My grades for **Project progress report**

Progress

10 / 10

Here you submit your progress report (a draft of your final project report, with placeholders to fill in for the latter) in pdf format

DALHOUSIE UNIVERSITY

Crime Zone Classifier(CZC) Project progress report

Submitted by:

Venkata Sreenivas Prasad Kasibhatla

B00972626 (vn769140@dal.ca)

Jahnavi Gajjala

B00969773 (jh707761@dal.ca)

Sichen Qiu

B00839638 (sc985823@dal.ca)

Submitted in fulfilment of the requirements for CSCI 4146&6409 to:

Prof. Evangelos Milios

Faculty of Computer Science

Dalhousie University

Halifax, Nova Scotia

February 26th. 2024

Abstract:

This project employs data science methodologies to confront the intricate challenge posed by city crime, with the objective of bolstering public safety, and contributing insights into urban planning and community development strategies.Based on analysis of dataset of crime reports procured from the Los Angeles Police Department, this project aims to discern and elucidate areas of pronounced criminal activity, cluster areas having criminal activity into into very high, high, medium, and low risk based on crime density, and furnish law enforcement agencies and policy formulators with practical recommendations. By leveraging advanced clustering algorithms and predictive analytics, this initiative is directed towards the optimization of resource distribution and the promotion of policies grounded in empirical evidence. Through the use of clustering algorithms and predictive analytics, this project is focused on optimizing the allocation of resources and promoting policies based on empirical evidence, to highlight the crime frequent zones and address the root socio-economic factors influencing crime trends.

Introduction:

Urban crime is a complex issue that has an impact on people's safety and quality of life in cities all over the world. Effective management of crime is a difficulty for law enforcement authorities due to its dynamic character, which is influenced by a variety of socio-economic factors. The project aims to leverage geospatial analysis to identify and visualize crime hotspots within a city and group these into hotspots using threshold based classification and density-based clustering techniques. This analysis will extend to categorize hotspots into very high, high, medium, and low risk based on crime density. By employing advanced data science techniques, we aim to analyze patterns in crime data, offering actionable insights that can aid in preventing future incidents.

Police can focus their efforts on areas that need the greatest attention by identifying crime hotspots. They can also potentially prevent crime by being strategically present in the community and engaging with the locals. The advantages also apply to community development and urban planning, enabling well-informed choices that can improve the quality of life in affected neighborhoods.

This initiative's main goal is to improve public safety by using data to inform decision-making. Given the resource limitations that law enforcement agencies frequently have, it is critical to focus resources in the regions where crime is most prevalent. Predictive analytics can be used in addition to traditional reactive approaches to crime prevention, enabling a more planned use of police resources. Furthermore, by comprehending the spatial distribution of crime, policymakers in the fields of social services and urban development can better address the root causes of crime.

Approach:

In this project, we initially performed data cleaning on the LAPD Crime records, data visualization and analysis of the visualization, and grouping the areas of criminal activity into very high risk, high risk, low risk and medium risk areas using threshold based classification.

Module Explanation:

Module 1 Data Cleaning:

- Drop columns with high percentage of missing values:
  Hence, We dropped the 'Crm Cd 2', 'Crm Cd 3', 'Crm Cd 4' columns.

```
# Check for missing values
print(data.isnull().sum())

DR_NO                  0
Date Rptd              0
DATE OCC               0
TIME OCC               0
AREA                   0
AREA NAME              0
Rpt Dist No            0
Part 1-2               0
Crm Cd                 0
Crm Cd Desc            0
Mocodes           125590
Vict Age               0
Vict Sex          119391
Vict Descent      119399
Premis Cd             10
Premis Desc          550
Weapon Used Cd    589089
Weapon Desc       589089
Status                 0
Status Desc            0
Crm Cd 1              11
Crm Cd 2          835674
Crm Cd 3          899144
Crm Cd 4          901293
LOCATION               0
Cross Street      759302
LAT                    0
LON                    0
dtype: int64
```

Figure 1: Number of null values in each of the columns

Make sure to change it to a proper table in the final report

- Filled missing values in **numerical** columns using the **median**.
- Filled missing values in **categorical** columns using the **mode**.
- Performed feature extraction from Datetime Fields by extracting day of the week and time of the day.
- Removed Duplicates.
- Handled Outliers by capping the values below the lower bound and above the upper bound.

