**<Data Science Toolbox: Python Programming >**

**PROJECT REPORT**

(Project Semester January-May2025)

***For the partial fulfilment of the BTech Computer Science Engineering***

***(Crime Incidents in 2024***

***District of Columbia)***

Submitted by

Priyam Tiwari

Registration No: 12315493

Programme and Section BTech CSE

Course Code: INT375

Under the Guidance of

**(Mrs. Aashima**

**Assistant Professor**

**SCSE (Data Science)**UID-28968

**Discipline of CSE/IT**

**Lovely School of Computer Science and Engineering**

**Lovely Professional University, Phagwara**

**CERTIFICATE**

This is to certify that Priyam Tiwari bearing Registration no. 12315493 has completed INT 375 project titled, **“*Crime incidents 2024*”** 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 and Engineering**

Lovely Professional University

Phagwara, Punjab.

Date: 9/4/2025

**DECLARATION**

I, Priyam Tiwari, student of BTech Computer Science Engineering 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: Signature

Registration No:12315493 Priyam Tiwari

**Table of Contents**

1. **Introduction**  
   1.1 Background  
   1.2 Objectives  
   1.3 Importance of Landslide Data Analysis
2. **Source of Dataset**  
   2.1 Origin of the Dataset  
   2.2 Dataset Description  
   2.3 Key Features and Variables
3. **Exploratory Data Analysis (EDA) Process**  
   3.1 Data Loading and Inspection  
   3.2 Handling Missing Values  
   3.3 Date Processing  
   3.4 Outlier Detection
4. **Detailed Analysis on Dataset**  
   4.1 Top 10 Countries by Landslide Frequency  
   a. Introduction  
   b. Description  
   c. Methodology and Functions  
   d. Results  
   e. Visualization  
   4.2 Fatalities vs Injuries  
   a. Introduction  
   b. Description  
   c. Methodology and Functions  
   d. Results  
   e. Visualization  
   4.3 Rainfall vs Other Triggers  
   a. Introduction  
   b. Description  
   c. Methodology and Functions  
   d. Results  
   e. Visualization

4.4 Quarterly Landslide Trends  
a. Introduction  
b. Description  
c. Methodology and Functions  
d. Results  
e. Visualization

4.5 Correlation Heatmap  
a. Introduction  
b. Description  
c. Methodology and Functions  
d. Results  
e. Visualization

4.6 Pair Plot of Numerical Variables  
a. Introduction  
b. Description  
c. Methodology and Functions  
d. Results  
e. Visualization

1. **Conclusion**  
   5.1 Summary of Findings  
   5.2 Key Insights
2. **Future Scope**  
   6.1 Enhancing Predictive Analysis  
   6.2 Integration with GIS Tools  
   6.3 Real-time Monitoring Applications
3. **References**  
   7.1 Dataset Source  
   7.2 Libraries and Tools Used

**CRIME INCIDENTS IN 2024 - ANALYSIS REPORT**

**1. Introduction:** This report provides a comprehensive analysis of crime incidents recorded throughout 2024 in the District of Columbia. The primary objective is to uncover patterns, trends, and insights using Exploratory Data Analysis (EDA) techniques. These findings are intended to support strategic crime prevention, resource allocation, and effective decision-making by law enforcement agencies, city planners, and public policy professionals.

Through this analysis, we aim to identify not just the most common crime types, but also when and where they occur most frequently. Understanding these patterns can lead to proactive measures that help improve community safety. The report includes detailed breakdowns by offense type, time of day and week, geographical distribution, and law enforcement response, supported by insightful visualizations and statistical summaries.

Crime analytics plays a critical role in modern policing. By applying data science techniques, we gain deeper insight into criminal behaviour, anticipate future occurrences, and develop community-specific solutions. This report leverages publicly available datasets and modern analytical tools to deliver a structured and informative review of crime across the district.

**2. Source of Dataset:** The dataset used for this analysis is sourced from the DC Open Data portal (<https://opendata.dc.gov/>) which provides access to a wide array of public datasets shared by the District of Columbia government. Specifically, the dataset comprises officially documented crime incidents collected from the Metropolitan Police Department.

The data includes:

* Crime type and description
* Geographic details (neighbourhood, ward, police district, and GPS coordinates)
* Date and time of incident
* Method and means of offense (e.g., weapon used)
* Case status (open, closed, pending)
* Response and clearance details

The dataset contains hundreds of thousands of records and is updated daily, allowing real-time tracking of criminal activity in Washington, D.C. This level of transparency and granularity empowers researchers, journalists, and policymakers to monitor crime trends with greater confidence.

