Crime Data Analysis

A Project Report

Submitted by

KARTHI G

KANISHK R

GURUDEV DJ

for

DEPARTMENT OF ARTIFICIAL INTELLIGENCE

KONGU ENGINEERING COLLEGE(Autonomous)

PERUNDURAI ERODE – 638 060

NOVEMBER 2024

DEPARTMENT OF ARTIFICIAL INTELLIGENCE

KONGU ENGINEERING COLLEGE

(Autonomous)

PERUNDURAI ERODE – 638 060

NOVEMBER 2024

Department of Artificial Intelligence

22ADF01 – Data Analysis Project Report

Signature of course in-charge Signature of the HOD

Submitted for the continuous Assessment viva voce examination held on _____

EXAMINER I **EXAMINER II**

ABSTRACT

The Crime Data Analysis System leverages the power of data analytics and visualization to help law enforcement agencies and decision-makers identify crime patterns and trends. By using a large dataset containing historical crime data, factors such as crime type, location, and time of occurrence are analyzed to find significant correlations. The system utilizes Geographic Information System (GIS) mapping to visually represent crime hotspots, making it easier to spot high-risk areas. Machine learning algorithms such as Decision Trees and Neural Networks are applied to predict potential future crimes based on past data. These predictions allow law enforcement agencies to optimize their patrol routes, allocate resources more effectively, and prevent crime before it occurs. The use of real-time data from sources like police reports ensures that the system remains up-to-date, providing valuable insights that adjust to new information.

In this project, Power BI was used as the primary tool for data processing, analysis, and visualization. A comprehensive crime dataset was imported into Power BI, which was then cleaned and prepared for analysis. Different visualizations, such as heat maps, bar charts, and pie charts, were created to represent crime trends across various dimensions—such as location, time, and crime type. Power BI's data modeling features allowed for filtering and slicing the dataset to explore specific areas of interest, such as high-crime neighborhoods or the frequency of certain crime types. By using Power BI's interactive dashboards, users can drill down into the data and discover deeper insights, making the system an effective solution for crime data analysis.

TABLE OF CONTENTS

CHAPTER No.	TITLE	PAGE NO
	ABSTRACT	I
	LIST OF FIGURES	II
1.	**INTRODUCTION**	5
	1.1 INTRODUCTION	
	1.1.1 DATA COLLECTION	
	1.2 PROBLEM STATEMENT	
	1.3 BUSINESS OBJECTIVE	
2.	**DATA PREPARATION AND MODELING**	12
	2.1 DATA CLEANING	
	2.2 DATA TRANSFORMATION	
	2.3 DATA DISTRIBUTION USING CHARTS	
	2.4 DATA MODELLING	
3.	**DATA ANALYSIS AND INTERPRETATION**	25
	3.1 DATA ANALYSIS	
	3.2 PUBLISHING DASHBOARDS	
	3.3 INFERENCE	
4.	**CONCLUSION**	39
	4.1 RECOMMENDATIONS	
5.	**REFERENCES**	40

CHAPTER 1

INTRODUCTION

1.1 INTRODUCTION

The primary goal of this Crime Data Analysis project is to enhance the interpretation of crime data to support better decision-making processes in fields such as law enforcement, urban planning, and policy-making. With the rise in criminal activities and the vast amount of data being generated, analyzing and interpreting this data has become essential for public safety. Crime data analysis can help in identifying patterns, trends, and geographic hotspots, which in turn can aid law enforcement agencies in strategic planning and crime prevention. By making crime data more accessible and interpretable, this project aims to improve the accuracy and relevance of the analysis.

The scope of the project includes developing algorithms and models that can automatically analyze crime patterns across various environments, such as urban areas or specific neighborhoods. Utilizing a dataset imported from Kaggle, the data is processed and cleaned using **Power BI**. The cleaned data is then used to create interactive visualizations, including charts, graphs, and maps, that make it easier to detect crime trends, hotspots, and frequency distributions. Power BI's advanced analytics and data modeling capabilities are used to explore deeper insights, such as seasonal crime trends and geospatial distribution of crimes. The expected outcomes of this project include improving the accuracy of crime pattern recognition, enabling law enforcement agencies to develop more effective strategies for crime prevention and community safety, and providing policymakers with the tools to make informed decisions. Through interactive dashboards and filters, users can explore the data in a meaningful way, leading to more targeted interventions and a reduction in crime rates across different regions.

1.2 DATA COLLECTION

The dataset for this Crime Data Analysis project was sourced from Kaggle and consists of 752,911 rows.

- **DR_NO**: This is a unique identifier assigned to each crime incident. It serves as a primary key for tracking and referencing individual cases within the database. Each crime has its own distinct DR_NO, ensuring no two incidents share the same identifier.
- **TIME OCC**: This field records the exact time the crime occurred, noted in a 24-hour format (e.g., 1345 for 1:45 PM). This timestamp is crucial for understanding crime patterns and helps authorities analyze peak hours for criminal activity.
- **AREA**: This numeric code represents the specific area or jurisdiction where the crime took place. Each area has a unique code, allowing data analysts to group incidents by region and enabling more localized analysis.
- **AREA NAME**: The name of the location, district, or neighborhood where the crime occurred. This information helps in identifying trends within specific communities and aids in regional comparisons for law enforcement strategies.
- **Rpt Dist No**: The Reporting District Number is a tracking code used by law enforcement to manage and organize crimes within smaller subdivisions or districts. It helps to narrow down the crime to a specific reporting area for more targeted resource allocation.
- **Crm Cd Desc**: A description of the crime committed, such as theft, assault, burglary, etc. This field categorizes crimes by type, allowing analysts to assess which crimes are most prevalent and identify trends in criminal behavior.
- **Mocodes**: This field denotes the specific method or modus operandi used to commit the crime. It helps in identifying patterns in criminal techniques, which can be valuable for crime-solving and predictive analysis.

- **Vict Age**: Records the age of the victim involved in the incident. Analyzing victim age ranges can reveal certain demographics at higher risk for specific crimes, aiding in focused prevention efforts.
- **Vict Sex**: Indicates the gender of the victim (options are Male, Female, or Unknown). This data is useful for identifying if certain crimes disproportionately affect one gender over another.
- **Vict Descent**: Provides information on the ethnicity or descent of the victim. This demographic data allows for analysis of crime impact on various ethnic groups, informing community-oriented policing efforts.
- **Premis Cd**: This code categorizes the type of location where the crime occurred (such as residential, public, commercial, or private locations). It helps in understanding the settings most susceptible to certain types of crimes.
- **Premis Desc**: A description of the premises where the crime took place, such as a house, park, street, or business establishment. This information aids in identifying vulnerable locations and developing location-specific security measures.
- **Weapon Desc**: If a weapon was used in the crime, this field describes it.
- **Status**: A code indicating the current status of the case, such as whether it is open, pending, or resolved. This code helps in tracking the progression of investigations and case closures.
- **Status Desc**: Provides a more detailed explanation of the case status, offering insights into specific reasons for case delays,
- **LOCATION**: This field provides the specific address or intersection where the crime occurred. Precise location information allows for spatial analysis and supports hotspot mapping to identify areas with frequent incidents.
- **LAT**: The latitude coordinate of the crime location, used for geospatial analysis. This coordinate, along with longitude, allows mapping of crime data to visualize geographic distribution and detect spatial crime patterns.
- **LON**: The longitude coordinate of the crime location, complementing latitude for precise geolocation.

1.3 PROBLEM STATEMENT

The vast amount of crime data gathered from various locations presents a formidable challenge for effective analysis and comprehensive understanding of crime patterns. Law enforcement agencies often encounter difficulties in making timely and informed decisions due to the data's volume, complexity, and the limited availability of advanced analytical tools. The data, which includes information on incidents, locations, timings, victim demographics, and criminal profiles, is often stored across disparate databases and formats, making it difficult to create a cohesive picture.

This project aims to address these issues by leveraging Power BI, a powerful business analytics tool, to integrate, process, visualize, and analyze extensive crime datasets. Power BI's advanced data visualization capabilities enable the transformation of raw crime data into clear, interactive dashboards and visual reports, allowing stakeholders to explore crime patterns intuitively and efficiently. By consolidating data from multiple sources into a unified, interactive platform, the project seeks to enhance the accessibility and usability of crime information, enabling law enforcement agencies to respond more effectively to emerging trends and high-crime areas.

Through this system, users can analyze crime trends over time, examine geographic hotspots, and identify the demographics most affected by different types of crime. Detailed visualizations, such as heat maps and time-series analyses, make it possible to detect seasonal or location-based crime surges and to anticipate future criminal activity patterns. This intelligence can then inform strategic decisions on resource allocation, enabling authorities to focus efforts on high-risk areas and times and ultimately improve crime prevention strategies. Additionally, by enhancing the transparency and accessibility of crime data, the system supports policy makers in developing more targeted and evidence-based policies, contributing to a safer and more secure community.

1.4 BUSINESS OBJECTIVE

- **Enhance crime data visibility**:

  The project will provide law enforcement agencies with clear, interactive visualizations of crime data, turning complex information into easy-to-understand formats such as heat maps and trend lines. This makes it easier to identify trends, recurring patterns, and hotspots, empowering agencies to monitor shifts in crime dynamics and respond in real time.

- **Support decision-making**:

  With insights derived from data analysis, agencies can make informed decisions about crime. This data-driven approach helps prioritize patrols, deploy resources strategically, and address specific crime trends, ensuring a more efficient response to public safety needs.