After cleaning the dataset:

```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 901357 entries, 0 to 901356
Data columns (total 26 columns):
 #   Column          Non-Null Count   Dtype
---  ------          --------------   -----
 0   DR_NO           901357 non-null  float64
 1   Date Rptd       901357 non-null  object
 2   DATE OCC        901357 non-null  datetime64[ns]
 3   TIME OCC        901357 non-null  object
 4   AREA            901357 non-null  int64
 5   AREA NAME       901357 non-null  object
 6   Rpt Dist No     901357 non-null  int64
 7   Part 1-2        901357 non-null  int64
 8   Crm Cd          901357 non-null  int64
 9   Crm Cd Desc     901357 non-null  object
 10  Mocodes         901357 non-null  object
 11  Vict Age        901357 non-null  float64
 12  Vict Sex        901357 non-null  object
 13  Vict Descent    901357 non-null  object
 14  Premis Cd       901357 non-null  float64
 15  Premis Desc     901357 non-null  object
 16  Weapon Used Cd  901357 non-null  float64
 17  Weapon Desc     901357 non-null  object
 18  Status          901357 non-null  object
 19  Status Desc     901357 non-null  object
 20  Crm Cd 1        901357 non-null  float64
 21  LOCATION        901357 non-null  object
 22  Cross Street    901357 non-null  object
 23  LAT             901357 non-null  float64
 24  LON             901357 non-null  float64
 25  Day of Week     901357 non-null  object
dtypes: datetime64[ns](1), float64(7), int64(4), object(14)
memory usage: 185.7+ MB
```

Figure 2: Information of the cleaned data

Make sure to change it to a proper table in the final report

Module 2 - Visualization:

- Checked the current dataset features correlation, find the Strong Positive Correlations for deep research .
- For example, 'Rpt Dist No' (Report District Number) and 'Part 1-2' seem to have a strong positive correlation, suggesting that as one increases, the other tends to increase as well.

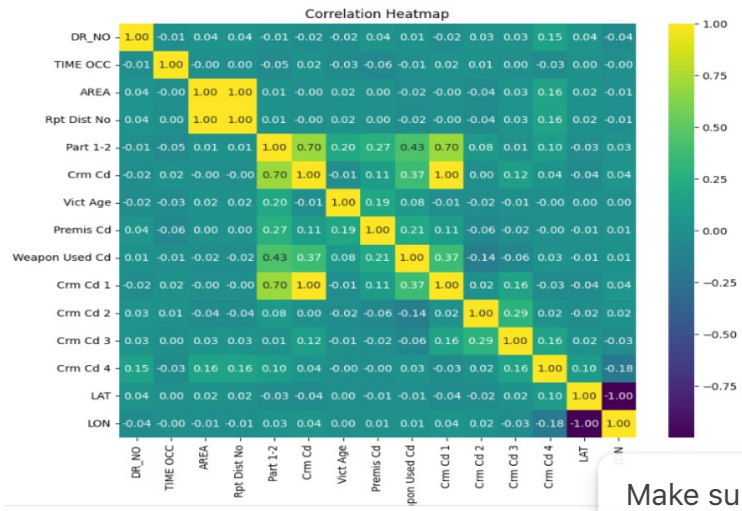


Figure 3: Heatmap showing the correlation of features with ea

Make sure the graphs are not blurred in the final report.

Geography Heatmap:

- Used the Folium library to build a geographical heatmap to provide th a more concrete and specific chart for viewing.
- Below graph uses the first 1000 row's latitude and longitude to set the crime position in actual map.

