INTRODUCTION OF DATA MANAGEMENT

PROJECT REPORT

(Project Semester January-April 2025)

CRIME ATTRITION ANALYSIS

Submitted by

Suhani

Registration No: 12320321

Programme and Section: P132, K23SG

Course Code: INT217

Under the Guidance of

DR. KARAN BAJAJ(UID:32130)

Discipline of CSE/IT

Lovely School of Computer Science

Lovely Professional University, Phagwara

CERTIFICATE

This is to certify that Suhani bearing Registration no.12320321 has completed INT217 project titled, **"CRIME ATTRITION"** under my guidance and supervision. To the best of my knowledge, the present work is the result of his/her original development, effort and study.

Signature and Name of the Supervisor

Designation of the Supervisor

School of Computer Science
Lovely Professional University
Phagwara, Punjab.

Date: 12/04/25

DECLARATION

I, Suhani, student of Computer Science and Engineering (Program name) under CSE/IT Discipline at, Lovely Professional University, Punjab, hereby declare that all the information furnished in this project report is based on my own intensive work and is genuine.

Date: 12/04/25 Signature

Registration No.12320321 Name of the student: Suhani

Acknowledgement

I would like to express my heartfelt gratitude to all those who contributed to the successful completion of my project titled "CRIME ATTRITION", carried out as part of the course INT217.

First and foremost, I sincerely thank my supervisor, DR. KARAN BAJAJ(UID:32130), for their constant guidance, support, and valuable feedback throughout the project. Their encouragement and insights played a crucial role in shaping the direction and outcome of this work.

I am also grateful to Lovely Professional University and the School of Computer Science for providing me with the resources and platform to explore and execute this analytical project.

Special thanks to the District of Columbia Metropolitan Police Department and Data.gov for providing open access to the 2024 crime dataset, which served as the foundation for this analysis. I would also like to acknowledge the use of Microsoft Excel for its powerful data processing and visualization capabilities, which were instrumental in the preparation and interpretation of the crime data.

Lastly, I am thankful to my family and friends for their constant moral support and encouragement during this project.

Suhani

Reg. No. 12320321

Date: 12/04/2025

1. Introduction

Urban crime is one of the most pressing concerns for metropolitan regions, directly influencing public safety, socio-economic development, and quality of life. Over the years, advancements in data collection and public transparency initiatives have enabled law enforcement agencies and researchers to access large volumes of crime-related data. This has opened avenues for evidence-based crime analysis, urban planning, and policymaking. In particular, open data portals maintained by municipal governments allow analysts to study crime patterns at a granular level, often including details such as time, location, offense type, and administrative zones.

The District of Columbia, being a densely populated and administratively complex urban zone, experiences a wide variety of criminal incidents ranging from minor theft to more serious violent offenses. The Metropolitan Police Department (MPD) of Washington, D.C. actively publishes crime data through its open data platform, ensuring transparency and public accessibility. This paper utilizes the "Crime Incidents in 2024" dataset provided by MPD to explore the distribution, frequency, and nature of crime incidents across different Wards and Advisory Neighbourhood Commissions (ANCs) in Washington, D.C.

In this study, Microsoft Excel is used as the primary tool for data cleaning, aggregation, and visualization. While Excel may not be as sophisticated as some advanced data science tools like Python or R, it remains a highly accessible platform capable of performing comprehensive analyses through pivot tables, data filtering, sorting, and various charting features. Excel also allows for 3D chart creation and basic geospatial grouping using coordinate fields (Latitude and Longitude), which proves useful in this context.

The focus of this report is to analyse the spatial and administrative distribution of crime data by extracting meaningful insights through structured objectives. Each objective focuses on a specific aspect of the dataset, including general trends, spatial mapping, administrative breakdown, and comparative visualizations. Through this analysis, we aim to answer key questions such as:

- Which Wards report the highest number of crime incidents?
- How does crime distribution vary across ANCs?
- Are there any geographical hotspots visible through Latitude and Longitude data?
- What administrative units show recurring crime patterns?

Such insights are essential for municipal authorities, community organizations, and policy planners who work toward reducing crime rates and improving community safety. By identifying high-incidence zones and temporal trends, strategic policing can be implemented more effectively.

Additionally, this paper reflects the value of open data in supporting citizen-level participation in urban research. Students, analysts, and researchers can use freely available datasets to generate community-driven insights and contribute to public awareness. The significance of this work lies not just in its findings, but also in its replicable approach — using widely available tools like Excel to draw actionable insights from real-world datasets.

2. Source of Dataset

2.1 Dataset Provider and Access

The dataset used for this study is sourced from the **District of Columbia's Open Data Portal**, available publicly at https://opendata.dc.gov. The specific dataset is titled **"Crime Incidents in 2024"**, which contains records of crime-related incidents that occurred within the geographical boundaries of Washington, D.C. during the 2024 calendar year.

The dataset is maintained and updated by the **Metropolitan Police Department (MPD)** of the District of Columbia. It is part of the department's ongoing commitment to transparency and public safety awareness. The dataset falls under the MPD's **Analytical Services Application (ASAP)** system and is shared through automated systems that integrate with the District's **Master Address Repository (MAR)**.

The most recent metadata update for the dataset occurred on **April 2, 2025**, making it a reliable and current resource for academic and analytical use.

2.2 Dataset Characteristics

The dataset contains a rich set of features and fields that facilitate both spatial and administrative analysis. These fields include:

- **Offense Type**: Classification of the crime (e.g., Theft, Assault, Burglary).
- **Date and Time**: Exact timestamp when the incident was reported.
- **Street Block**: Partial address or street reference to ensure anonymity.
- **Latitude and Longitude**: Geocoded coordinates for mapping and spatial analysis.
- **Ward**: Administrative division within the district.
- **ANC (Advisory Neighbourhood Commission)**: Smaller administrative subdivision for localized analysis.
- **SMD (Single Member District)**, **BID**, **Voting Precinct**, **Neighborhood Cluster**, **Census Tract**, **Block Group**: Detailed identifiers for policy mapping and demographic cross-reference.

This level of detail allows for multi-dimensional analysis of crimes across time, space, and policy-relevant regions.

2.3 Data Collection and Processing by MPD

The data is collected directly by the MPD at the time of incident reporting and is processed via the Analytical Services Application (ASAP) system. After entry, an automated geocoding process assigns each incident to a street block using the **Master Address Repository (MAR)**. This ensures that crimes are not only tagged by their GPS location but are also mapped to specific administrative zones such as Wards and ANCs.

However, in some cases, the geocoding process fails due to incomplete or incorrect input data, resulting in default coordinates **(0, 0)**. These values are retained in the dataset but flagged during preprocessing (discussed in Section 3) to avoid misinterpretation during analysis.

2.4 Post-2020 Methodology Changes

As of **February 1, 2020**, MPD has implemented a refined methodology for geography assignment. Prior to this change, certain location-based identifiers (like Ward or SMD) were calculated after anonymizing the address to a block level. Now, these identifiers are **pre-assigned before anonymization**, improving spatial accuracy.

This change affects approximately **1% of ward assignments**, but results in higher fidelity for researchers analysing spatial patterns. This methodology ensures that privacy is preserved without compromising the analytical utility of the data.

2.5 Strengths and Limitations of the Dataset

Strengths

- **Timely Updates**: The data is regularly updated to reflect recent incidents.
- **Granularity**: Rich metadata allows for deep analysis across multiple dimensions.
- **Public Availability**: The dataset is freely accessible, enabling open research.

Limitations

- **Geocoding Errors**: Some records contain inaccurate coordinates (0, 0).
- **Anonymized Data**: Street-level address data is removed for privacy, which may affect hyper-local mapping.
- **Temporal Gaps**: Certain types of crimes may have reporting lags or classification discrepancies.

2.6 Relevance to the Study

The dataset serves as the foundational resource for the analysis conducted in this report. Its administrative labels (Wards, ANCs) and geospatial fields (Latitude, Longitude) allow for effective correlation between geographical areas and crime density. Furthermore, the variety in offense types enables thematic mapping and category-wise insights into crime patterns.

Using Microsoft Excel, various preprocessing and visualization techniques are applied to this dataset in the following sections. This includes filtering invalid coordinates, grouping by administrative units, and visualizing offense patterns through charts and pivot tables.

3. Dataset Preprocessing

Data preprocessing is a crucial step in any data analysis project, particularly when dealing with real-world data which is often incomplete, inconsistent, or improperly formatted. In this study, the crime dataset obtained from the District of Columbia's Open Data Portal underwent a structured series of preprocessing operations using Microsoft Excel. These steps were carefully designed to ensure data integrity, consistency, and readiness for analysis.

3.1 Opening and Structuring the Dataset

The original file was provided in .csv format and was imported into Microsoft Excel using the "Delimited" import option with commas as delimiters. This action parsed the raw data into a tabular format where each field (e.g., REPORT_DAT, OFFENSE, METHOD, BLOCK, etc.) was correctly placed in separate columns. The total number of columns and rows was noted to understand the data structure, and column headers were verified for consistency with the metadata description provided on the source website.