- **Predict future crimes**:

  Insights into crime types, demographics, and locations allow agencies to craft targeted safety strategies. By understanding local crime drivers, law enforcement can develop initiatives aimed at addressing underlying issues, contributing to long-term community safety improvements.

- **Improve public safety strategies**:

  Insights into crime types, demographics, and locations allow agencies to craft targeted safety strategies. By understanding local crime drivers, law can develop initiatives , contributing to long-term community safety improvements.

- **Optimize resource deployment**:

  The platform enhances resource management by identifying high-crime areas and peak times for criminal activity, allowing authorities to deploy patrols and resources precisely where they are most needed. By using data to map hotspots and predict active periods, law enforcement can allocate personnel strategically, focusing on areas and times with the highest potential impact. This targeted approach not only maximizes the effectiveness of each resource but also reduces response times, enabling faster interventions that help deter crime and improve public safety.

- **Aid policy development**:

  The platform empowers policymakers by providing them with reliable, data-driven insights, which are essential for formulating effective crime-related policies and urban safety measures. By analyzing crime trends, demographic impacts, and geographic patterns, policymakers can gain a clearer understanding of areas requiring focused attention. This data-driven approach supports the creation of targeted policies, such as allocating resources to high-crime areas, implementing preventive measures for vulnerable communities, and designing urban infrastructure improvements to enhance public safety. The insights from this platform not only help shape current policies but also aid in forecasting future needs, ensuring that policy decisions are proactive, informed, and responsive to community safety concerns.

CHAPTER 2

DATA PREPARATION AND MODELING

2.1 DATA CLEANING

Handling Missing Values

In any dataset, missing values can lead to incomplete analysis and incorrect conclusions. Power BI's Query Editor provides flexible options for handling missing data, helping ensure that the dataset is as accurate and complete as possible. Depending on the analysis, you can decide whether to fill in missing values or exclude incomplete records.

- **Replacement Strategy**: If data is missing in specific columns, you might choose to replace those values with statistically meaningful substitutes like the column average, median, or a constant default value. For example, if a dataset includes crime statistics, missing values in age could be replaced with the average age to maintain demographic balance.
- **Row Removal**: For analyses where precision is critical, rows with missing values may need to be removed to avoid skewing results. This approach works well for smaller datasets where each data point's completeness is essential.
- **Method**: Select the columns with missing data → Use "Remove Rows" to delete incomplete entries or "Replace Values" to insert substitutes, ensuring consistency across the dataset.

Removing Duplicates

Duplicate data points can distort analysis by artificially inflating values or creating redundancy in visualizations. Power BI's Query Editor has options for identifying and removing duplicates, preserving the dataset's integrity.

- **Importance of Deduplication**: Removing duplicates ensures that calculations like counts, averages, and totals are accurate. For example, if a dataset includes multiple entries for the same crime incident, removing these duplicates will prevent overestimations of crime frequency.
- **Method**: Select one or more columns that should be unique (such as a crime ID or date-time of occurrence) → Use "Remove Duplicates" to clear repeated entries. This step ensures each data point is counted only once.

Data Type Conversion

Ensuring each column has the correct data type is crucial, as inconsistent types can lead to errors during analysis or visualization. Power BI allows conversion between types like text, date, and number to facilitate accurate calculations and meaningful visualizations.

- **Data Type Relevance**: Converting columns to the correct data type enables functions specific to that type. For instance, dates should be in date format to allow for time-based analysis, while numeric fields need to be in number format to support aggregation.
- **Method**: Select the desired column → Choose "Data Type" in the Query Editor and specify the appropriate format (e.g., Text, Date, or Number). This ensures each field behaves correctly in formulas and visualizations.

Splitting Columns

Data often comes in consolidated formats that may need to be separated for more granular analysis. For example, a single timestamp might need to be divided into separate date and time columns to analyze daily or hourly trends.

- **Application**: Splitting columns enhances analysis precision, allowing, for example, crime data to be analyzed by time of day or specific dates. Address fields may also need splitting for individual city, state, or ZIP code analysis.
- **Method**: Select the target column → Use the "Split Column" feature in Query Editor, specifying the delimiter (such as space, comma, or hyphen) to break data into separate parts.

Removing Unnecessary Columns

To improve processing speed and avoid distractions in your analysis, it's beneficial to remove irrelevant columns that don't contribute to your analytical goals. By focusing on essential data, Power BI performs faster and presents clearer insights.

- **Rationale**: Removing non-essential columns reduces dataset clutter, leading to faster query processing and easier report creation. This is particularly helpful for large datasets with many attributes.
- **Method**: Select columns that are unnecessary for your analysis → Use "Remove Columns" to delete them, streamlining the dataset.

Transforming Data

Power BI supports various data transformation operations to standardize values, improve readability, and ensure consistent formatting. This includes actions like trimming whitespace, merging columns, and formatting text. These transformations enhance the data's usability, ensuring that it is clean, well-structured, and ready for analysis.

- **Trim**: Use the “Trim” function to remove any extra spaces from text entries. This ensures consistency and prevents unnecessary gaps in data that could lead to misinterpretation or errors in analysis.
- **Merge Columns**: If you have data split across multiple columns that should be combined (e.g., street number and street name), use the “Merge Columns” function to combine them into a single column, improving data organization and clarity.
- **Format Text**: Power BI allows you to format text entries consistently, such as converting them to uppercase, lowercase, or proper case. This is useful for standardizing names or locations and improving the readability of your data.

Standardizing Data

In datasets with inconsistent formats, differing entries for the same category can cause confusion and skew analysis results. Power BI's tools allow you to standardize entries across columns, ensuring uniformity and reducing the risk of duplicate categories.

- **Use Case**: When data entries vary in format—such as "N.Y.," "NYC," and "New York City" all representing the same location—standardizing them into a single label ("New York") maintains a cohesive dataset. This approach prevents duplicate categories from appearing in visualizations, leading to more accurate insights.
- **Method**: In Power BI’s Query Editor, the "Replace Values" function is used to unify variations. By selecting the target column and specifying the desired replacement values, you can create a consistent format across similar entries, making aggregated data clearer and enhancing the accuracy of reports.

Filtering Rows

Filtering allows you to exclude irrelevant or erroneous data points that may skew results. This is especially useful for isolating data that meets specific conditions or for removing outliers that don't fit the dataset's purpose.

- **Examples of Filtering**: In crime analysis, you might filter out records with incomplete dates or locations, or filter by a specific crime type to focus on targeted insights.
- **Method**: Select the desired column → Use "Filter Rows" to apply conditions and exclude rows that don't meet specific criteria, refining the dataset for more accurate analysis.

Dealing with Outliers

Outliers are data points that deviate significantly from the norm and can distort analysis if not properly handled. Power BI provides tools to detect and address outliers, either by excluding them or by using statistical adjustments.

- **Detecting Outliers**: Outliers can be identified through visualizations like box plots or statistical measures, such as values beyond 1.5 times the interquartile range. Removing or adjusting outliers is especially relevant in metrics like income or crime rates, where extreme values might skew averages.
- **Method**: Use Power BI's visualization or statistical tools to identify outliers, then decide whether to exclude them or adjust their influence in calculations, preserving the accuracy of analytical insights.

2.2 DATA TRANSFORMATION

Data transformation is a critical step in preparing the dataset for analysis, ensuring that the data is clean, accurate, and ready to generate meaningful insights. It involves integrating multiple data sources, cleaning the data, creating new features, structuring the data, and validating its accuracy. Here's an expanded overview of the key steps involved in transforming crime data:

Integration of Data Sources

- **Combining Datasets**: By merging data from various sources (e.g., police reports, crime logs, geographic data), we create a comprehensive and unified view of crime incidents. This allows for a more holistic analysis that considers multiple factors such as location, time, and type of crime.
- **Data Linking**: In cases where datasets are not directly comparable, we use linking keys (e.g., unique crime IDs, geographical identifiers) to integrate data from different regions or reporting agencies into a single dataset.

Data Cleaning

- **Missing Values**: Missing data is a common challenge in crime datasets. It is addressed by either replacing the missing values with averages, medians, or other calculated default values, or by removing rows containing missing information. This ensures that the analysis does not get distorted by incomplete data.
    - **Method**: Use Power BI's Query Editor to replace missing values using "Replace Values" or remove null rows with "Remove Rows."
- **Removing Duplicates**: Duplicated entries can distort the analysis, leading to inaccurate conclusions. Removing duplicates ensures that each crime incident is counted only once, making the dataset more reliable.

- **Method**: In Power BI, use the “Remove Duplicates” function in Query Editor to eliminate repeated entries based on a unique identifier like the crime ID.

- **Data Type Conversion**: Ensuring that each column in the dataset has the correct data type (e.g., numerical, date, text) is essential for accurate analysis. Incorrect data types can cause errors in calculations, visualizations, and statistical methods.
  - **Method**: In Query Editor, use the “Data Type” option to convert columns into the appropriate format (e.g., changing text to numbers or dates).

Feature Engineering

- **Time-Based Features**: Temporal features such as the year, month, and day of a crime incident are critical for analyzing crime trends over time. By extracting these features from the crime occurrence time, we can identify seasonal patterns and fluctuations in crime rates.
  - **Example**: Extracting the month and year allows analysis of seasonal crime trends (e.g., higher crime rates during holidays or specific months).
- **Geographic Indicators**: Using latitude and longitude coordinates helps in spatial analysis. By plotting these coordinates on a map, we can identify high-crime areas (hotspots), allowing for targeted interventions and resource allocation.
  - **Example**: Mapping crime incidents using latitude and longitude helps visualize patterns in urban areas or specific neighborhoods, making it easier to deploy law enforcement strategically.