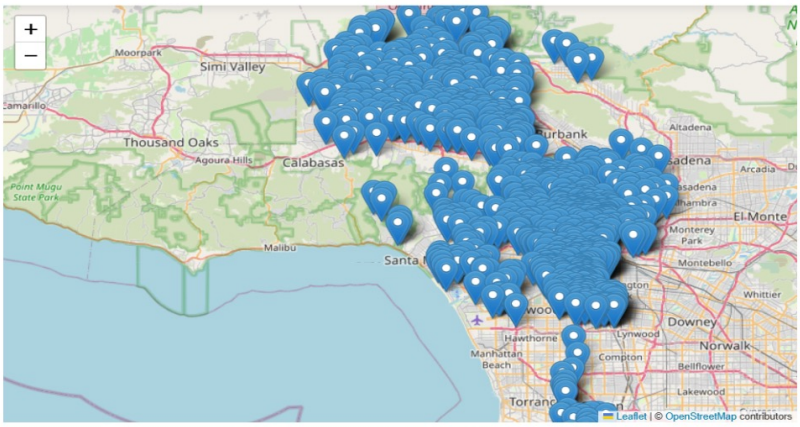
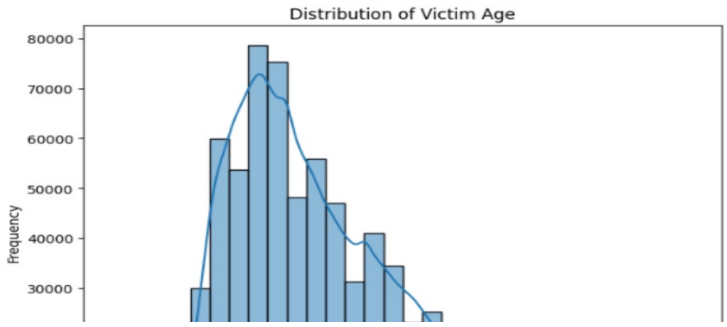


Figure 4: Geographical heatmap using latitude and longitude of the crime occurrence

Module 3: Single feature analysis



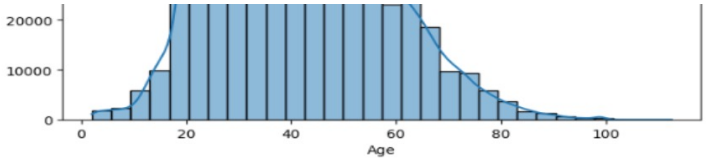


Figure 5: Histogram showing the distribution of victim age with the frequency of the crimes

- The histogram serves as an excellent illustration of our analysis, demonstrating the relationship between age and crime frequency.
- From the chart, it is clearly observable that the majority of crimes are perpetrated by individuals in the 20-40 year age range.
- The data appears to be skewed right, indicating that younger individuals are more often victims compared to older individuals.

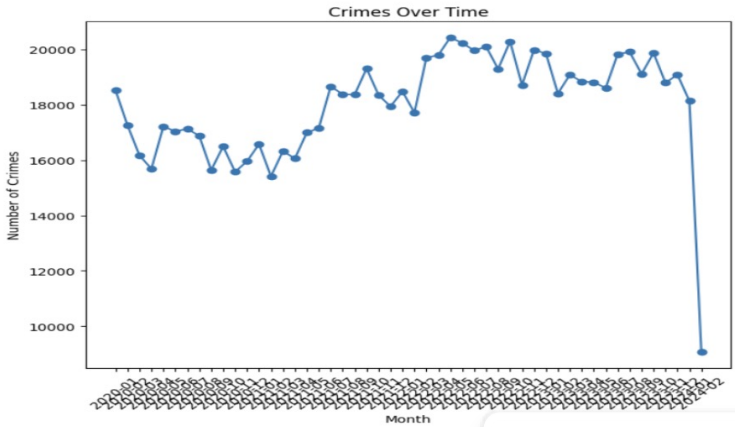


Figure 6: Plot showing the distribution of crimes over time with the

The X axes is not readable

- This line graph shows the number of crimes over a period of time.
- The last data point shows a sharp decline, which is due to incomplete data for the most recent month.