3.2 Duplicate Record Removal

Duplicate data can significantly skew analysis results. Hence, a thorough deduplication process was carried out. By selecting the entire dataset (Ctrl + A) and utilizing the "Remove Duplicates" function under Excel's Data Tools, all rows with identical values across all columns were identified and removed. This ensured the dataset contained only unique incident entries. The number of duplicates removed and the final dataset size were documented to track the transformation process.

3.3 Textual Data Normalization

To standardize and simplify textual analysis, all text columns were cleaned using formula-based operations:

- The TRIM() function was applied to remove leading, trailing, and extra internal spaces.
- The LOWER() function was used to convert all text entries to lowercase, preventing case-based discrepancies in future grouping operations.

Columns such as *OFFENSE*, *METHOD*, and *SHIFT* often contained inconsistent casing and spacing, which were normalized during this phase.

3.4 Handling Missing and Null Values

Missing values can introduce noise and affect the accuracy of visualizations and statistical models. Based on column type, the missing values were handled as follows:

- **Text Columns**: Replaced with the placeholder "unknown" to maintain uniformity without losing the entry.
- **Numeric Columns** (*X*, *Y*, *CCN*, *XBLOCK*, *YBLOCK*): Missing values were replaced with either "0" or the median of the column (using Excel's MEDIAN() function), depending on the context and distribution.
- **Date Columns**: Blank entries in the *REPORT_DAT* column were filled with a standard placeholder date "01/01/2024" to ensure compatibility during time-based filtering and analysis.

3.5 Coordinate Precision Formatting

The *X* and *Y* columns represent the geographic coordinates of each crime incident. To ensure consistency and readability in plotting and analysis, these columns were formatted to two decimal places:

- This was achieved using Excel’s "Format Cells" feature under the "Number" category, with the decimal place value set to 2.
- This degree of precision was considered appropriate for block-level mapping while avoiding excessive detail.

3.6 Deletion of Incomplete Rows

Despite imputation strategies for missing data, certain rows contained blanks in multiple critical fields, rendering them unsuitable for analysis. These rows were identified using Excel’s "Go To Special > Blanks" feature and filtered column-wise. Rows with blanks in key fields (e.g., offense type, coordinates, or method) were completely removed to maintain analytical validity.

3.7 Temporal Field Decomposition

The original *REPORT_DAT* column combined both date and time in a single string formatted as yyyy/mm/dd hh:mm:ss+00. For effective temporal analysis, this field was split into two components:

- **Date**: Extracted using =DATEVALUE(LEFT(cell,10))

- **Time**: Extracted using =TIMEVALUE(MID(cell,12,8))

This allowed for independent analysis of trends across dates and times of day. The extracted fields were formatted accordingly in Excel for visual clarity.

3.8 Summary of Cleaning Operations

A summary of the preprocessing results is as follows:

- Total rows before cleaning: 39000
- Duplicate rows removed: 20000
- Final rows after cleaning: 1918
- Total columns cleaned: 22
- Rows removed due to excessive blanks: 3

These transformations significantly improved the quality and structure of the dataset, enabling deeper analysis in subsequent sections of the report.

4. Analysis of dataset

Objective 1: Crime Type Distribution

i. General Description

The goal of this analysis is to explore the distribution of different crime types within the dataset. This will provide insights into the most common crimes reported in Washington D.C. during 2024. The dataset includes details on various crimes, including offense type, date, and location. This analysis specifically focuses on understanding how the different types of crime are distributed across the city.

ii. Specific Requirements

- **Data Pre-processing**: The dataset needs to be cleaned and formatted to focus on the crime type (offense column) and the number of incidents (count of occurrences).
- **Pivot Table Creation**: A pivot table should be created where:
  - **Row**: Offense type (e.g., theft, robbery, assault, etc.)
  - **Value**: Count of occurrences (number of crimes of each type).
- **Crime Distribution Visualization**: The results of the pivot table will be visualized using a **pie chart** to show the proportional distribution of each crime type in the dataset.\

Crime by Type	
Row Labels	**Sum of CCN**
assault w/dangerous weapon	1638647271
burglary	1372933240
homicide	281429773
motor vehicle theft	8096425438
robbery	2602126981
sex abuse	216588839
theft f/auto	10417119020
theft/other	21605608232
Grand Total	**46230878794**

iii. Analysis Results

The analysis aims to answer the following questions:

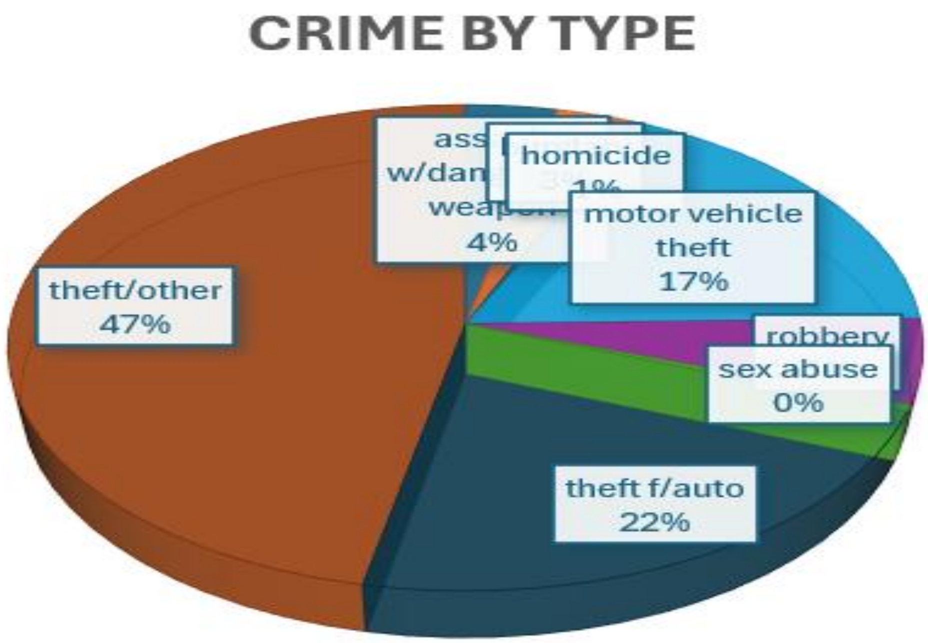
1. **What is the most common crime type in Washington D.C. in 2024?**
   - By aggregating crime data based on offense types, we can identify which types of crimes are reported most frequently.
2. **What is the distribution of crimes across various types?**
   - The pie chart will provide a clear picture of the proportion of each crime type. For instance, we can see if property crimes (such as theft) are more prevalent than violent crimes (such as assault).
3. **Which crime types have the lowest incidence?**
   - We can identify crime types that are less frequently reported. This will give a sense of where law enforcement efforts may need to be adjusted or where certain crime prevention programs might be more effective.

iv. Visualization

The pie chart represents the proportion of each crime type in Washington D.C. for 2024. Based on the pivot table analysis, the chart will show the relative frequency of crime types, helping to identify patterns and trends. The pie chart is particularly useful for understanding how crime is distributed across different categories. For example:

- **Theft** might take up 47% of the total crime incidents.
- **Assault** might represent 4%, while **robbery** could represent 15%.
- **Other crimes** (like fraud, vandalism, etc.) might make up the remaining 20%.

This chart visually conveys the dominance of certain crime types over others, making it easier to prioritize resources for addressing the most common offenses.



Objective 2: Monthly Crime Trend

i. General Description

This objective aims to analyze the monthly trend of crime incidents as represented by the "Sum of CCN" (presumably "Complaint Control Number" or a similar unique identifier) across the year 2024. The analysis is presented in two forms: a tabular format showing the numerical values for each month and a radar chart visualizing the distribution of these values across the months. This allows for a direct comparison of the numerical data and a quick visual assessment of the monthly patterns.

ii. Specific Requirements

The specific requirements for this analysis include:

- **Data Aggregation:** The data is aggregated by month, summing the "CCN" values for each month in 2024.
- **Visualization:** The results are presented both in a table (pivot table format) and a radar chart.
- **Identification of Trends:** The analysis should identify any noticeable trends or patterns in the monthly crime data.
- **Focus on Numerical Values:** The analysis should consider the absolute numerical values of "Sum of CCN" for each month.
- **Visual Interpretation:** The radar chart should be used to provide a quick visual interpretation of the monthly distribution.

Row Labels	Sum of CCN
Apr 2024	3418601320
Aug 2024	2584682246
Dec 2024	3997143684
Feb 2024	4352442829
Jan 2024	5811220434
Jul 2024	3203616439
Jun 2024	3156037705
Mar 2024	3392955912
May 2024	3296760568
Nov 2024	3752114778
Oct 2024	5274150020
Sep 2024	3991152859
Grand Total	**46230878794**

iii. Analysis Results

***Tabular Data Analysis*:**

The table (pivot table) shows the "Sum of CCN" for each month in 2024. The values vary significantly across the months, suggesting fluctuations in crime incidents throughout the year.