Data Structuring

- **Normalization**: Standardizing formats (e.g., date formats, address formats) and naming conventions (e.g., consistent naming for crime types or locations) ensures that the dataset is uniform and easier to interpret.
    - **Example**: Converting all date formats to "YYYY-MM-DD" ensures consistency across the dataset, allowing for easier comparisons and calculations.
- **Hierarchical Structuring**: Organizing the dataset in a way that reflects relationships between different attributes helps provide context and deeper insights. For example, grouping crime types under broader categories (e.g., violent crimes, property crimes) can help in better analysis and policy-making.
    - **Example**: Structuring crime incidents by area, time, and crime type allows for a layered view of the dataset, making it easier to identify trends and patterns.

Data Aggregation

- **Group By Operations**: Aggregating data based on certain attributes (e.g., crime type, location, time) allows for summarizing large datasets into more manageable insights. This helps in identifying overall trends such as crime rates over time or by geographic area.
    - **Example**: Grouping data by month to calculate monthly crime totals or by neighborhood to identify areas with higher crime rates.
- **Calculating Metrics**: Deriving key metrics such as total crimes per area, crime rates per population, and percentage changes in crime frequency provides valuable insights for decision-makers. These metrics are crucial for assessing crime trends and allocating resources effectively.
    - **Example**: Calculating the crime rate per 1,000 people in a neighborhood to compare it to other areas.

Data Formatting

- **Standardizing Formats**: Ensuring consistency in data formatting, such as unifying date formats and currency formats, makes the data more readable and easier to analyze.
    - **Example**: Standardizing dates to “DD-MM-YYYY” format ensures that all records use the same format, reducing potential errors during analysis.
- **Text Formatting**: Cleaning up and clarifying descriptive columns (such as crime descriptions or victim information) enhances the readability of the dataset, ensuring that key information is understood and accurately interpreted.
    - **Example**: Converting crime type descriptions to a standard format, such as "Theft" instead of "Larceny/Theft."

Data Validation

- **Cross-Referencing**: After transforming the data, it is essential to cross-check it against the original sources to ensure accuracy. This step verifies that the transformation process has not introduced errors or inconsistencies.
    - **Example**: Comparing transformed data with official records to ensure that all crime incidents match the original reports.
- **Statistical Validation**: Statistical methods (such as summary statistics, distributions, and variance analysis) are used to confirm that the data adheres to expected patterns and characteristics. This helps detect outliers or anomalies in the data that might need further investigation or correction.
    - **Example**: Using statistical tests to verify that crime rates fall within expected ranges based on historical data.

2.3 DATA DISTRIBUTION USING CHARTS

Visualizing data distribution through charts is essential for understanding trends, identifying patterns, and making informed decisions based on crime data. Below are the various types of visualizations that can help analyze and interpret crime data:

Bar Charts

- **Purpose**: Display the frequency of different crime types.
- **Insight**: Bar charts provide a clear comparison of crime frequencies, helping identify the most common crimes. This insight is valuable for law enforcement agencies to prioritize prevention and intervention efforts in the most prevalent crime categories.
    - **Example**: A bar chart could display crime types such as theft, assault, and robbery, showing which crime is most frequent.

Pie Charts

- **Purpose**: Illustrate the proportion of various crime categories.
- **Insight**: Pie charts offer an intuitive way to show how different crimes contribute to the overall dataset. This visualization helps stakeholders understand the relative size of each category and where resources might be best allocated.
    - **Example**: A pie chart could show the proportion of property crimes, violent crimes, and drug-related offenses within a city.

Line Charts

- **Purpose**: Show trends in crime rates over time.
- **Insight**: Line charts are ideal for illustrating how crime rates fluctuate over time. By plotting crime incidents across months, seasons, or years, line charts can help identify peak periods for specific types of crime.

- **Example**: A line chart could show the rise in thefts during holiday seasons or the decrease in crime during a police crackdown.

Heat Maps

- **Purpose**: Visualize crime density across geographical locations.
- **Insight**: Heat maps provide a visual representation of crime hotspots, helping law enforcement agencies focus resources and efforts in high-risk areas. The density of crimes in various locations can be clearly seen, facilitating better spatial planning.
  - **Example**: A heat map could display crime locations within a city, with darker areas representing higher crime rates.

Histograms

- **Purpose**: Represent the distribution of numerical data, such as victim age or time of occurrence.
- **Insight**: Histograms show the distribution of numerical values and help identify patterns in the dataset. For example, histograms can help detect which age groups are most affected by specific crimes or the times of day when crime rates peak.
  - **Example**: A histogram could show the distribution of crimes by the age of victims, identifying which age groups are more vulnerable.

Scatter Plots

- **Purpose**: Analyze the relationship between two variables, such as crime rate vs. area population.
- **Insight**: Scatter plots help identify correlations between different variables. By visualizing how crime rates relate to factors like population density, income level, or neighborhood size, scatter plots can reveal underlying patterns and guide decision-making.

- **Example**: A scatter plot could illustrate the relationship between crime rates and neighborhood population size, indicating whether larger areas experience more crimes.

Box Plots

- **Purpose**: Summarize the distribution of a dataset and identify outliers.
- **Insight**: Box plots provide a summary of the data distribution, including the median, quartiles, and potential outliers. This visualization is particularly useful for detecting anomalies or extreme values in the crime data, such as unusual spikes in certain crime types.
  - **Example**: A box plot could show the distribution of crime incidents across different neighborhoods, highlighting outliers or areas with unusually high or low crime rates.

2.4 DATA MODELLING

Data modeling is a crucial step in transforming raw data into a structured format that supports effective analysis and visualization.

1. **Data Relationships**
   - **Definition:** Establishing connections between such as crimes, locations.
   - **Purpose:** Enables comprehensive analysis by allowing users to explore how different variables interact with each other.
2. **Data Tables**
   - **Fact Tables:** Contain quantitative data for analysis (e.g., number of crimes).
   - **Dimension Tables:** Hold descriptive attributes related to the facts (e.g., crime type, victim demographics).
   - **Purpose:** Provides a structured database that simplifies querying and enhances performance.

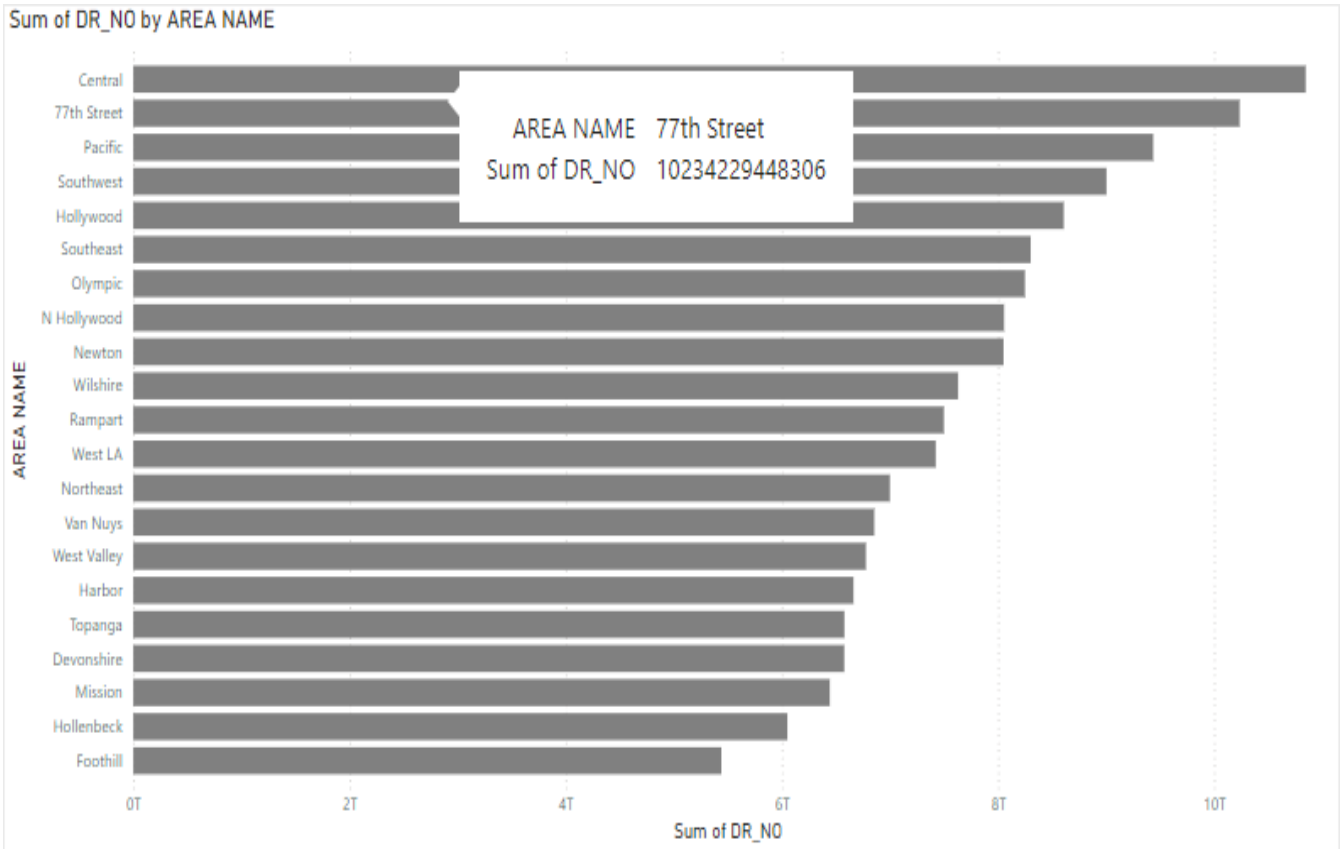
3. **Normalization**
    - **Definition:** Organizing data to minimize redundancy and improve data integrity.
    - **Purpose:** Ensures that data is stored efficiently and reduces the risk of inconsistencies.
4. **Calculated Columns and Measures**
    - **Calculated Columns:** Created to derive new data points from existing data (e.g., calculating the age of victims).
    - **Measures:** Aggregated calculations (e.g., total number of crimes)
    - **Purpose:** Enhances analytical capabilities by allowing custom calculations within the model.
5. **Hierarchies**
    - **Definition:** Structuring data into levels (e.g., Year > Month > Day) to facilitate drill-down analysis.
    - **Purpose:** Enables users to navigate data at various levels of granularity, enhancing insights.
6. **Data Types and Formats**
    - **Definition:** Ensuring each data column is assigned the appropriate data type (e.g., Date, Number, Text).
    - **Purpose:** Improves the accuracy of calculations and visualizations by avoiding data type mismatches.
7. **Performance Optimization**
    - **Techniques:** Implementing indexing, aggregating data, and reducing the volume of data loaded into the model.
    - **Purpose:** Enhances query performance and responsiveness of dashboards and reports.