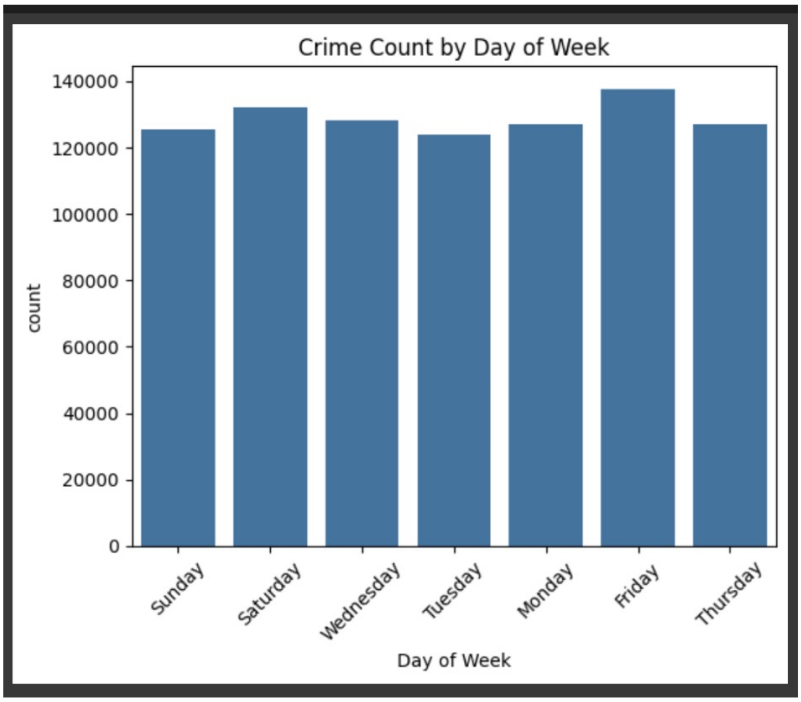


Figure 7: Plot showing the distribution of crimes count by day of the week with the frequency of the crimes

- The bars seem relatively even in height, suggesting that the crime count does not fluctuate significantly by day of week.

Module 4 - Threshold based Classification:

- Used a simple threshold classification to identify areas with higher frequencies of crime. Classified these areas into broad categories like very high, high, medium, and low incidence zones.
- First we counted the number of crimes in each area and merged the crime_counts back with the original data by creating a new column for crime_counts.
- We defined thresholds for very high, high, medium, and low incidence zones by calculating the percentiles.

7

- We used 25, 50, 75 percentiles for the threshold classification.
- We applied the thresholds to categorize the 'RiskLevel'.

	AREA NAME	CrimeCount	RiskLevel
0	Wilshire	43191	Medium
1	Central	61416	Very High
2	Southwest	50654	Very High
3	Van Nuys	38056	Medium
4	Hollywood	47508	High
...	...	...	...
901352	Foothill	29796	Low

901353	Hollywood	47508	High
901354	Central	61416	Very High
901355	Van Nuys	38056	Medium
901356	Southwest	50654	Very High

901357 rows × 3 columns

Figure 8: Threshold based classification of the risk level of the crime spots.

Project resources:
For this project, we use Python as our programming language. We choose to use Google Colab and Jupyter Notebook as Development Environments.
Below is the Python library we used in the project:
Pandas: Streamlines data manipulation and analysis with flexible structures like DataFrames, aiding in data cleaning, filtering, and visualization.
NumPy: Provides efficient arrays for high-performance numerical computations, serving as the foundation for Python's data science ecosystem.
Matplotlib: Enables a wide range of data visualizations, from simple charts to complex plots, crucial for insight into data trends.
Folium: Facilitates interactive geospatial data visualization on maps, making the embedding of complex maps in Python straightforward.

The training dataset comes from posting the Crime data with below Features:

8

Feature Name	Description
DR_NO	Unique report number for each crime incident.
Date Rptd	The date when the crime was reported.
DATE OCC	The date when the crime occurred.
TIME OCC	The time when the crime occurred.
AREA	Numeric code representing the area where the crime was reported.
AREA NAME	The name of the area or neighbourhood where the crime occurred.
Rpt Dist No	Reporting district number, a specific code for a smaller segment within the area.
Part 1-2	Indicator of whether the crime is a Part 1 (more serious) or Part 2 offence.
Crm Cd	Numeric code for the crime category.
Crm Cd Desc	Description of the crime category.
Mocodes	Modus operandi codes providing details on the crime.
Vict Age	Age of the victim.
Vict Sex	Sex of the victim.
Vict Descent	Ethnic/racial descent of the victim.
Premis Cd	Numeric code indicating the type of location where the crime occurred.
Premis Desc	Description of the location type.
Weapon Used Cd	Code for the type of weapon used (if any).
Weapon Desc	Description of the weapon used.
Status	Code indicating the status of the crime report (e.g., investigated, closed).
Status Desc	Description of the report status.
Crm Cd 1-4	Up to four additional crime codes if multiple offences occurred in the same incident.