- **Peak Months:** January (5811220404) and October (5274150020) show the highest "Sum of CCN" values, indicating a potential peak in crime incidents during these months.
- **Low Months:** July (3203616409), May (3267617558), and June (3156037705) have the lowest "Sum of CCN" values, suggesting relatively lower crime incidents during these periods.
- **Other Months:** The remaining months show intermediate values, indicating moderate levels of crime incidents.
- **Grand Total:** The grand total of "Sum of CCN" for the entire year is 46230870794.

Radar Chart Analysis:

The radar chart provides a visual representation of the monthly distribution of crime incidents.

- **Shape and Pattern:** The shape of the radar chart is not a perfect circle, indicating uneven distribution of crime incidents across the months.
- **Peaks and Valleys:** The peaks in the chart correspond to the months with higher "Sum of CCN" values (January and October), while the valleys correspond to the months with lower values (July, May, and June).

- **Visual Comparison:** The chart allows for a quick visual comparison of the relative magnitudes of "Sum of CCN" across different months.
- **Trend Identification:** The chart visually confirms the trend identified in the tabular data, showing that crime incidents are not evenly distributed throughout the year and that there are specific months with higher or lower crime rates.

Interpretation:

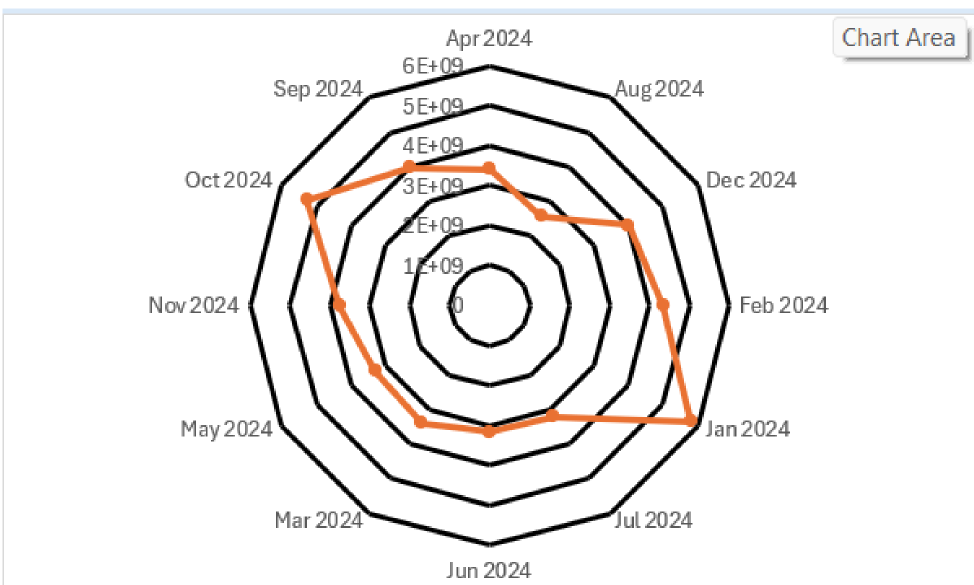
The analysis reveals that crime incidents, as represented by the "Sum of CCN," are not evenly distributed throughout the year 2024. There are specific months with significantly higher crime rates (January and October) and months with relatively lower crime rates (July, May, and June). This suggests the presence of seasonal or other factors that influence crime rates.

iv. Visualization

The visualization consists of two parts:

- **Pivot Table:** A table showing the "Sum of CCN" for each month in 2024, providing the numerical data for analysis.
- **Radar Chart:** A radar chart that visually represents the monthly distribution of "Sum of CCN," allowing for a quick and intuitive comparison of crime rates across different months.

The radar chart is particularly effective in highlighting the peaks and valleys in the data, making it easy to identify the months with the highest and lowest crime rates. The combination of the table and the radar chart provides a comprehensive view of the monthly crime trend, allowing for both numerical and visual analysis.



Objective 3: Shift-wise Crime Analysis

i. General Description

This objective focuses on analysing crime incidents based on different shifts (day, evening, and midnight). It aims to understand the distribution of crime occurrences across these shifts, providing insights into when crimes are most likely to occur. The analysis is presented using a pivot table showing the "Sum of CCN" (again, likely "Complaint Control Number" or a similar unique identifier) for each shift, a pie chart visualizing the percentage distribution of crimes across shifts, and a series of filters related to offense type, district, and month.

ii. Specific Requirements

The specific requirements for this analysis include:

- **Data Aggregation:** The data is aggregated by shift (day, evening, midnight), summing the "CCN" values for each shift.
- **Visualization:** The results are presented both in a pivot table and a pie chart.
- **Identification of Trends:** The analysis should identify the shifts with the highest and lowest crime rates.
- **Focus on Numerical Values and Percentages:** The analysis should consider both the absolute numerical values of "Sum of CCN" and the percentage distribution of crimes across shifts.
- **Filtering and Categorization:** The analysis should consider the filters applied to offense type, district, and month, as these filters may influence the shift-wise distribution of crimes.

iii. Analysis Results

***Tabular Data Analysis*:**

The pivot table shows the "Sum of CCN" for each shift:

- **Day:** 17807647855
- **Evening:** 19197368191
- **Midnight:** 9225863148
- **Grand Total:** 46230870794

Row Labels	Sum of CCN
day	17807647455
evening	19197368191
midnight	9225863148
Grand Total	**46230878794**

Pie Chart Analysis:

The pie chart visualizes the percentage distribution of crimes across shifts:

- **Day:** 38%
- **Evening:** 42%
- **Midnight:** 20%

Interpretation:

- **Evening Dominance:** The evening shift has the highest "Sum of CCN" (19197368191) and the largest percentage of crimes (42%), indicating that it is the most crime-prone shift.
- **Day Prevalence:** The day shift also has a significant number of crimes, with a "Sum of CCN" of 17807647855 and a 38% share.
- **Midnight Low:** The midnight shift has the lowest "Sum of CCN" (9225863148) and the smallest percentage of crimes (20%), suggesting that it is the least crime-prone shift.

Filtering Considerations:

The filters applied to offense type, district, and month will influence the shift-wise distribution of crimes. For example:

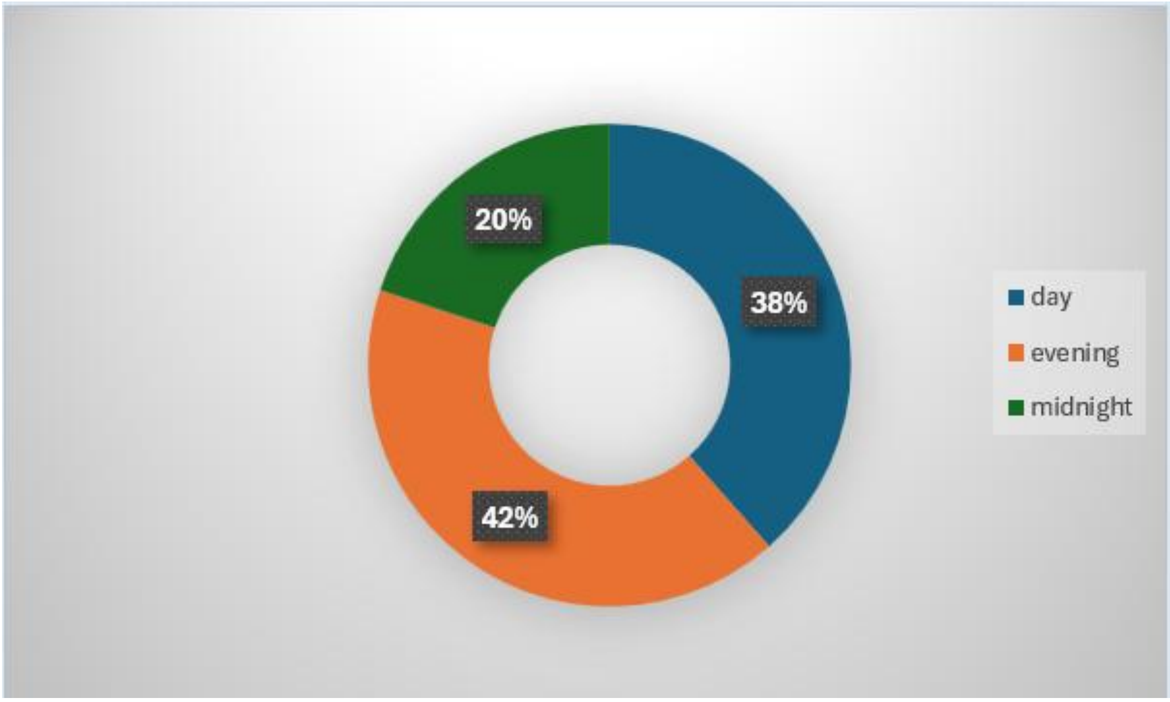
- **Offense Type:** Certain offenses might be more prevalent during specific shifts (e.g., burglaries at night).
- **District:** Crime patterns might vary across different districts, with some districts experiencing more crimes during specific shifts.
- **Month:** Seasonal variations in crime might influence the shift-wise distribution (e.g., more outdoor crimes during warmer months).

iv. Visualization

The visualization consists of two primary components:

- **Pivot Table:** A table showing the "Sum of CCN" for each shift, providing the numerical data for analysis.
- **Pie Chart:** A pie chart that visually represents the percentage distribution of crimes across shifts, making it easy to compare the relative proportions of crimes during different shifts.