CHAPTER 3
DATA ANALYSIS AND INTERPRETATION

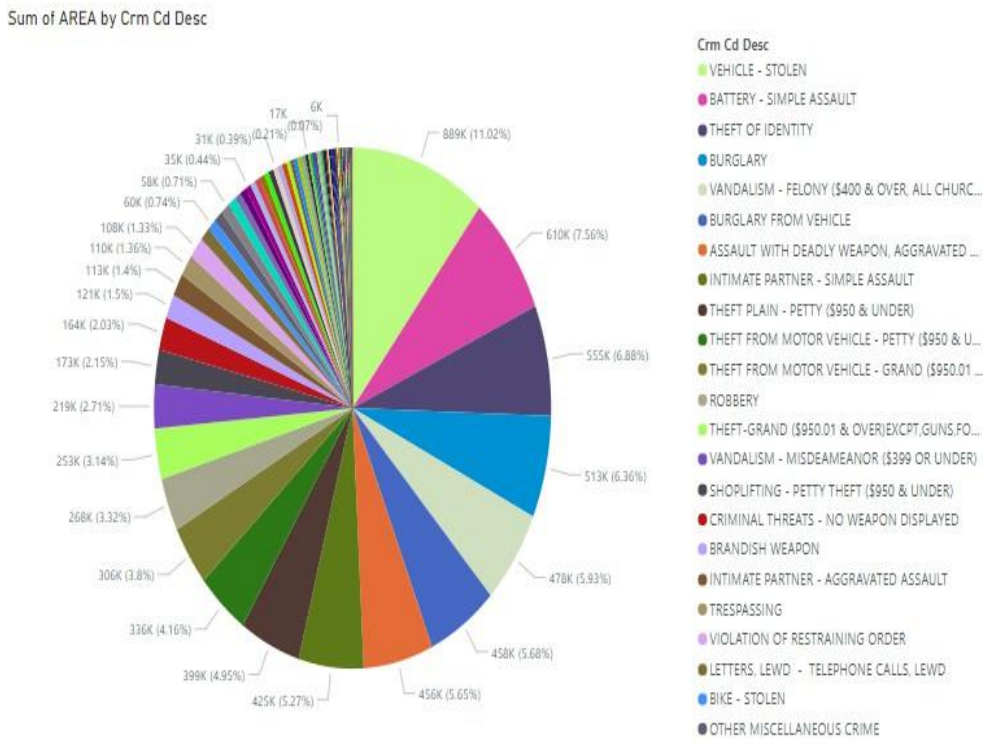
3.1 DATA ANALYSIS

There are various analytical approaches available for examining the data, and several critical questions have been explored in this process.

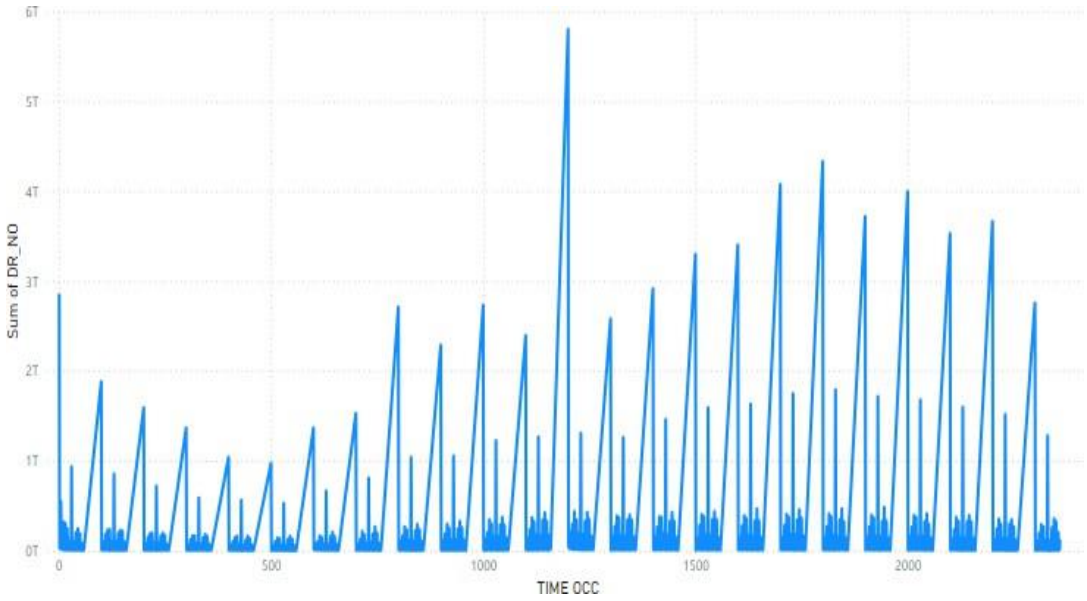
1. **What is the total number of crimes reported in each area?**



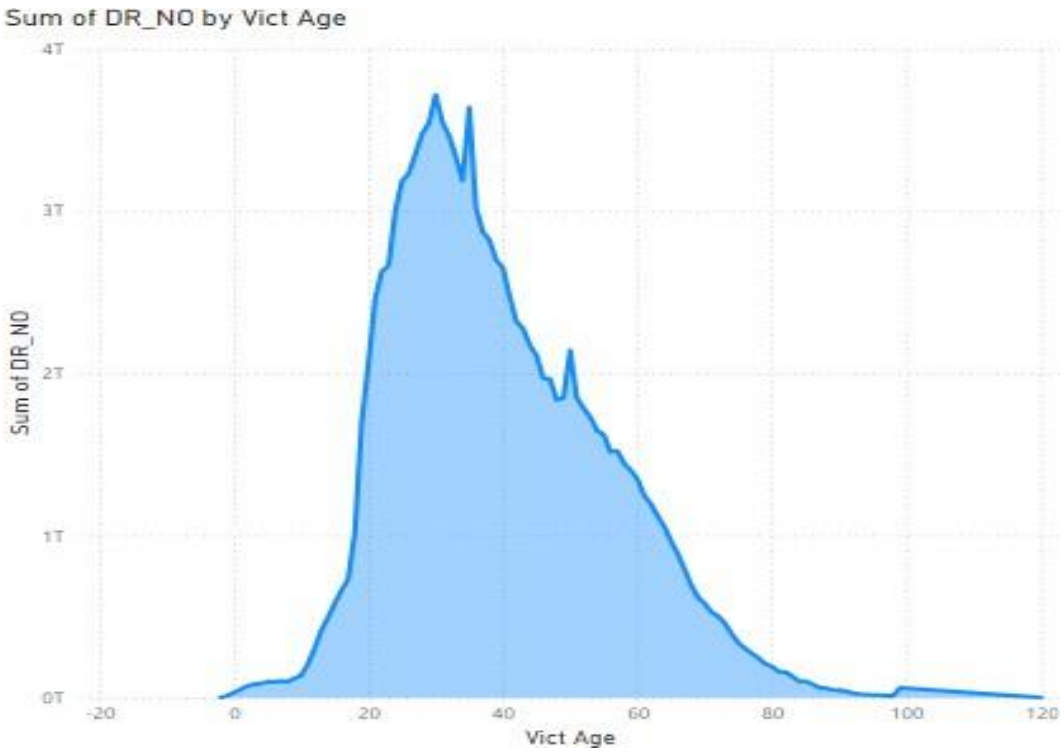
2. What is the distribution of crime types across different areas?