LOCATION	Street address or intersection where the crime occurred.
Cross Street	Nearest cross street (if applicable).
LAT	Latitude coordinate of the crime location.
LON	Longitude coordinate of the crime location.

Dataset detail shows below:

	DR_NO	Date Rptd	DATE OCC	TIME OCC	AREA	AREA NAME	Rpt Dist No	Part 1-2	Crm Cd	Crm Cd Desc	...	Weapon Used Cd	Weapon Desc	Status	Status Desc	Crm Cd 1	LOCATION	Cross Street	LAT	LON	Day of Week
0	190325475.0	03/01/2020 12:00:00 AM	2020-03-01	21:30:00	7	Wilshire	784	1	510	VEHICLE - STOLEN	...	400.0	STRONG-ARM (HANDS, FIST, FEET OR BODILY FORCE)	AA	Adult Arrest	510.0	1900 S LONGWOOD AV	BROADWAY	34.0375	-118.3506	1
1	200100753.0	02/09/2020 12:00:00 AM	2020-02-08	18:00:00	1	Central	182	1	330	BURGLARY FROM VEHICLE	...	400.0	STRONG-ARM (HANDS, FIST, FEET OR BODILY FORCE)	IC	Invest Cont	330.0	1000 S FLOWER ST	BROADWAY	34.0444	-118.2628	0
2	200320258.0	11/11/2020 12:00:00 AM	2020-11-04	17:00:00	3	Southwest	356	1	480	BIKE - STOLEN	...	400.0	STRONG-ARM (HANDS, FIST, FEET OR BODILY FORCE)	IC	Invest Cont	480.0	1400 W 37TH ST	BROADWAY	34.0210	-118.3002	4
3	200907217.0	05/12/2023 12:00:00 AM	2020-03-10	20:37:00	9	Van Nuys	964	1	343	SHOPLIFTING-GRAND THEFT (.$950.01 & OVER)	...	400.0	STRONG-ARM (HANDS, FIST, FEET OR BODILY FORCE)	IC	Invest Cont	343.0	14000 RIVERSIDE DR	BROADWAY	34.1575	-118.4387	10
4	220614831.0	08/18/2022	2020-08-	17:00:00	6	Hollywood	666	2	354	THEFT OF IDENTITY	...	400.0	STRONG-ARM (HANDS, FIST, FEET OR	IC	Invest Cont	354.0	1900 TRANSIENT	BROADWAY	34.0944	-118.3277	17

			12:00:00 AM	17											BODILY FORCE)									
...	...		...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
001352	231600525.0	03/22/2023 12:00:00 AM	2023-03-22	10:00:00	16	Foothill	1652	1	230	ASSAULT WITH DEADLY WEAPON, AGGRAVATED ASSAULT	...	400.0	STRONG-ARM (HANDS, FIST, FEET OR BODILY FORCE)	IC	Invest Cont	230.0	12800 FILMORE ST	BROADWAY	34.2760	-118.4110	22
001353	230613716.0	12/07/2023 12:00:00 AM	2023-12-01	10:00:00	6	Hollywood	647	2	901	VIOLATION OF RESTRAINING ORDER	...	400.0	STRONG-ARM (HANDS, FIST, FEET OR BODILY FORCE)	IC	Invest Cont	901.0	1500 N BRONSON AV	BROADWAY	34.0980	-118.3180	1
001354	230115220.0	07/02/2023 12:00:00 AM	2023-07-01	00:01:00	1	Central	154	1	352	PICKPOCKET	...	400.0	STRONG-ARM (HANDS, FIST, FEET OR BODILY FORCE)	IC	Invest Cont	352.0	500 S MAIN ST	BROADWAY	34.0467	-118.2485	1
001355	230905456.0	03/05/2023 12:00:00 AM	2023-03-05	09:00:00	9	Van Nuys	914	2	745	VANDALISM - MISDEAMEANOR ($399 OR UNDER)	...	400.0	STRONG-ARM (HANDS, FIST, FEET OR BODILY FORCE)	IC	Invest Cont	745.0	14500 HARTLAND ST	BROADWAY	34.1951	-118.4487	5
001356	230310716.0	11/10/2023 12:00:00 AM	2023-11-09	23:00:00	3	Southwest	305	1	331	THEFT FROM MOTOR VEHICLE - GRAND ($950.01 AND ...	...	400.0	STRONG-ARM (HANDS, FIST, FEET OR BODILY FORCE)	IC	Invest Cont	331.0	4100 S HOBART BL	BROADWAY	34.0991	-118.3078	9