Additionally, the image includes filters related to offense type, district, and month. These filters allow for a more granular analysis of the shift-wise crime distribution, enabling users to explore how specific factors influence crime patterns across different shifts.



Objective 4: Top 5 Districts

i. General Description

This objective focuses on identifying the top 5 districts with the highest crime rates, as measured by the "Sum of CCN" (likely "Complaint Control Number" or a similar unique identifier). It aims to pinpoint the geographical areas with the most reported crime incidents. The analysis is presented using a pivot table showing the "Sum of CCN" for each of the top 5 districts and a horizontal bar chart visualizing the same data.

ii. Specific Requirements

The specific requirements for this analysis include:

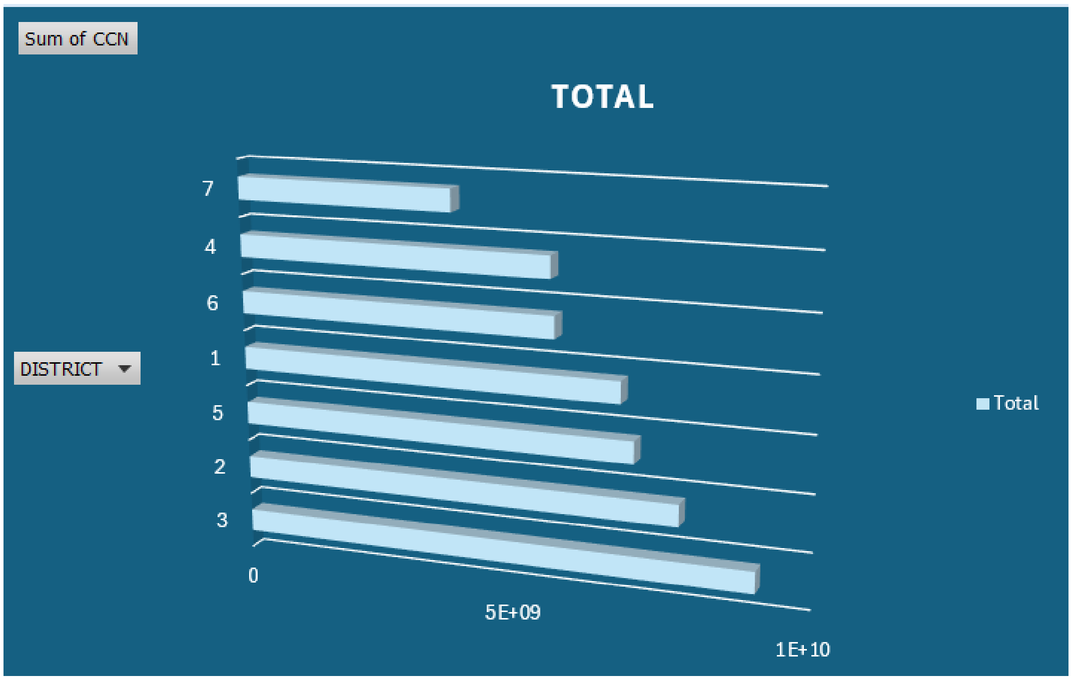
- **Data Aggregation:** The data is aggregated by district, summing the "CCN" values for each district.
- **Selection of Top 5:** The analysis specifically focuses on identifying and presenting the top 5 districts with the highest "Sum of CCN."
- **Visualization:** The results are presented both in a pivot table and a horizontal bar chart.
- **Focus on Numerical Values:** The analysis should consider the absolute numerical values of "Sum of CCN" for each of the top 5 districts.
- **Geographical Identification:** The analysis should identify the districts by their numerical labels (1 through 5 in the image).

iii. Analysis Results

***Tabular Data Analysis*:**

The pivot table shows the "Sum of CCN" for the top 5 districts:

- **District 1:** 9163341497
- **District 2:** 7867280855
- **District 3:** 7088106871
- **District 4:** 6854067043
- **District 5:** 5711883600
- **District 6:** 5641344293
- **District 7:** 3904771630
- **Grand Total:** 46230870794



Bar Chart Analysis:

The horizontal bar chart visually represents the "Sum of CCN" for each of the top 5 districts.

- **Relative Length of Bars:** The length of each bar corresponds to the "Sum of CCN" for that district, allowing for a quick visual comparison of crime rates across the top 5 districts.
- **Ordering:** The districts are ordered from highest to lowest "Sum of CCN," making it easy to identify the district with the highest crime rate.

Interpretation:

- **District 1 Dominance:** District 1 has the highest "Sum of CCN" (9163341497), indicating it is the district with the highest crime rate among the top 5.
- **Decreasing Trend:** The "Sum of CCN" decreases as you move from District 1 to District 5, suggesting a decreasing trend in crime rates among the top districts.

- **Numerical Labels:** The districts are identified by numerical labels (1 through 5), indicating that the actual geographical names of the districts are not provided in this analysis.

iv. Visualization

The visualization consists of two components:

- **Pivot Table:** A table showing the "Sum of CCN" for each of the top 5 districts, providing the numerical data for analysis.

Row Labels	Sum of CCN
3	9163341497
2	7867283855
5	7088186871
1	6854067043
6	5711883605
4	5641344293
7	3904771630
Grand Total	**46230878794**

- **Horizontal Bar Chart:** A bar chart that visually represents the "Sum of CCN" for each of the top 5 districts, allowing for a quick and intuitive comparison of crime rates across the districts.

The bar chart effectively visualizes the relative magnitudes of crime rates across the top 5 districts, making it easy to identify the district with the highest crime rate and compare the crime rates of the other districts.

Objective 5: Neighbourhood-wise Analysis

i. General Description

This objective focuses on analysing crime incidents at a neighbourhood level, likely represented by "Clusters" in the dataset. It aims to understand the distribution of crime occurrences across different neighbourhoods, providing insights into which areas have higher or lower crime rates. The analysis is presented using a pivot table showing the "Sum of CCN" (likely "Complaint Control Number" or a similar unique identifier) for each cluster and a column chart visualizing the same data.

ii. Specific Requirements

The specific requirements for this analysis include:

- **Data Aggregation:** The data is aggregated by neighbourhood "Cluster", summing the "CCN" values for each cluster.
- **Visualization:** The results are presented both in a pivot table and a column chart.

- **Focus on Numerical Values:** The analysis should consider the absolute numerical values of "Sum of CCN" for each cluster.
- **Identification of Trends:** The analysis should identify clusters with significantly higher or lower crime rates.
- **Neighbourhood Identification:** The analysis should identify neighbourhoods by their cluster labels (Cluster 1, Cluster 2, etc.).

iii. Analysis Results

Tabular Data Analysis:

The pivot table shows the "Sum of CCN" for each cluster. The values vary significantly across the clusters, suggesting differences in crime rates across neighbourhoods.

- **Cluster 1:** 12770000650
- **Cluster 2:** 3280592320
- **Cluster 3:** 2694985392
- **Cluster 17:** 1205063796
- **Cluster 18:** 2071246878
- **Cluster 25:** 3575289414
- **Cluster 26:** 1689209934
- **Cluster 27:** 1061324025
- **Cluster 28:** 4099776997
- **Cluster 29:** 72319249
- **Cluster 30:** 6985567754
- **Cluster 31:** 8912216884
- **Other Clusters:** (Values for other clusters are also present but not listed here for brevity)

Column Chart Analysis:

The column chart visually represents the "Sum of CCN" for each cluster.

- **Height of Columns:** The height of each column corresponds to the "Sum of CCN" for that cluster, allowing for a quick visual comparison of crime rates across neighbourhoods.
- **Variation in Heights:** The chart shows significant variation in the heights of the columns, indicating differences in crime rates across clusters.
- **Identification of Outliers:** The chart makes it easy to identify clusters with exceptionally high or low crime rates.