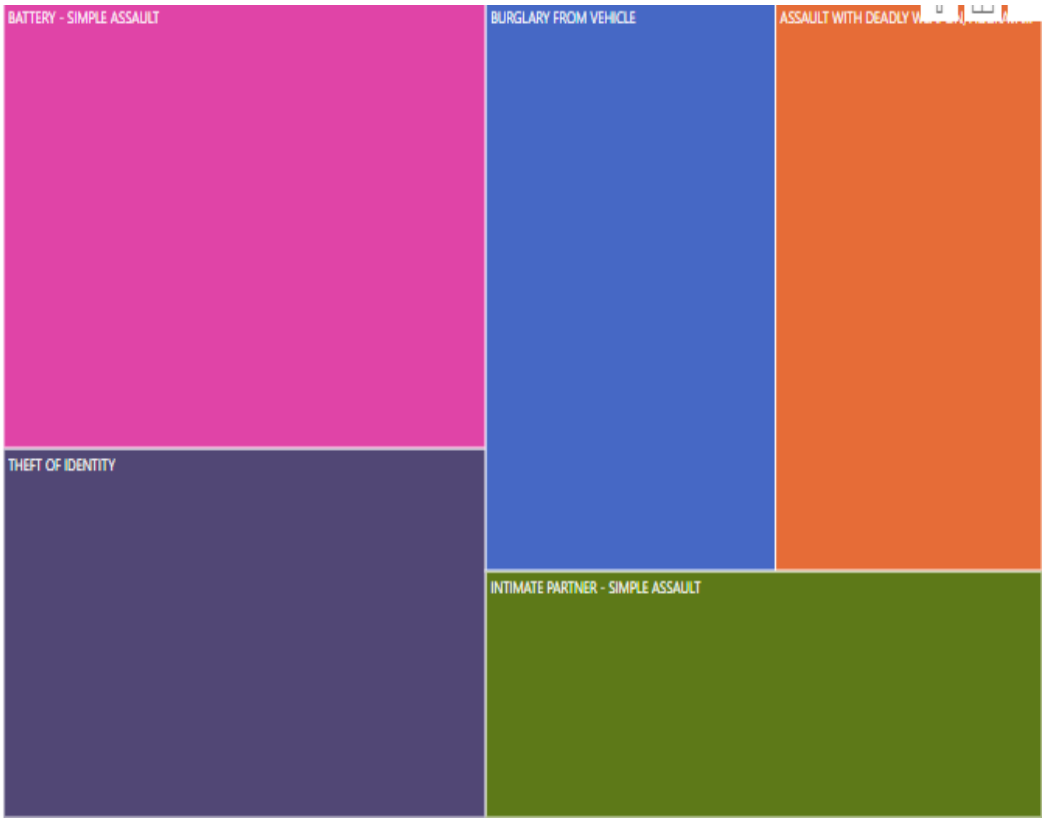
3. How does crime frequency vary by time of day?



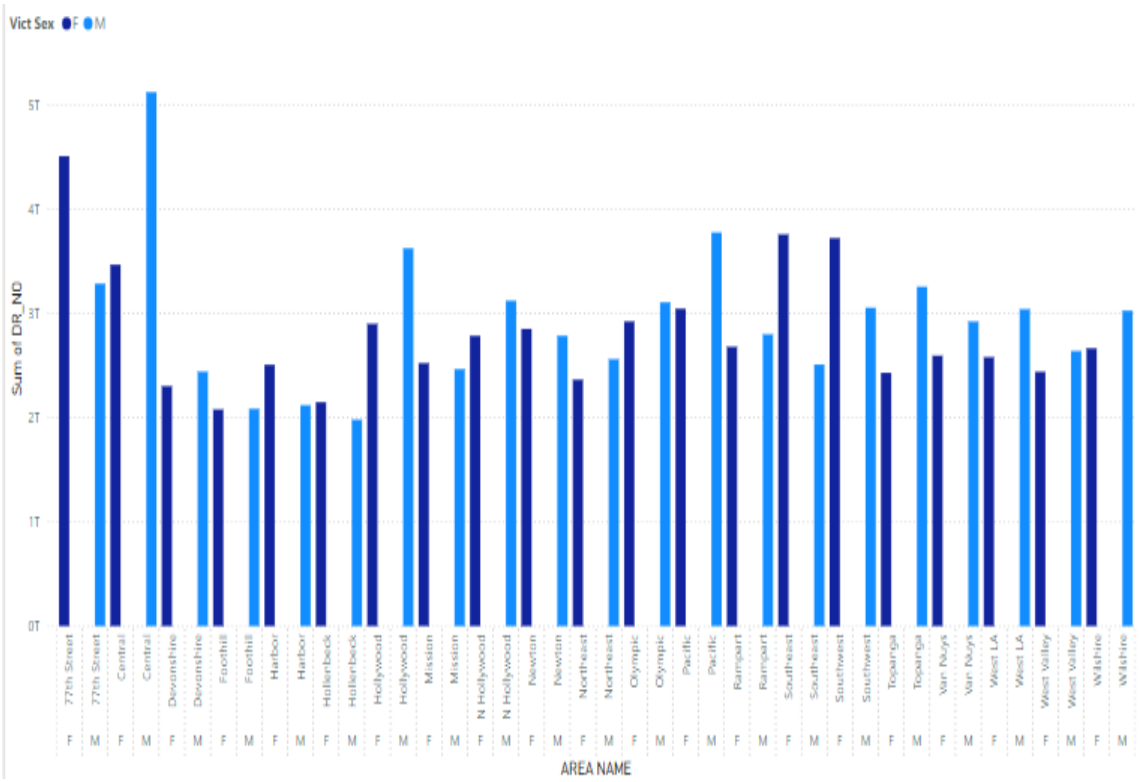
4. How do crime rates differ across various age groups of victims?



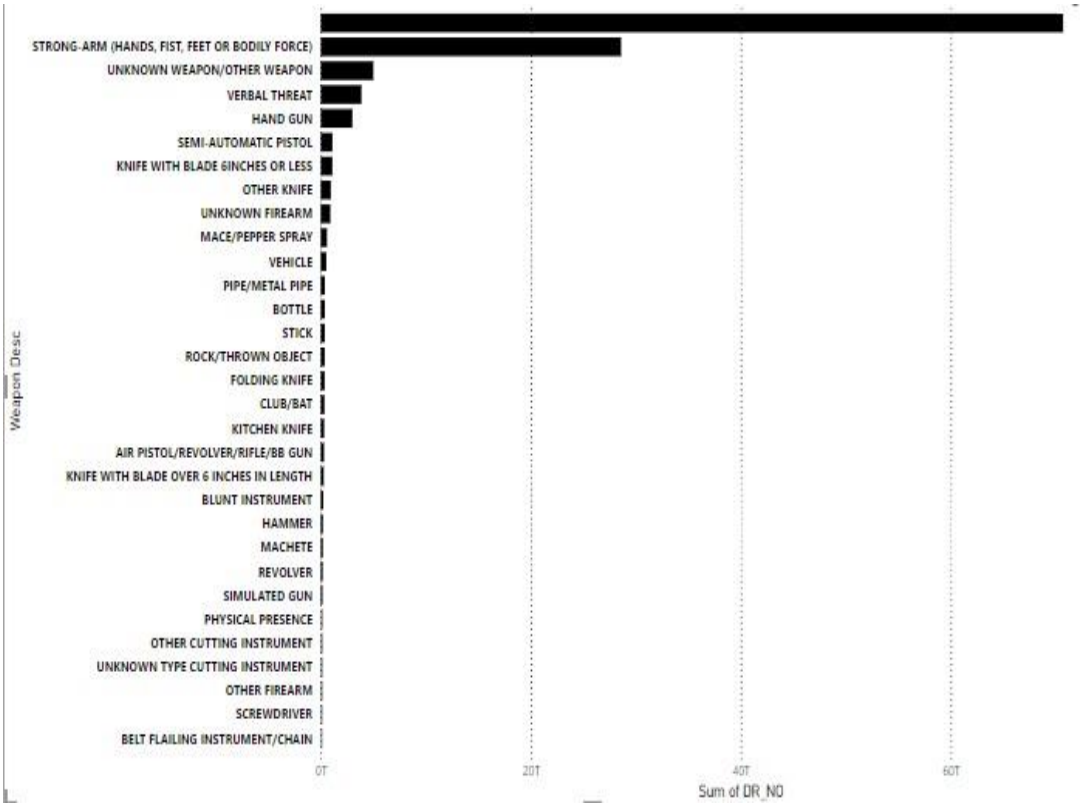
5. What are the top 5 most common crime categories?