59357 rows × 26 columns

Figure 9: Part of the Dataset

Conclusions and Future Work:
We were able to complete the minimum functionality for our project. The tasks completed were: Data Collection, Data Cleaning, Data Visualisation, Analysis, Threshold based classification of the areas of crime into very high, high, medium and low risk areas. For the future work we will implement the following:

Expected Functionality:

1. **Model development (clustering techniques):**
   - Implement density-based clustering (e.g., DBSCAN) to identify and visualize crime hotspots.Analyse spatial patterns to understand the distribution of different types of crimes.
2. **Heatmaps and visualization:**
   - Generate heat maps to visualize areas with high crime densities.
3. **Validation and testing:**
   - Use a portion of your data for training (setting thresholds) and another portion for validation.

Bonus Functionality:
1. **Temporal Analysis of Crime Data:**

- Conduct time series analysis to identify trends and patterns over time and how the hotspots evolved over time. Explore variations in crime rates by time of day, week, and season.

2. **Correlation with External Datasets:**
   - Integrate external demographic and socioeconomic data to examine correlations of unemployment or budget allocation with crime rates.
3. **Location expansion other than LA**

Slides

10 / 10

Here you submit the slides of your project progress presentation in pdf format

Crime Zone Classifier

PROJECT GROUP MEMBERS

Sichen Qiu - SC985823

Jahnavi Gajjala - JH707761

Venkata Sreenivas Prasad – VN769140

1

BUSINESS PROBLEM

- In metropolitan cities like Los Angeles, there might be numerous crime patterns, which can be both enduring and constantly evolving, are common in urban areas.
- Design and implement a system that uses past crime data to precisely identify and categorize crime hotspots in a metropolitan region.
- Categorizing different areas based on their likelihood of experiencing criminal activities.

2

WHY?

1. Public Safety Enhancement
2. Effective Resource Allocation
3. Informed Community Awareness
4. Aid in Urban Planning

3

Source Used:

https://data.lacity.org/Public-Safety/Crime-Data-from-2020-to-Present/2nrs-mtv8/data_preview

What's in this dataset:

Rows: 900K

Columns: 28

Time Range: February 10,2020 to January 18,2024

Each row is a crime incident

4

INPUT DATA

Few of the Features:

```
df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 883987 entries, 0 to 883986
Data columns (total 28 columns):
 #   Column          Non-Null Count   Dtype
---  ------          --------------   -----
```

DR_NO : Division of Records Number

Date Rptd : Crime Report Date

DATE OCC : Crime Occurance date

TIME OCC : Crime Occurance Time

AREA : The LAPD has 21 Community Police Stations referred to as Geographic Areas within the department.

AREA NAME : The 21 Geographic Areas or Patrol Divisions

Rpt Dist No : A 4-digit code that represents a sub-area.

Crm Cd : Indicates the crime committed.

Crm Cd Desc : Defines the Crime Code provided.

Vict Age : Victim age

```
 ---  ------          --------------  -----
 0    DR_NO           883987 non-null  int64
 1    Date Rptd       883987 non-null  object
 2    DATE OCC        883987 non-null  object
 3    TIME OCC        883987 non-null  int64
 4    AREA            883987 non-null  int64
 5    AREA NAME       883987 non-null  object
 6    Rpt Dist No     883987 non-null  int64
 7    Part 1-2        883987 non-null  int64
 8    Crm Cd          883987 non-null  int64
 9    Crm Cd Desc     883987 non-null  object
 10   Mocodes         760974 non-null  object
 11   Vict Age        883987 non-null  int64
 12   Vict Sex        766927 non-null  object
 13   Vict Descent    766919 non-null  object
 14   Premis Cd       883977 non-null  float64
 15   Premis Desc     883447 non-null  object
 16   Weapon Used Cd  306604 non-null  float64
 17   Weapon Desc     306604 non-null  object
 18   Status          883987 non-null  object
 19   Status Desc     883987 non-null  object
 20   Crm Cd 1        883976 non-null  float64
 21   Crm Cd 2        64497 non-null   float64
 22   Crm Cd 3        2176 non-null    float64
 23   Crm Cd 4        62 non-null      float64
 24   LOCATION        883987 non-null  object
 25   Cross Street    139779 non-null  object
 26   LAT             883987 non-null  float64
 27   LON             883987 non-null  float64
dtypes: float64(8), int64(7), object(13)
memory usage: 188.8+ MB
```