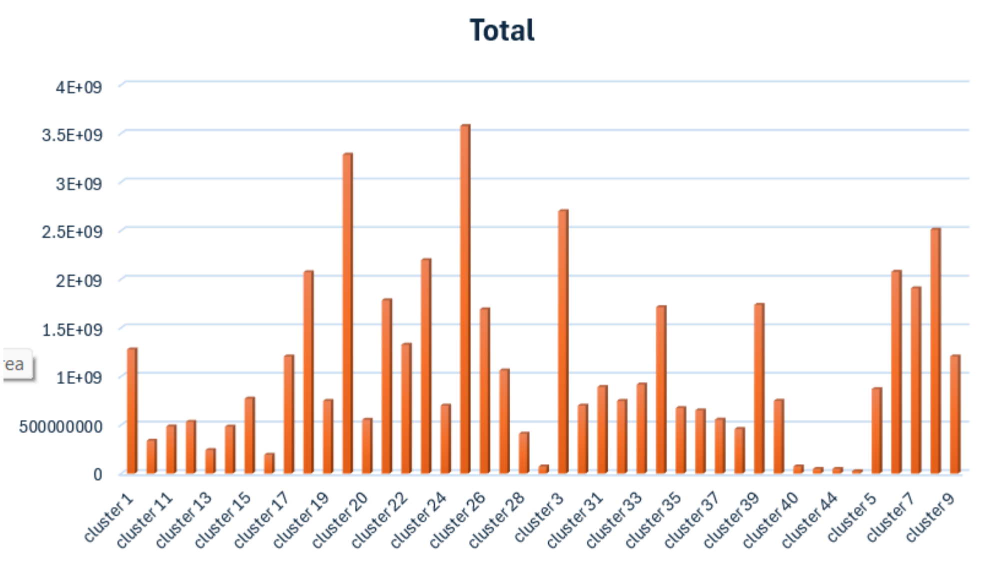
Interpretation:

- **Cluster 1 High Crime:** Cluster 1 has the highest "Sum of CCN" (12770000650), indicating it is the neighbourhood with the highest crime rate.
- **Variable Crime Rates:** The column chart and pivot table show that crime rates vary significantly across different clusters, suggesting that some neighbourhoods are more prone to crime than others.
- **Cluster Labels:** The neighbourhoods are identified by cluster labels (Cluster 1, Cluster 2, etc.), indicating that the actual geographical names of the neighbourhoods are not provided in this analysis.

iv. Visualization

The visualization consists of two components:

- **Pivot Table:** A table showing the "Sum of CCN" for each cluster, providing the numerical data for analysis.
- **Column Chart:** A column chart that visually represents the "Sum of CCN" for each cluster, allowing for a quick and intuitive comparison of crime rates across neighbourhoods.



The column chart effectively visualizes the relative magnitudes of crime rates across different clusters, making it easy to identify neighbourhoods with high or low crime rates. The combination of the table and the chart provides a comprehensive view of the neighbourhood-wise crime distribution.

Objective 6: Ward-wise Crime

i. General Description

This objective focuses on analysing crime incidents at a ward level, showing the distribution of various offense types across different wards. It aims to understand the specific types of crimes prevalent in each ward and identify patterns or trends. The analysis is presented using a pivot table showing the "Count of OFFENSE" for different offense categories across wards and a 3D surface chart visualizing the same data.

ii. Specific Requirements

The specific requirements for this analysis include:

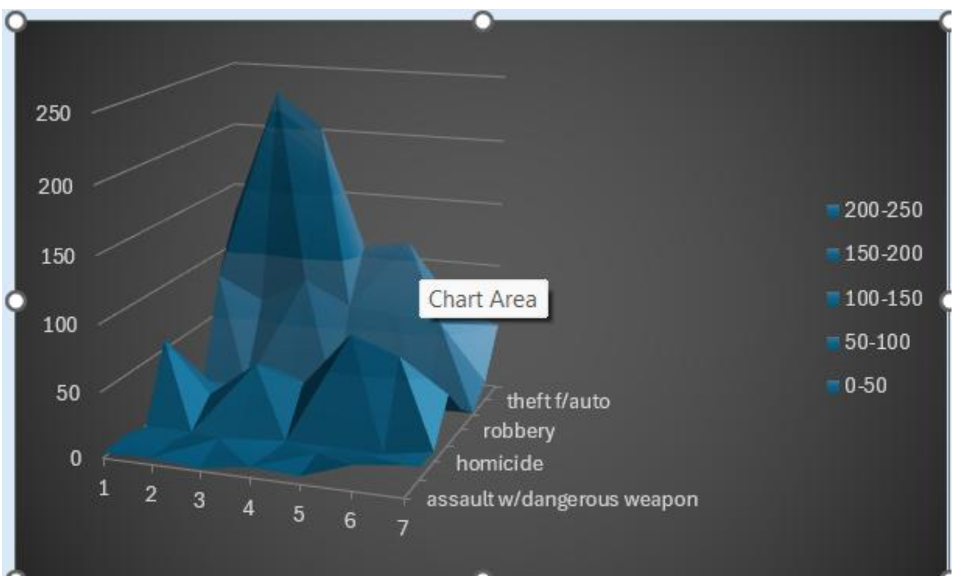
- **Data Aggregation:** The data is aggregated by ward, counting the occurrences of each offense type in each ward.
- **Visualization:** The results are presented both in a pivot table and a 3D surface chart.
- **Focus on Offense Types:** The analysis should consider the distribution of various offense types across wards.
- **Identification of Trends:** The analysis should identify wards with high occurrences of specific offense types.
- **Ward Identification:** The analysis should identify wards by their numerical labels (1 through 7 in the image).

iii. Analysis Results

***Tabular Data Analysis*:**

The pivot table shows the "Count of OFFENSE" for different offense types across wards:

- **Offense Types:** assault w/dangerous weapon, burglary, homicide, motor vehicle theft, robbery, sex abuse, theft f/auto, theft/other
- **Wards:** 1 to 7



Observations:

- **Ward 3 High Crime:** Ward 3 has the highest total count of offenses (380).
- **Theft/Other Dominance:** The "theft/other" category has the highest overall count (895) and is prevalent across all wards.
- **Robbery Variation:** The "robbery" count varies significantly across wards, with Ward 3 having the highest count (46).
- **Homicide Low:** The "homicide" count is relatively low across all wards.
- **Assault Variation:** The "assault w/dangerous weapon" count also varies across wards, with Ward 3 having the highest count (24).

3D Surface Chart Analysis:

The 3D surface chart visually represents the distribution of offenses across wards.

- **Axes:** The chart has three axes: ward number, offense type, and count of offenses.
- **Peaks and Valleys:** The peaks in the chart represent high counts of offenses, while the valleys represent low counts.
- **Overall Pattern:** The chart shows the overall pattern of offense distribution across wards, highlighting the wards and offense types with high or low counts.

Interpretation:

- **Ward 3 High Crime Zone:** Ward 3 stands out as a high-crime zone with high counts across multiple offense types.
- **Theft/Other Common:** The "theft/other" category is the most common offense type across all wards.
- **Offense Type Variation:** The chart and table show that the distribution of different offense types varies across wards, suggesting that different areas have different crime profiles.

Count of OFFENSE	Column Labels								
Row Labels	assault w/dangerous weapon	burglary	homicide	motor vehicle theft	robbery	sex abuse	theft f/auto	theft/other	Grand Total
1	4	5		58	7	2	81	127	284
2	5	4		18	3	3	64	229	326
3	2	12		46	32	1	84	203	380
4	8	4	2	34	12	1	66	107	234
5	6	10	2	77	22	1	63	113	294
6	19	13	1	64	16		59	65	237
7	24	9	8	39	16	1	15	51	163

iv. Visualization

The visualization consists of two components:

- **Pivot Table:** A table showing the "Count of OFFENSE" for different offense types across wards, providing the numerical data for analysis.
- **3D Surface Chart:** A 3D surface chart that visually represents the distribution of offenses across wards, highlighting the wards and offense types with high or low counts.

The 3D surface chart effectively visualizes the complex relationship between wards and offense types, making it easy to identify patterns and trends. The combination of the table and the chart provides a comprehensive view of the ward-wise crime distribution.

Objective 7: Ward-wise Location Spread

i. General Description

This objective focuses on analysing the geographical spread of crime incidents across different wards, likely using latitude and longitude data. It aims to understand the spatial distribution of crimes and identify any patterns or concentrations within each ward. The analysis is presented using a pivot table showing the "Sum of LATITUDE" and "Sum of LONGITUDE" for each ward and a tree map visualizing the same data.

ii. Specific Requirements

The specific requirements for this analysis include:

- **Data Aggregation:** The data is aggregated by ward, summing the latitude and longitude values for each ward.
- **Visualization:** The results are presented both in a pivot table and a tree map.
- **Focus on Spatial Distribution:** The analysis should consider the spatial distribution of crime incidents across wards.
- **Identification of Patterns:** The analysis should identify any patterns or concentrations of crime incidents within each ward.
- **Ward Identification:** The analysis should identify wards by their numerical labels (-7 to 8 in the image).

iii. Analysis Results

Tabular Data Analysis:

The pivot table shows the "Sum of LATITUDE" and "Sum of LONGITUDE" for each ward (-7 to 8).

- **Latitude and Longitude Sums:** The table shows the aggregated latitude and longitude values for each ward. These values are likely used to represent the central point or centroid of crime incidents within each ward.