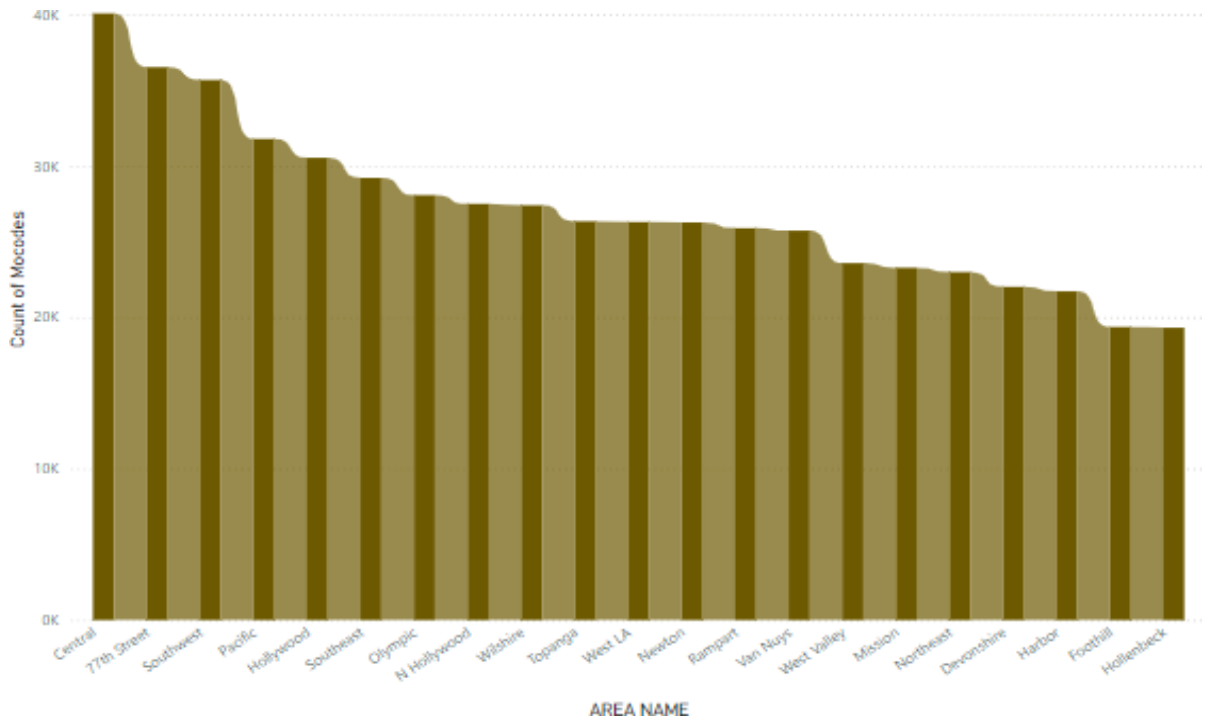
6. Find gender wise crime rate in specific area?



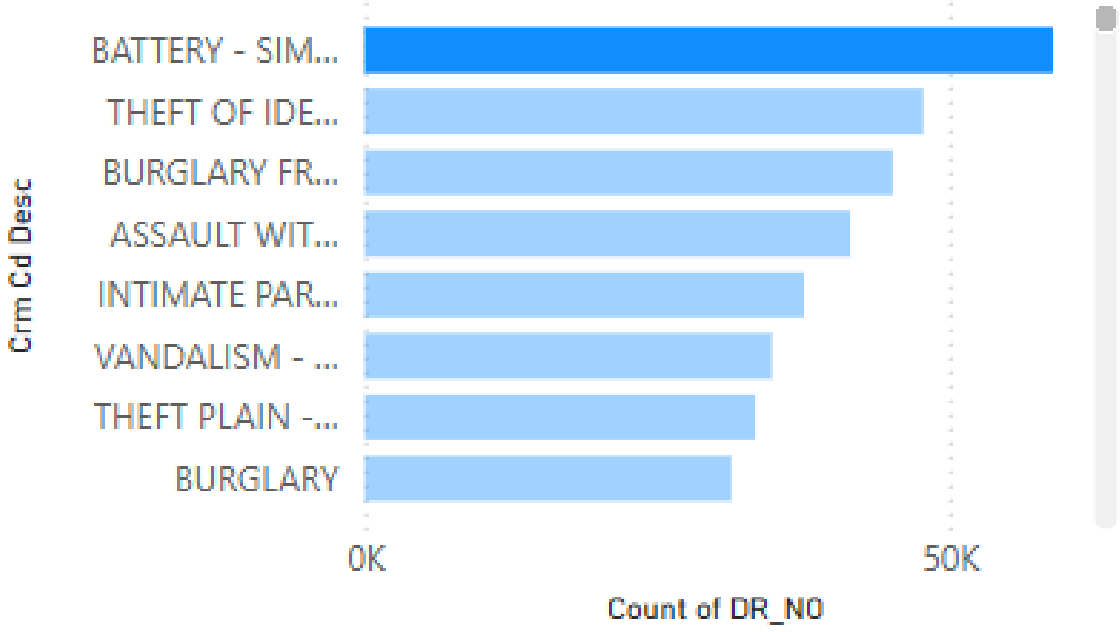
7. What types of weapon are used for the crimes?



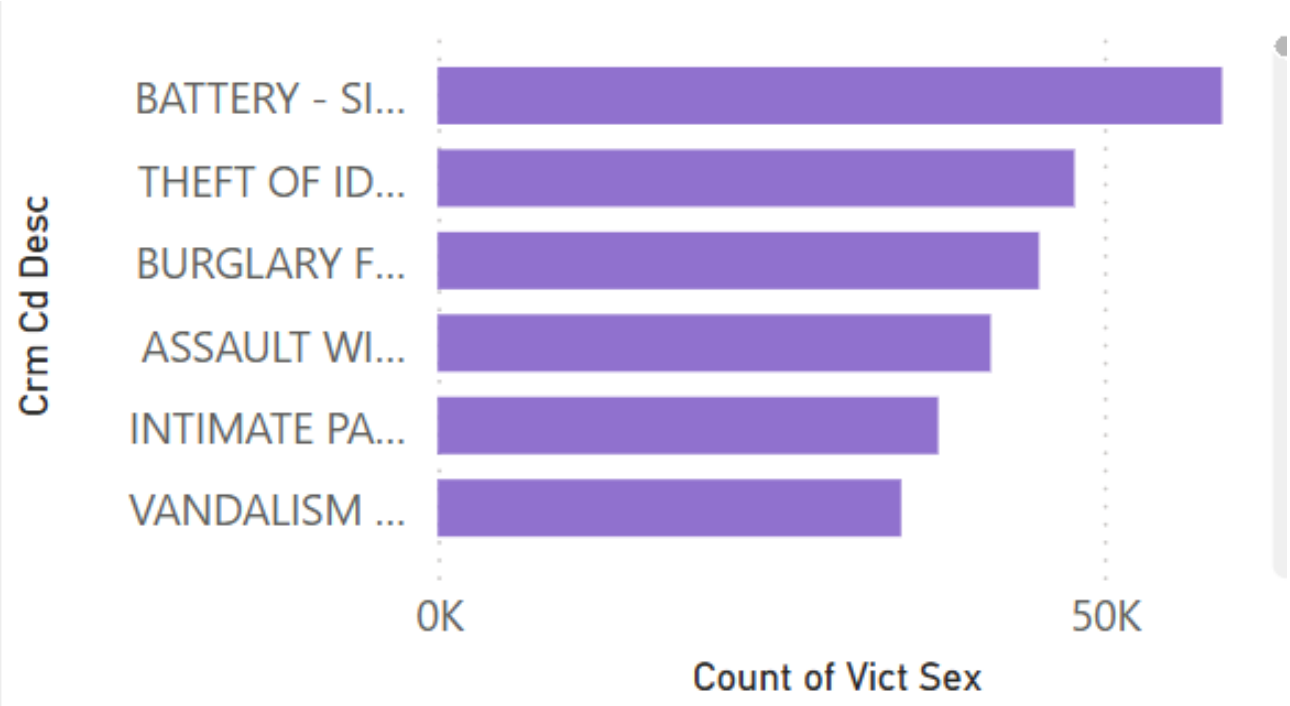
8. Find the highest mocodes with respect to the area name ?



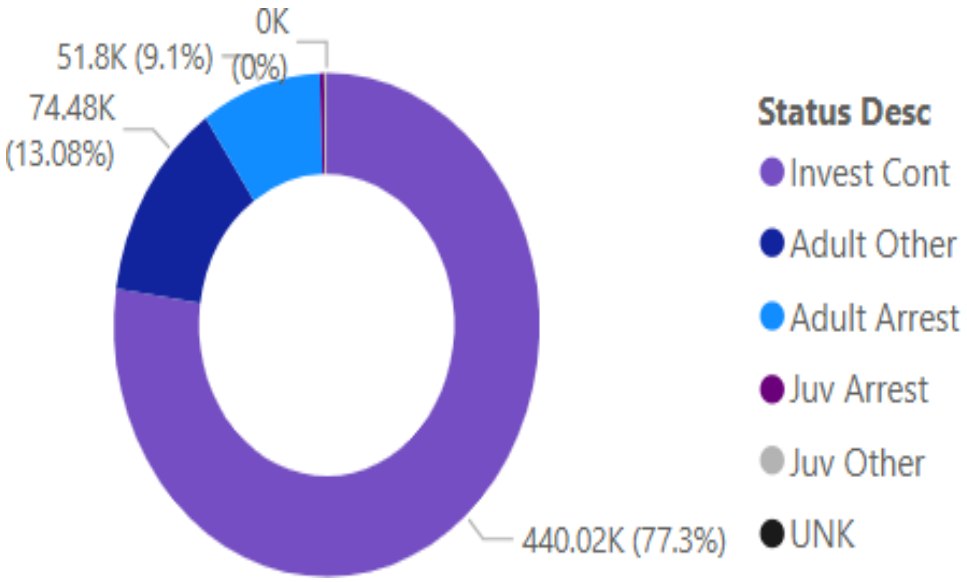
9. What is the distribution of crimes by type?



