SIDEWALK                                             449
OTHER BUSINESS                                       362
VEHICLE, PASSENGER/TRUCK                             276
GARAGE/CARPORT                                       161
DRIVEWAY                                             145
PARKING UNDERGROUND/BUILDING                         131
Name: count, dtype: int64
Victim Sex Distribution:
Vict Sex
M    3422
F    2983
X     723
Name: count, dtype: int64
Victim Descent Distribution:
Vict Descent
H    2445
W    1767
B    1286
X     764
O     597
A     188
K      20
C      19
F      18
V       9
I       8
J       3
P       2
Z       1
Name: count, dtype: int64
Victim Age Stats:
count    8233.00000
mean       29.36876
std        21.67559
min         0.00000
25%         2.00000
50%        30.00000
75%        45.00000
max        99.00000
```

```
Name: count, dtype: int64
Most Dangerous Locations:
LOCATION
6TH                            ST           27
6TH                                         20
5TH                                         18
3RD                            ST           16
7TH                                         16
WESTERN                        AV           15
800 N  ALAMEDA                        ST    15
7TH                            ST           15
SLAUSON                        AV           14
VERMONT                                     12
Name: count, dtype: int64
Top Crime Type per Victim Descent:
Vict Descent
A                                 BURGLARY FROM VEHICLE
B        ASSAULT WITH DEADLY WEAPON, AGGRAVATED ASSAULT
C                                     THEFT OF IDENTITY
F     THEFT FROM MOTOR VEHICLE - GRAND ($950.01 AND ...
H                              BATTERY - SIMPLE ASSAULT
I                                         BIKE - STOLEN
J                      THEFT PLAIN - PETTY ($950 & UNDER)
K                                 BURGLARY FROM VEHICLE
O                                 BURGLARY FROM VEHICLE
P                                         BIKE - STOLEN
V     THEFT FROM MOTOR VEHICLE - GRAND ($950.01 AND ...
W                                 BURGLARY FROM VEHICLE
X     VANDALISM - FELONY ($400 & OVER, ALL CHURCH VA...
Z                                 BURGLARY FROM VEHICLE
Name: Crm Cd Desc, dtype: object
Cross-tab: Crime Type vs Weapon
Weapon Desc                                     AIR PISTOL/REVOLVER/RIFLE/BB GUN  ...  VERBAL THREAT
Crm Cd Desc                                                                       ...
ARSON                                                                          0  ...              1
ASSAULT WITH DEADLY WEAPON ON POLICE OFFICER                                   0  ...              0
ASSAULT WITH DEADLY WEAPON, AGGRAVATED ASSAULT                                 9  ...              1
ATTEMPTED ROBBERY                                                              0  ...              4
BATTERY - SIMPLE ASSAULT                                                       0  ...              0
BATTERY POLICE (SIMPLE)                                                        0  ...              1
BATTERY WITH SEXUAL CONTACT                                                    0  ...              0
BOMB SCARE                                                                     0  ...              1
BRANDISH WEAPON                                                                2  ...              8
BUNCO, GRAND THEFT                                                             0  ...              0

[10 rows x 58 columns]
```