Tree map Analysis:

The tree map visually represents the spatial distribution of crime incidents across wards.

- **Size of Rectangles:** The size of each rectangle in the tree map represents the relative number of crime incidents in each ward. Larger rectangles indicate wards with more crime incidents.

- **Colour Coding:** The colour coding in the tree map likely represents a specific attribute, such as the number of incidents or the density of incidents.
- **Ward Labels:** The tree map includes labels for each ward (-7 to 8).

Interpretation:

- **Ward Variation:** The tree map shows significant variation in the size of rectangles, indicating differences in the number of crime incidents across wards.
- **Spatial Concentration:** The tree map suggests that crime incidents are not evenly distributed across wards, with some wards having higher concentrations of crime incidents.
- **Centroid Data:** The pivot table provides the aggregated latitude and longitude data, which can be used to calculate the centroid of crime incidents within each ward. This data can be used to further analyze the spatial distribution of crimes

  .

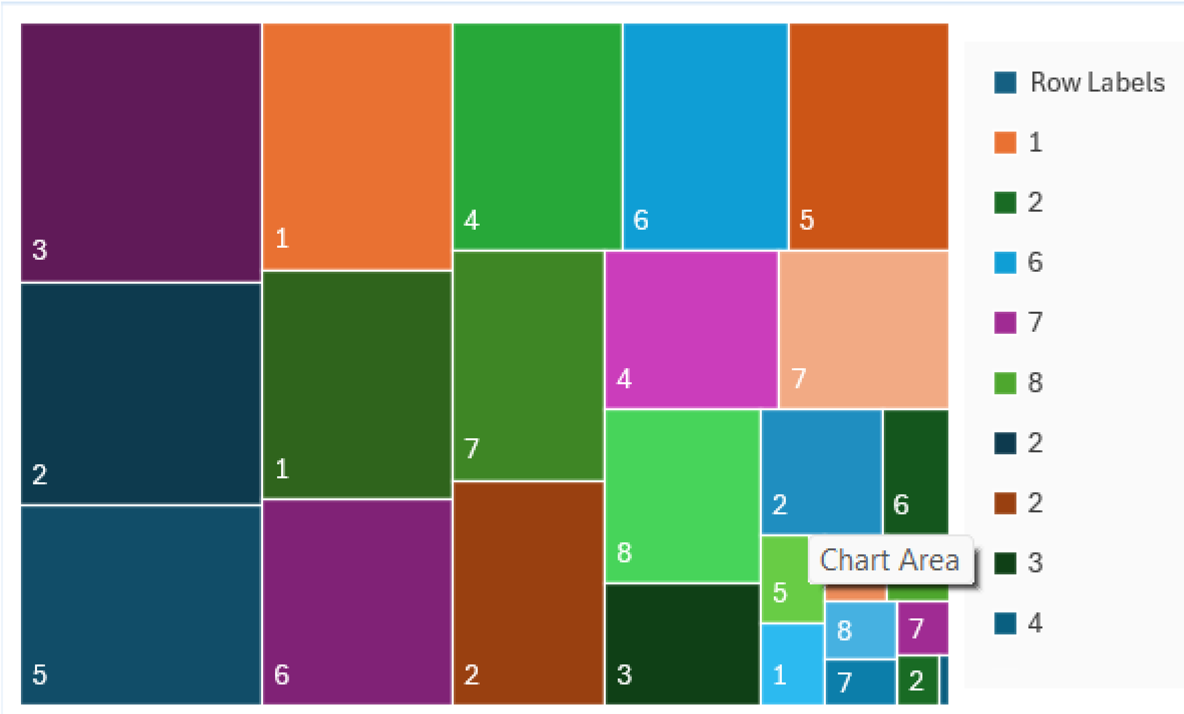
iv. Visualization

The visualization consists of two components:

- **Pivot Table:** A table showing the "Sum of LATITUDE" and "Sum of LONGITUDE" for each ward, providing the numerical data for analysis.

Row Labels	Sum of LATITUDE	Sum of LONGITUDE
⊟1	**11044.25303**	**-21868.68015**
2	505.6995073	-1001.284084
6	8905.546616	-17633.6706
7	661.0947699	-1308.662462
8	971.9121368	-1925.063
⊟2	**12687.07504**	**-25119.19608**
2	8092.157147	-16024.87183
3	4478.02138	-8863.13651
4	116.8965135	-231.1877381
⊟3	**14789.13583**	**-29270.96157**
1	10236.57798	-20259.23462
2	3579.730307	-7086.420731
5	972.8275403	-1925.306216
⊟4	**9114.579753**	**-18022.51142**
1	1245.842037	-2464.817774
4	6505.705311	-12862.79094
5	1363.032405	-2694.902701
⊟5	**11440.46178**	**-22634.38073**
5	8678.386453	-17167.66463
6	1984.136491	-3927.185596
7	777.9388343	-1539.530509
⊟6	**9215.146488**	**-18237.056**
7	8243.436797	-16312.63828
8	971.7096917	-1924.417728
⊟7	**6331.805388**	**-12549.46515**
8	6331.805388	-12549.46515
Grand Total	**74622.45731**	**-147702.2511**

- **Tree map:** A tree map that visually represents the spatial distribution of crime incidents across wards, highlighting the wards with higher or lower concentrations of crime incidents.



The tree map effectively visualizes the relative number of crime incidents across wards, making it easy to identify wards with high or low crime rates. The combination of the table and the tree map provides a comprehensive view of the ward-wise location spread.

Objective 8: X & Y Sum by Ward

i. General Description

This objective focuses on analysing the sum of X and Y coordinates (likely representing spatial data, similar to latitude and longitude) for crime incidents across different wards. It aims to understand the spatial distribution of crimes based on these summed coordinates and identify any patterns or concentrations within each ward. The analysis is presented using a pivot table showing the "Sum of X" and "Sum of Y" for each ward and a stacked horizontal bar chart visualizing the same data.

ii. Specific Requirements

The specific requirements for this analysis include:

- **Data Aggregation:** The data is aggregated by ward, summing the X and Y coordinate values for each ward.
- **Visualization:** The results are presented both in a pivot table and a stacked horizontal bar chart.

- **Focus on Spatial Coordinates:** The analysis should consider the spatial distribution of crime incidents based on the summed X and Y coordinates.
- **Identification of Patterns:** The analysis should identify any patterns or concentrations of crime incidents within each ward based on the summed coordinates.
- **Ward Identification:** The analysis should identify wards by their numerical labels (-8 to 8 in the image).
-

iii. Analysis Results

Tabular Data Analysis:

The pivot table shows the "Sum of X" and "Sum of Y" for each ward (-8 to 8).

- **Sum of X and Sum of Y:** The table shows the aggregated X and Y coordinate values for each ward. These values are likely used to represent the central point or centroid of crime incidents within each ward, similar to latitude and longitude in the previous objective.

Row Labels	Sum of X	Sum of Y
⊟1	**117859318.3**	**40560739.1**
3	105077725.8	36146296.76
4	12781592.54	4414442.346
⊟2	**125123679.1**	**43129563.58**
1	5188064.042	1784787.95
2	83101492.91	28687967.87
3	36834122.1	12656807.76
⊟3	**45925414.24**	**15808308.65**
2	45925414.24	15808308.65
⊟4	**67931989.77**	**23416702.5**
2	1193933.615	403928.1
4	66738056.16	23012774.4
⊟5	**113031593.8**	**38814481.93**
3	9962565.956	3450297.534
4	13990977.44	4806173.994
5	89078050.42	30558010.4
⊟6	**111898657.2**	**38466803.05**
1	91501430.86	31463741.21
5	20397226.34	7003061.844
⊟7	**99566807.27**	**34209759.92**
1	6810279.973	2333043.04
5	8000629.945	2736424.214
6	84755897.35	29140292.67
⊟8	**85140046.89**	**29362268.32**
1	9964649.36	3467249.662
6	10007429.98	3452099.873
7	65167967.55	22442918.78
Grand Total	**766477506.6**	**263768627.1**

Stacked Horizontal Bar Chart Analysis:

The stacked horizontal bar chart visually represents the "Sum of X" and "Sum of Y" for each ward.

- **Stacked Bars:** Each bar represents a ward, with the "Sum of X" and "Sum of Y" values stacked on top of each other.
- **Length of Bars:** The total length of each bar represents the combined magnitude of "Sum of X" and "Sum of Y" for that ward.
- **Colour Coding:** The colour coding distinguishes between "Sum of X" and "Sum of Y" values.
- **Comparison:** The chart allows for a quick visual comparison of the relative magnitudes of "Sum of X" and "Sum of Y" across different wards.