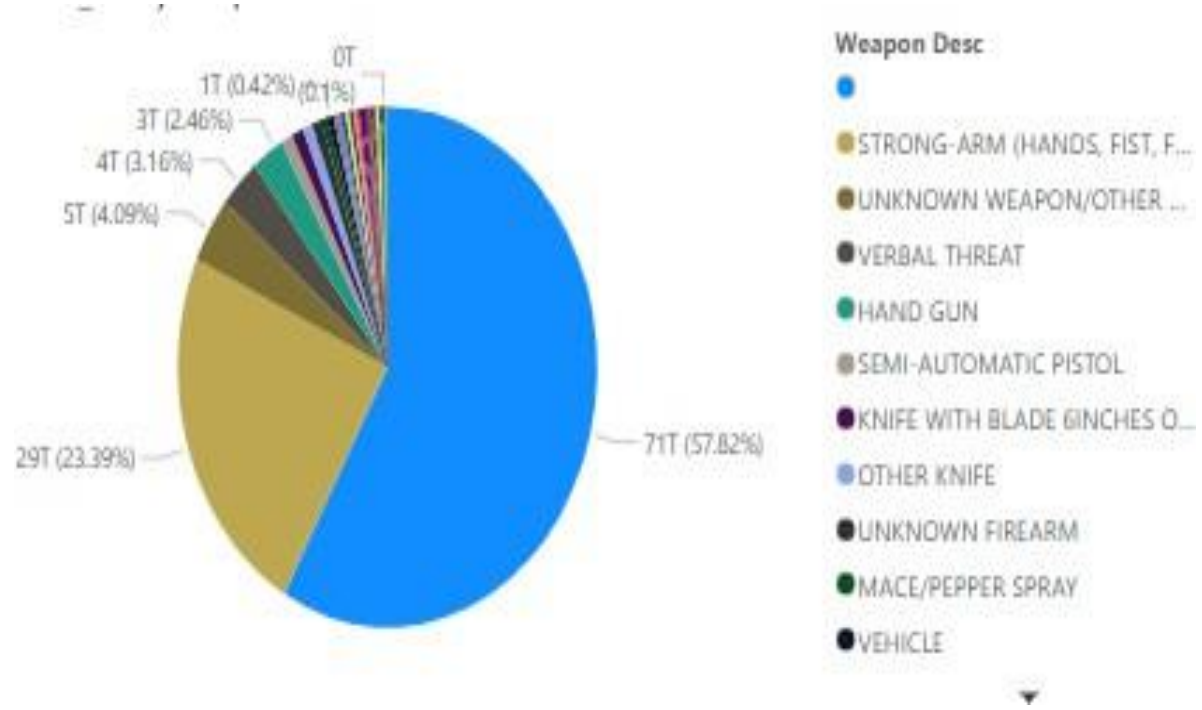
10. How does victim gender vary across different crime types?



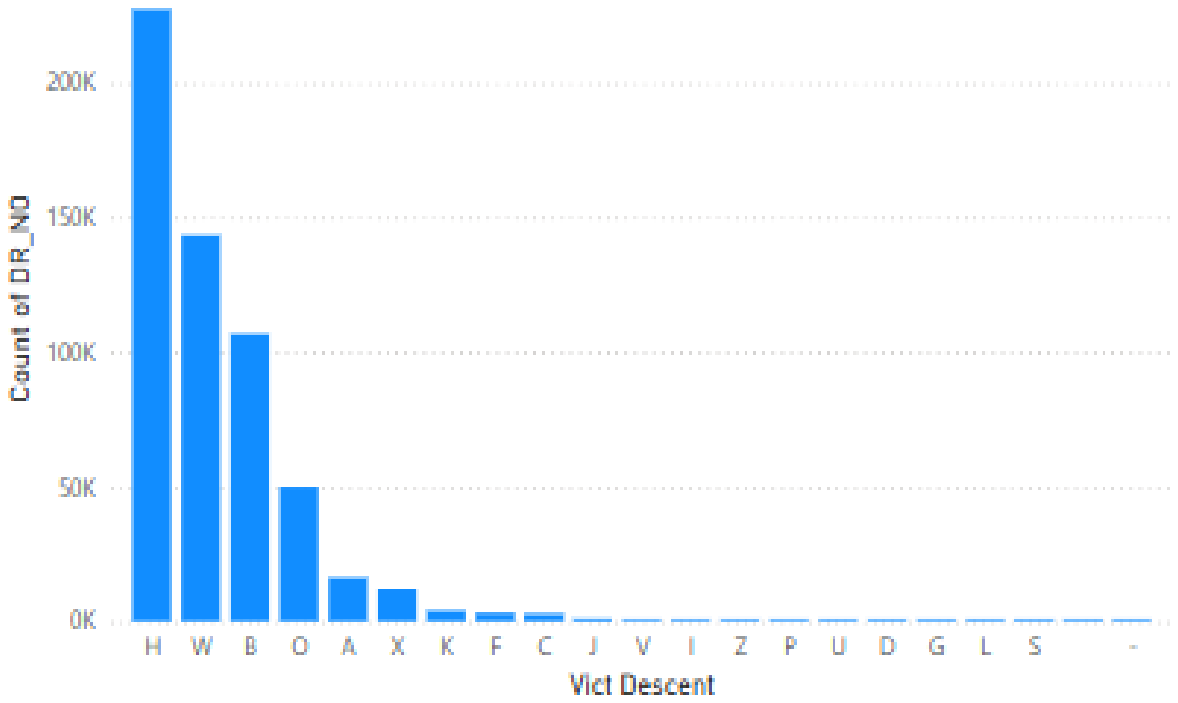
11. What is the status distribution of reported crimes?



12. What types of weapons are most commonly associated with crimes?



13. How does crime frequency vary by victim's descent?



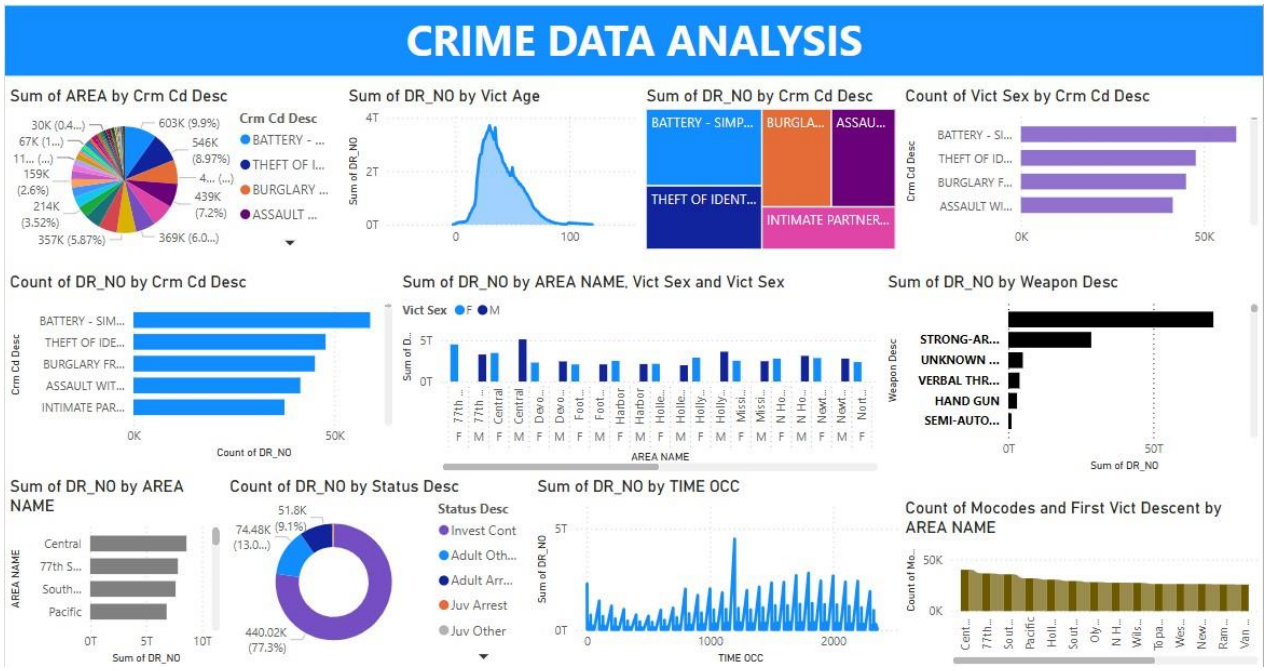
14. Which areas have the highest average victim age?

AREA NAME	Average of Vict Age
77th Street	38.52
Central	38.24
Devonshire	42.38
Foothill	40.51
Harbor	40.54
Hollenbeck	39.39
Hollywood	37.84
Mission	38.97
N Hollywood	40.29
Newton	37.38
Northeast	41.18
Olympic	38.78
Pacific	41.08
Rampart	37.60
Southeast	37.93
Southwest	36.17
Topanga	41.24
Van Nuys	41.03
West LA	43.15
West Valley	42.49
Wilshire	39.95
Total	**39.57**

15. What is the most common crime status in each area?

AREA NAME	AA	AO	CC	IC	JA	JO	**Total**
77th Street	3287	6265		26670	187	95	**36504**
Central	2393	3856	2	33750	40	11	**40052**
Devonshire	1842	3279		16731	68	84	**22004**
Foothill	2380	3091		13692	72	120	**19355**
Harbor	2839	3820		14916	101	32	**21708**
Hollenbeck	1767	3079		14309	95	54	**19304**
Hollywood	2391	2538		25512	41	39	**30521**
Mission	3120	3850		16094	90	111	**23265**
N Hollywood	3153	4337		19900	57	36	**27483**
Newton	2199	2685	1	21264	91	14	**26254**
Northeast	1879	3165		17812	73	46	**22975**
Olympic	2564	4250		21146	39	56	**28055**
Pacific	2091	2019		27584	50	27	**31771**
Rampart	2731	3903		19119	87	36	**25876**
Southeast	1817	3282		23918	143	34	**29194**
Southwest	3212	4021		28191	186	47	**35657**
Topanga	3025	3945		19108	141	110	**26329**
Van Nuys	2778	3801		19073	40	19	**25711**
West LA	1508	2449		22174	84	67	**26282**
West Valley	2897	4001		16526	120	25	**23569**
Wilshire	1929	2848		22532	55	24	**27388**
Total	**51802**	**74484**	**3**	**440021**	**1860**	**1087**	**569257**

3.2 PUBLISHING DASHBOARDS



3.3 INFERENCE

1. What is the total number of crimes reported in each area?

- The Central area has a lot of people, which can lead to more crimes.
- The Central area may have more economic problems, which can increase crime, while Foothill is more stable.
- Foothill might have better community safety programs, helping to keep crime rates low.