Fig 1: Data information

5

Tasks completed Till Date

- **Data Gathering**
- **Data Cleaning**
- **Data Visualization**

- Completion of the minimal functionality

6

DATA CLEANING

Before

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 901357 entries, 0 to 901356
Data columns (total 28 columns):
 #   Column          Non-Null Count   Dtype
---  ------          --------------   -----
 0   DR_NO           901357 non-null  int64
 1   Date Rptd       901357 non-null  object
 2   DATE OCC        901357 non-null  object
 3   TIME OCC        901357 non-null  int64
 4   AREA            901357 non-null  int64
 5   AREA NAME       901357 non-null  object
 6   Rpt Dist No     901357 non-null  int64
 7   Part 1-2        901357 non-null  int64
 8   Crm Cd          901357 non-null  int64
 9   Crm Cd Desc     901357 non-null  object
 10  Mocodes         775767 non-null  object
 11  Vict Age        901357 non-null  int64
 12  Vict Sex        781966 non-null  object
 13  Vict Descent    781958 non-null  object
 14  Premis Cd       901347 non-null  float64
 15  Premis Desc     900007 non-null  object
 16  Weapon Used Cd  312268 non-null  float64
 17  Weapon Desc     312268 non-null  object
 18  Status          901357 non-null  object
```

Null values

```
# Check for missing values
print(data.isnull().sum())

DR_NO                 0
Date Rptd             0
DATE OCC              0
TIME OCC              0
AREA                  0
AREA NAME             0
Rpt Dist No           0
Part 1-2              0
Crm Cd                0
Crm Cd Desc           0
Mocodes          125590
Vict Age              0
Vict Sex         119391
Vict Descent     119399
Premis Cd            10
Premis Desc         550
Weapon Used Cd   589089
Weapon Desc      589089
Status                0
```

After

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 901357 entries, 0 to 901356
Data columns (total 26 columns):
 #   Column          Non-Null Count   Dtype
---  ------          --------------   -----
 0   DR_NO           901357 non-null  float64
 1   Date Rptd       901357 non-null  object
 2   DATE OCC        901357 non-null  datetime64[ns]
 3   TIME OCC        901357 non-null  object
 4   AREA            901357 non-null  int64
 5   AREA NAME       901357 non-null  object
 6   Rpt Dist No     901357 non-null  int64
 7   Part 1-2        901357 non-null  int64
 8   Crm Cd          901357 non-null  int64
 9   Crm Cd Desc     901357 non-null  object
 10  Mocodes         901357 non-null  object
 11  Vict Age        901357 non-null  float64
 12  Vict Sex        901357 non-null  object
 13  Vict Descent    901357 non-null  object
 14  Premis Cd       901357 non-null  float64
 15  Premis Desc     901357 non-null  object
 16  Weapon Used Cd  901357 non-null  float64
 17  Weapon Desc     901357 non-null  object
 18  Status          901357 non-null  object
```

```
19  Status Desc     901357 non-null  object
20  Crm Cd 1        901346 non-null  float64
21  Crm Cd 2        65683 non-null   float64
22  Crm Cd 3        2213 non-null    float64
23  Crm Cd 4        64 non-null      float64
24  LOCATION        901357 non-null  object
25  Cross Street    142055 non-null  object
26  LAT             901357 non-null  float64
27  LON             901357 non-null  float64
dtypes: float64(8), int64(7), object(13)
memory usage: 192.6+ MB
```

Fig 2: Before cleaning the Dataset

```
Crm Cd 1            11
Crm Cd 2        835674
Crm Cd 3        899144
Crm Cd 4        901293
LOCATION             0
Cross Street    759302
LAT                  0
LON                  0
dtype: int64
```

Fig 3: Null values information

```
19  Status Desc     901357 non-null  object
20  Crm Cd 1        901357 non-null  float64
21  LOCATION        901357 non-null  object
22  Cross Street    901357 non-null  object
23  LAT             901357 non-null  float64
24  LON             901357 non-null  float64
25  Day of Week     901357 non-null  object
dtypes: datetime64[ns](1), float64(7), int64(4), object(14)
```

Fig 4: After cleaning the Dataset

7

STEPS IN DATA CLEANING

1) Removed the columns [Crm Cd 2,Crm Cd 3, Crm Cd 4], as all these columns had 95% + Nan values.