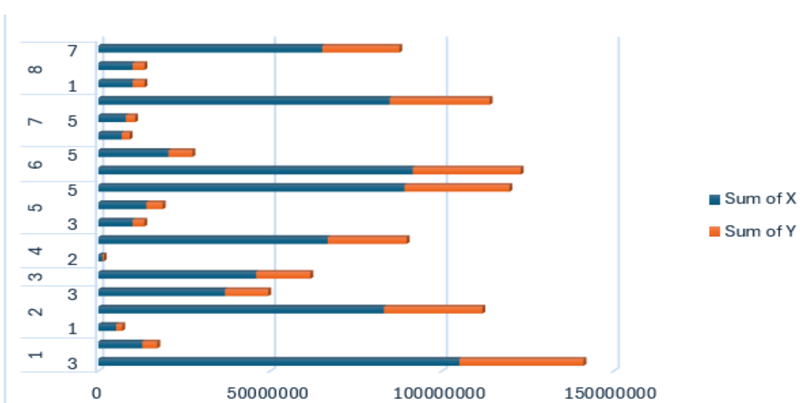
Interpretation:

- **Ward Variation:** The stacked horizontal bar chart shows significant variation in the length of bars, indicating differences in the combined magnitude of "Sum of X" and "Sum of Y" across wards.
- **Coordinate Distribution:** The chart shows the relative distribution of "Sum of X" and "Sum of Y" within each ward, suggesting differences in the spatial orientation of crime incidents across wards.
- **Centroid Data:** The pivot table provides the aggregated X and Y coordinate data, which can be used to calculate the centroid of crime incidents within each ward. This data can be used to further analyze the spatial distribution of crimes.

iv. Visualization

The visualization consists of two components:

- **Pivot Table:** A table showing the "Sum of X" and "Sum of Y" for each ward, providing the numerical data for analysis.
- **Stacked Horizontal Bar Chart:** A stacked horizontal bar chart that visually represents the "Sum of X" and "Sum of Y" for each ward, highlighting the differences in the spatial distribution of crime incidents across wards.



The stacked horizontal bar chart effectively visualizes the relative magnitudes and distributions of "Sum of X" and "Sum of Y" across wards, making it easy to identify wards with high or low spatial coordinates. The combination of the table and the chart provides a comprehensive view of the X and Y sum by ward.

Objective 9: ANC Count by Ward & Group

i. General Description

This objective focuses on analysing the "ANC Count" (likely representing "Advisory Neighbourhood Commission" or a similar grouping) across different wards and groups. It aims to understand the distribution of ANC counts across these categories and identify any patterns or concentrations. The analysis is presented using a pivot table showing the "Count of ANC" for different groups across wards and a 3D surface chart visualizing the same data.

ii. Specific Requirements

The specific requirements for this analysis include:

- **Data Aggregation:** The data is aggregated by ward and group, counting the occurrences of ANC values.
- **Visualization:** The results are presented both in a pivot table and a 3D surface chart.
- **Focus on ANC Counts:** The analysis should consider the distribution of ANC counts across wards and groups.
- **Identification of Patterns:** The analysis should identify wards and groups with high or low ANC counts.
- **Ward and Group Identification:** The analysis should identify wards and groups by their numerical labels (1 to 8 in the image).

iii. Analysis Results

Tabular Data Analysis:

The pivot table shows the "Count of ANC" for different groups across wards:

- **Groups:** 1 to 8 (likely representing different ANC groups)
- **Wards:** 1 to 7

Observations:

- **Ward 3 High ANC Count:** Ward 3 has the highest total ANC count (380).
- **Group 2 High ANC Count:** Group 2 has the highest total ANC count (313).

- **Variation Across Wards:** ANC counts vary significantly across wards, suggesting differences in the distribution of ANC groups across these areas.
- **Variation Across Groups:** ANC counts also vary across groups, suggesting differences in the prevalence of these groups within the dataset.

3D Surface Chart Analysis:

The 3D surface chart visually represents the distribution of ANC counts across wards and groups.

- **Axes:** The chart has three axes: ward number, group number, and count of ANC values.
- **Peaks and Valleys:** The peaks in the chart represent high ANC counts, while the valleys represent low ANC counts.
- **Overall Pattern:** The chart shows the overall pattern of ANC count distribution across wards and groups, highlighting the wards and groups with high or low counts.

Interpretation:

- **Ward 3 Dominance:** Ward 3 stands out as a high ANC count zone with high counts across multiple groups.
- **Group 2 Prevalence:** Group 2 has a high prevalence across different wards.
- **ANC Distribution Differences:** The chart and table show that the distribution of ANC counts varies across wards and groups, suggesting that different areas have different ANC profiles.

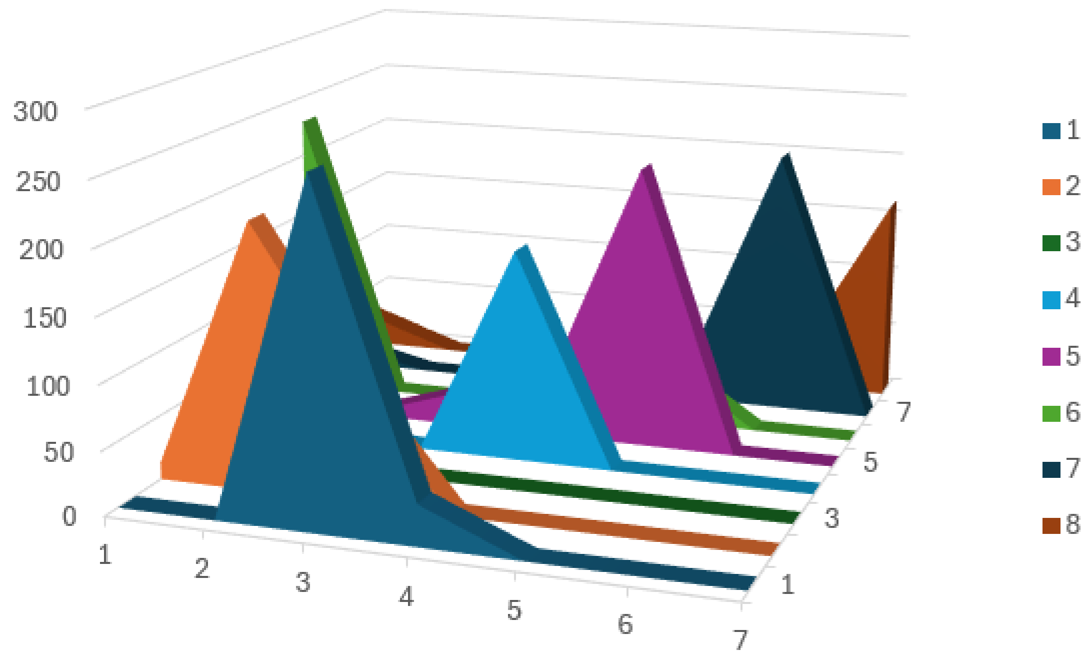
iv. Visualization

The visualization consists of two components:

- **Pivot Table:** A table showing the "Count of ANC" for different groups across wards, providing the numerical data for analysis.

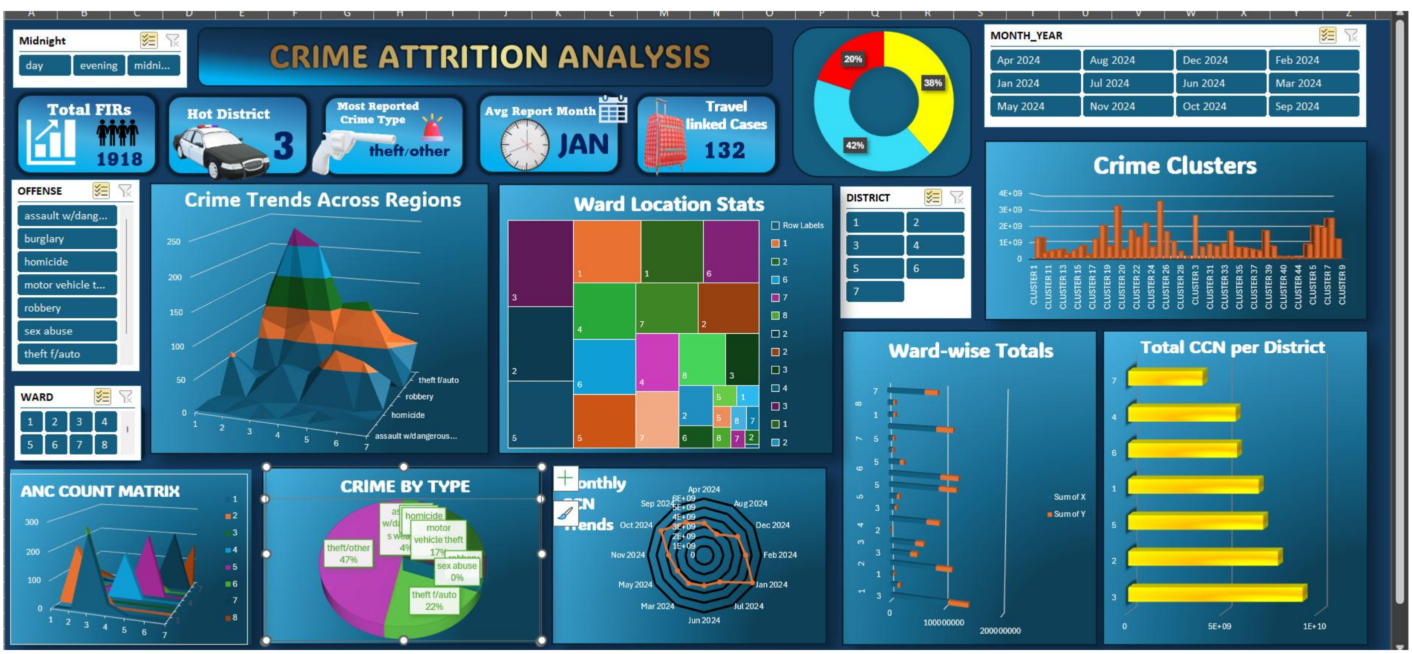
Count of ANC	Column Labels								
Row Labels	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**Grand Total**
1		13				229	17	25	284
2		208	115	3					326
3	263	92			25				380
4	32			167	35				234
5					223	51	20		294
6							212	25	237
7								163	163
Grand Total	**295**	**313**	**115**	**170**	**283**	**280**	**249**	**213**	**1918**