2. What is the distribution of crime types across different areas?

- The high number of vehicle thefts suggests a potential lack of security measures in the affected areas, making them attractive to thieves.
- Areas with higher vehicle theft rates may experience increased crime overall, indicating a need for enhanced law enforcement and community safety initiatives.
- Public awareness campaigns and improved parking facilities could be essential in reducing vehicle theft incidents in these regions.

3. How does crime frequency vary by time of day?

- Crime frequency tends to peak during late evening and early morning hours, suggesting that these times may pose higher risks for criminal activity.
- Increased criminal activity during these hours may indicate a correlation with reduced visibility and fewer people in public spaces, highlighting the need for enhanced patrols during these times.
- Understanding these patterns can help law enforcement agencies allocate resources more effectively, ensuring a stronger presence in high-risk areas during peak crime hours.

4. How do crime rates differ across various age groups of victims?

- The highest number of crime victims are aged 30, indicating that this age group may be more vulnerable or targeted for crimes.
- Very few crimes are reported for individuals over 90 years old, suggesting that elderly individuals may be less involved in criminal incidents, either as victims or due to their demographic characteristics.
- The age distribution highlights potential areas for focused crime prevention strategies aimed at younger adults, particularly around the age of 30.

5. What are the top 5 most common crime categories?

- Battery Simple Assault is the most common crime, suggesting frequent physical confrontations.
- Identity theft ranks high, showing a growing need for stronger data protection measures.
- Assault with Deadly Weapon highlights a significant level of violence, calling for improved safety efforts.

6. Find gender wise crime rate in specific area?

- Female crime rates are highest in the Central area, indicating it may be a hotspot for crimes against women.
- Male crime rates are highest in the 77th Street area, showing it as a key location for crimes involving men.
- These patterns suggest potential gender-based vulnerabilities or crime trends in these specific areas, which could guide targeted interventions.

7. What types of weapon are used for the crimes?

- **Strong Arm** is the most frequently used weapon in crimes, indicating a trend towards offenses relying on physical force over firearms or bladed weapons.
- This dominance of **Strong Arm** cases suggests that many crimes are likely close-contact, involving personal threats or assaults.
- The lower use of more lethal weapons like firearms may indicate a lesser presence of violent crime involving potentially fatal weapons in the region.
- These insights could guide law enforcement to focus on public safety measures, especially in high-density areas, to deter and reduce physical assault incidents.

8. Find the highest mocodes with respect to the area name ?

- Central area has the highest number of reported crimes associated with the **Mocodes**, indicating a significant concentration of criminal activity in this region.
- The high **Mocode** count in Central suggests a potential need for increased law enforcement presence and community safety measures to address the underlying issues contributing to crime.
- Comparing other areas to Central can help identify trends and patterns, allowing for targeted interventions and resource allocation to regions with rising crime rates.
- Understanding the factors contributing to the high **Mocode** rate in Central can lead to effective crime prevention strategies and community engagement initiatives to enhance safety.

9. What is the distribution of crimes by type?

- The Central area shows the highest concentration of reported crimes, as seen in the frequency of crime reports, suggesting a need for targeted law enforcement presence and community safety measures.
- Analyzing the distribution of crime types reveals that certain offenses are more common, allowing authorities to prioritize these areas with tailored interventions.
- Patterns observed in modification codes across different regions provide insights into specific crime behaviors, guiding focused preventative strategies.
- Additionally, mapping crime locations identifies city hotspots, enabling targeted patrols and resource allocation to improve safety in high-crime areas.

10. How does victim gender vary across different crime types?

- Analyzing victim gender across different crime types reveals gender-based patterns in victimization, which can guide targeted awareness and prevention efforts.
- Certain crimes show a higher prevalence among either male or female victims, highlighting potential vulnerabilities and areas where support may be needed.

- Identifying these trends enables law enforcement and social services to allocate resources more effectively, tailoring responses to the demographics most affected by specific crimes.
- This approach fosters a proactive community safety strategy, providing education and support to those at greater risk, improving overall crime prevention.

11. What is the status distribution of reported crimes?

- Analyzing crime status distribution shows the efficiency of case resolutions.
- Distinguishes between ongoing investigations and resolved cases.
- Helps law enforcement assess performance and allocate resources effectively.
- A higher proportion of resolved cases indicates effective management.
- Unresolved cases highlight areas needing improvement.
- Supports transparency, builds public trust, and enhances community safety strategies.

12. What types of weapons are most commonly associated with crimes?

- Analyzing weapon types in crimes helps identify the most commonly used weapons.
- Visualizing the data by weapon type reveals trends and prevalence.
- Identifies which weapons are frequently involved in criminal activities.
- Provides insights that can guide law enforcement in implementing targeted weapon control measures.
- Helps allocate resources for controlling and preventing the use of specific weapons in crimes.
- Supports proactive crime prevention strategies by addressing the most common weapon-related threats.

13. How does crime frequency vary by victim's descent?

- Analyzing crime frequency by victim's descent identifies demographic patterns in victimization.
- A bar chart visualizes the variation in crime frequency across different descent categories.
- Highlights specific groups that may experience higher rates of victimization.
- Provides insights for law enforcement and social services to design targeted outreach and support programs.
- Facilitates the allocation of resources to address the needs of at-risk demographic groups.
- Informs evidence-based strategies to enhance community safety and support services.

14. Which areas have the highest average victim age?

- Identifies locations with older or younger victim demographics.
- A bar chart or table displays the average victim age across areas.
- Highlights regions where specific age groups may be more affected by crime.
- Helps law enforcement focus on age groups vulnerable to certain types of crime.
- Assists community services in developing age-specific support programs.
- Guides resource allocation, focusing preventive measures in areas with at-risk age groups.

15. What is the most common crime status in each area?

- Reveals the progression and resolution of incidents across locations.
- A stacked column chart or matrix visualizes the distribution of statuses (e.g., reported, under investigation, closed).
- Identifies areas with higher numbers of unresolved or open cases.
- Aids law enforcement in prioritizing areas with unresolved cases for efficient case management.
- Provides data for policymakers to assess the effectiveness of crime resolution processes.

CHAPTER 4

CONCLUSION

4.1 RECOMMENDATIONS

In conclusion, this Crime Data Analysis project serves as a critical tool for enhancing the understanding of crime patterns and providing actionable insights for law enforcement agencies, urban planners, and policymakers. By leveraging Power BI's advanced data processing and visualization capabilities, the project not only simplifies the analysis of large, complex datasets but also transforms raw crime data into clear, actionable information. The integration of various data fields, including crime types, locations, timings, and demographic information, enables a comprehensive analysis that highlights trends, hotspots, and vulnerable populations. This allows for more informed decision-making and targeted interventions to prevent and reduce crime.

The expected outcome of this project is to provide law enforcement agencies with the tools needed to proactively address crime, allocate resources efficiently, and ultimately improve public safety. By providing policymakers with a deeper understanding of crime dynamics, the project also contributes to the development of more effective, evidence-based policies. The ability to analyze crime data visually and interactively is an essential step toward creating safer communities through data-driven strategies, which can have a long-term positive impact on both crime prevention and community well-being.

CHAPTER 5

REFERENCES

1) McClendon, Lawrence, and Natarajan Meghanathan. "Using machine learning algorithms to analyze crime data." *Machine Learning and Applications: An International Journal (MLAIJ)* 2.1 (2015): 1-12.
2) Shermila, A. Mary, Amrith Basil Bellarmine, and Nirmala Santiago. "Crime data analysis and prediction of perpetrator identity using machine learning approach." *2018 2nd international conference on trends in electronics and informatics (ICOEI)*. IEEE, 2018.
3) Safat, Wajiha, Sohail Asghar, and Saira Andleeb Gillani. "Empirical analysis for crime prediction and forecasting using machine learning and deep learning techniques." *IEEE access* 9 (2021): 70080-70094.
4) Kim, Suhong, et al. "Crime analysis through machine learning." *2018 IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)*. IEEE, 2018.
5) Bandekar, Shraddha Ramdas, and C. Vijayalakshmi. "Design and analysis of machine learning algorithms for the reduction of crime rates in India." *Procedia Computer Science* 172 (2020): 122-127.
6) Aziz, Rabia Musheer, et al. "Machine learning-based soft computing regression analysis approach for crime data prediction." *Karbala International Journal of Modern Science* 8.1 (2022): 1-19.
7) Ateş, Emre Cihan, Gazi Erkan Bostancı, and Serdar Msg. "Big data, data mining, machine learning, and deep learning concepts in crime data." *Journal of Penal Law and Criminology* 8.2 (2020): 293-319.
8) Mahmud, Sakib, Musfika Nuha, and Abdus Sattar. "Crime rate prediction using machine learning and data mining..*(IC3 2020)*. Springer Singapore, 2021.