2) Missing values in the **Continous features** were filled with the **median**.

3) Missing values in the **Categorical features** were filled with the **mode**.

4) Convertion of the date type from String to dateTime datatype.

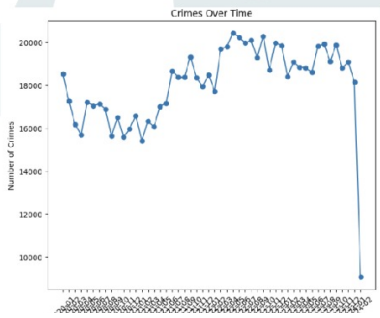
5) Adding a column which will specify which day of the week the crime occurred.

6) Dropping the duplicates.

7) Removing the outliers.

8

DATA VISUALIZATION



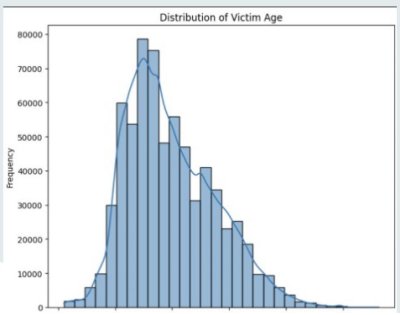


Fig 5: Line graph

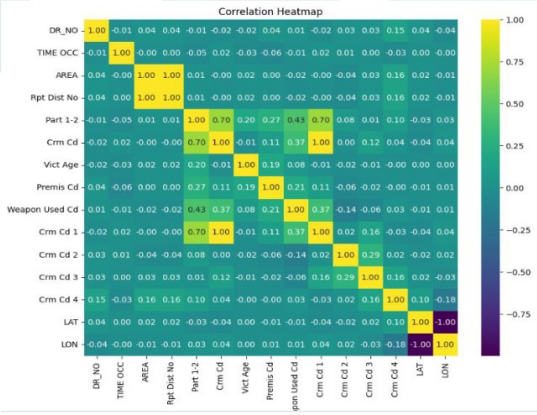
This line graph shows the number of crimes over a period of time, which seems to be measured monthly. There is a noticeable trend or pattern in crime occurrence over time, but the specific time frame is not legible. The last data point shows a sharp decline, which could be due to incomplete data for the most recent month or a sudden drop in crime incidents.

Fig 6: Histogram

The histogram depicts the distribution of victims' ages. It is bimodal with two peaks, suggesting that there are two age groups that are more frequently victimized than others. The data appears to be skewed right, indicating that younger individuals are more often victims compared to older individuals.

9

DATA VISUALIZATION



- Checked the current dataset features correlation, find the Strong Positive Correlations for deep research .
- For example, 'Rpt Dist No' (Report District Number) and 'Part 1-2' seem to have a strong positive correlation, suggesting that as one increases, the other tends to increase as well.

10

Minimum Functionality (Threshold based classification)

1) Calculating the crime counts by grouping the common area names.
2) Setting a threshold.
3) Based on the threshold (25, 50,75 percentile), distributing the area names into low, medium, high and very high.

10

TARGET

Expected Functionality:

1. **Model development (clustering techniques):**

● Implement density-based clustering (e.g., DBSCAN) to identify and visualise crime hotspots.Analyse spatial patterns to understand the distribution of different types of crimes.

2. **Heatmaps and visualisation:**

● Generate heat maps to visualise areas with high crime densities.

3. **Validation and testing:**

● Use a portion of your data for training (setting thresholds) and another portion for validation.

Bonus Functionality:

1. **Temporal Analysis of Crime Data:**

● Conduct time series analysis to identify trends and patterns over time and how the hotspots evolved over time. Explore variations in crime rates by time of day, week, and season.

2. **Correlation with External Datasets:**

● Integrate external demographic and socioeconomic data to examine correlations of unemployment or budget allocation with crime rates.

3. **Location expansion other than LA**

11