- **3D Surface Chart:** A 3D surface chart that visually represents the distribution of ANC counts across wards and groups, highlighting the wards and groups with high or low counts.



The 3D surface chart effectively visualizes the complex relationship between wards and groups, making it easy to identify patterns and trends. The combination of the table and the chart provides a comprehensive view of the ANC count by ward and group.

5.Conclusion



The 2024 Crime Data Dashboard presents a thorough and insightful exploration of crime trends within the District of Columbia. Through in-depth data analysis and visual representation, this dashboard not only identifies key areas of concern but also serves as a foundational tool for informed decision-making in public safety and policy formulation. The findings derived from the analysis are vital for enabling law enforcement agencies, policymakers, and community leaders to develop targeted interventions that address the root causes of criminal activity and enhance community well-being.

5.1. Geographical Hotspots: Focus on District 3

One of the most striking revelations from the data is the persistently high crime concentration in District 3. This area has consistently reported elevated numbers of incidents, making it a clear hotspot that requires urgent and focused attention. The prevalence of crime in this district suggests systemic issues—possibly including socioeconomic disparities, inadequate infrastructure, or limited community policing presence—that need to be addressed. Targeted interventions such as increased patrolling, community policing initiatives, installation of surveillance systems, and neighbourhood revitalization efforts can help mitigate the criminal activity in this high-risk area. Moreover, a spatial analysis of surrounding zones suggests potential spillover effects, indicating that crime prevention in District 3 could also positively impact neighbouring regions.

5.2. Crime Type Distribution: Addressing the Rise of Opportunistic Offenses

The analysis highlights "Theft/Other" as the most dominant crime type across the city. This category often includes crimes of opportunity such as thefts from vehicles, package thefts, and unattended property being stolen. The frequency of such incidents indicates not only a lapse in preventive infrastructure but also a behavioural trend among offenders taking advantage of low-risk, high-reward scenarios. To address this, public awareness campaigns focused on personal property security, coupled with strategic placement of law enforcement in high-theft areas, could prove effective. Additionally, promoting neighbourhood watch programs and encouraging the use of technology such as smart doorbells and security cameras may further deter opportunistic criminal behaviour.

5.3. Temporal Trends: Vulnerability During Specific Periods

Temporal analysis uncovers notable patterns in crime occurrence, with January emerging as a month of heightened criminal activity. This spike may be influenced by post-holiday economic strain, reduced daylight hours, or seasonal changes in community behaviour. Similarly, crime concentration during evening hours suggests a vulnerability window when surveillance and law enforcement visibility should be heightened. Law enforcement agencies can use this insight to realign patrol schedules, increase street lighting, and encourage businesses and public areas to remain vigilant during these peak times. Additionally, community-based initiatives such as evening safety walks or patrols could enhance the sense of security during vulnerable hours

5.4. Community Involvement: Leveraging ANC Engagement

The analysis also underscores the significance of Advisory Neighbourhood Commissions (ANCs) and their role in community engagement. Certain wards with high ANC activity show a greater correlation with reported crime incidents, indicating both heightened community awareness and possibly stronger mechanisms for reporting and recording crimes. This presents an opportunity: ANCs can act as key partners in disseminating crime prevention information, facilitating community policing efforts, and fostering trust between residents and law enforcement. Strengthening the relationship between ANCs and police departments can lead to collaborative initiatives such as town halls, safety audits, and youth engagement programs aimed at crime reduction.

5.5. Emphasizing Data-Driven Policing Strategies

Perhaps the most important takeaway from this dashboard is the power and necessity of data-driven policing. The ability to visualize and interpret crime trends over time, space, and category provides a significant advantage in formulating strategic responses. Traditional blanket approaches to crime reduction often waste valuable resources; by contrast, targeted strategies based on data ensure maximum impact with optimal resource deployment. This approach also promotes transparency and accountability in policing, as decisions are backed by tangible evidence rather than assumptions. Integrating this dashboard into routine police operations, performance tracking, and crime forecasting models can substantially enhance the overall effectiveness of public safety efforts.

6. Future Scope

While the present dashboard provides a robust analysis of crime patterns in Washington, D.C. for the year 2024, there remains significant scope for further development and research. Future enhancements can increase the analytical power of the dashboard and expand its utility for law enforcement, policymakers, and the public.

1. Real-Time Data Integration

Currently, the analysis is based on static 2024 data. Future versions of the dashboard can integrate real-time data feeds directly from the Metropolitan Police Department's databases or APIs. This will allow for dynamic visualization of ongoing incidents, enabling real-time situational awareness and quicker response strategies.

2. Predictive Crime Modelling

Advanced statistical models and machine learning algorithms can be incorporated to forecast potential crime hotspots and high-risk time windows. Predictive policing techniques using historical trends, socio-economic data, and seasonal effects can greatly enhance crime prevention efforts.

3. Inclusion of Demographic and Socioeconomic Layers

Future studies can incorporate demographic variables such as income level, education, population density, and unemployment rates across wards. This would allow a deeper understanding of the underlying causes of crime and help design targeted community interventions.

4. Mobile-Friendly Dashboard and Public Access Tools

To enhance accessibility, the dashboard can be converted into a mobile-responsive format or app, allowing public users and law enforcement officers in the field to interact with the data on the go. Features like location-based alerts or citizen reporting modules can be considered.

5. Enhanced Geospatial Analysis

Future iterations may use GIS tools for more advanced spatial clustering, heatmaps, and crime corridor detection. Integration with satellite imagery and 3D mapping can also aid in understanding the urban environment’s influence on criminal activity.

6. Comparative and Longitudinal Studies

By collecting data across multiple years, analysts can conduct trend analysis to understand whether crime is increasing, decreasing, or shifting geographically. Comparative analysis with other cities or regions can also help in identifying successful crime prevention models.

7. Community Feedback and Participatory Analysis

Incorporating qualitative inputs from local communities through surveys or interactive comment systems can enhance the interpretation of the data. Understanding residents' perceptions can offer ground-level insights that are often missed in quantitative analysis alone.

In summary, the future scope emphasizes the integration of technology, interdisciplinary data, and public engagement to create a more holistic and proactive crime analytics system.

7. References

- District of Columbia Metropolitan Police Department, "Crime Incidents in 2024," *Data.gov*, [Online]. Available: https://catalog.data.gov/dataset/crime-incidents-in-2024. [Accessed: Apr. 10, 2025].
- Microsoft Corporation, *Microsoft Excel*, Version 2302, Redmond, WA, USA: Microsoft, 2024. [Software].
- U.S. Census Bureau, "QuickFacts: District of Columbia," *United States Census Bureau*, [Online]. Available: https://www.census.gov/quickfacts/fact/table/districtofcolumbia. [Accessed: Apr. 10, 2025].
- S. S. Sridhar and A. K. Sinha, "Crime Data Analysis Using Visualization Tools," in *Proc. 2022 Int. Conf. Data Science and Applications (ICDSA)*, Gwalior, India, Dec. 2022, pp. 45–50. doi: 10.1109/ICDSA55671.2022.00014.

8.Other Activities

- Google Drive Link: https://docs.google.com/spreadsheets/d/1vHWuJSh0i-DrE0nKv7IHrN1nfdz_YEbv/edit?usp=drive_link&ouid=107031321188897402278&rtpof=true&sd=true

- LinkedIn Post Link: https://www.linkedin.com/posts/suhanirawat2305_dataanalytics-exceldashboard-crimedata-activity-7316839374514495489-nZ_w?utm_source=social_share_send&utm_medium=member_desktop_web&rcm=ACoAAEzCvT0B0uXaw4pneCrR0JkXZLpjPiqTONk