Before analysis, the dataset underwent a validation process through data profiling techniques to ensure accuracy, consistency, and completeness. Missing values were handled, and formats were standardized to align with analytical requirements. Columns were renamed where necessary to ensure consistency, and any outlier entries were flagged for review.

**3. EDA Process:** the Exploratory Data Analysis (EDA) process involved a series of systematic steps to extract meaningful patterns and prepare the data for detailed investigation. The major steps included

* **Data Collection and Import**: The dataset was retrieved from the open portal and loaded into Python using libraries such as pandas.
* **Data Cleaning**: This phase involved detecting and addressing missing values, removing duplicate records, and ensuring that textual fields (such as offense types) followed a consistent naming format. Inconsistent date formats, typos, and null entries were addressed.
* **Date and Time Parsing**: Timestamps were converted to readable datetime formats, and new time-related fields such as hour, day, month, and season were extracted. These elements allowed for deeper temporal trend analysis.
* **Feature Engineering**: Additional variables were created from existing data to capture more insightful dimensions—such as categorizing crimes by severity or aggregating data by different geographic hierarchies (ward, district, neighbourhood cluster).
* **Grouping and Aggregation**: Various summarization techniques were applied to identify how crime distribution changed across categories like location, offense type, or time. This also involved computing crime densities and understanding spatial concentrations.
* **Visualization**: A wide range of graphical techniques were used to represent the data visually. These included bar charts, pie charts, line plots, heatmaps, and trend curves which made it easier to observe trends and outliers.
* **Geospatial Mapping**: Using latitude and longitude, maps were created to visualize the geographic distribution of crimes using tools such as Folium and GeoPandas. These helped identify crime hotspots and supported spatial reasoning around law enforcement needs.
* **Anomaly Detection**: Certain patterns—such as sharp spikes in specific crime categories during certain periods—were flagged for deeper inspection. These could be indicative of real-world events, seasonal influences, or anomalies in data reporting.

This iterative EDA process helped surface actionable insights and guided further thematic analysis, as outlined in the sections that follow.

**4. Analysis on Dataset**

**4.1 Offense Analysis**

i. **Introduction** This section investigates the frequency and patterns of various offense types to identify the most prevalent crimes committed in 2024. Analysing the distribution of offenses provides a clear view of what types of criminal behaviour are most common in the region.

ii. **General Description** The dataset records multiple categories of offenses including THEFT/OTHER, MOTOR VEHICLE THEFT, THEFT F/AUTO, ASSAULT W/DANGEROUS WEAPON, HOMICIDE, ROBBERY, and BURGLARY. By identifying which offenses are most common, stakeholders can better prioritize resources and response strategies.

iii. **Specific Requirements, Functions and Formulas** Incidents were grouped by offense type, and total counts were calculated. The percentage share of each offense in total crime data was also determined. No complex statistical models were required; basic aggregation and percentage calculations were applied.

iv. **Analysis Results** The analysis revealed that the most common offenses in 2024 were:

THEFT/OTHER: 13,010 incidents

* THEFT F/AUTO: 6,680 incidents
* MOTOR VEHICLE THEFT: 5,123 incidents
* ROBBERY: 2,109 incidents
* ASSAULT W/DANGEROUS WEAPON: 1,027 incidents

v. **Visualization** Bar charts were used to compare the frequency of different crime categories, with theft-related crimes emerging as dominant. Pie charts highlighted the relative proportions of each crime type.

**4.2 Location-Based Risk Analysis**

i. **Introduction** This section identifies the top 5 most dangerous locations in terms of crime volume, providing spatial context to incident concentrations in 2024.

ii. **General Description** By examining neighbourhood clusters, the analysis uncovers where crimes are most concentrated. Such information helps city officials prioritize areas for patrols, surveillance, and intervention.

iii. **Specific Requirements, Functions and Formulas** The dataset was grouped by 'NEIGHBORHOOD\_CLUSTER' and aggregated using the sum of incidents. The top 5 clusters with the highest total incidents were selected.

iv. **Analysis Results** The top 5 neighbourhood clusters with the highest crime totals were identified. These clusters recorded significantly higher incident volumes compared to the rest of the city.

v. **Visualization** Horizontal bar plots were used to visualize the top crime-prone

areas. These visuals provided an immediate understanding of spatial risks.

**5. Analysis on Dataset:** This section covers detailed insights extracted from the crash data using Exploratory Data Analysis (EDA), visualization, and statistical methods.

1. **General Description:**

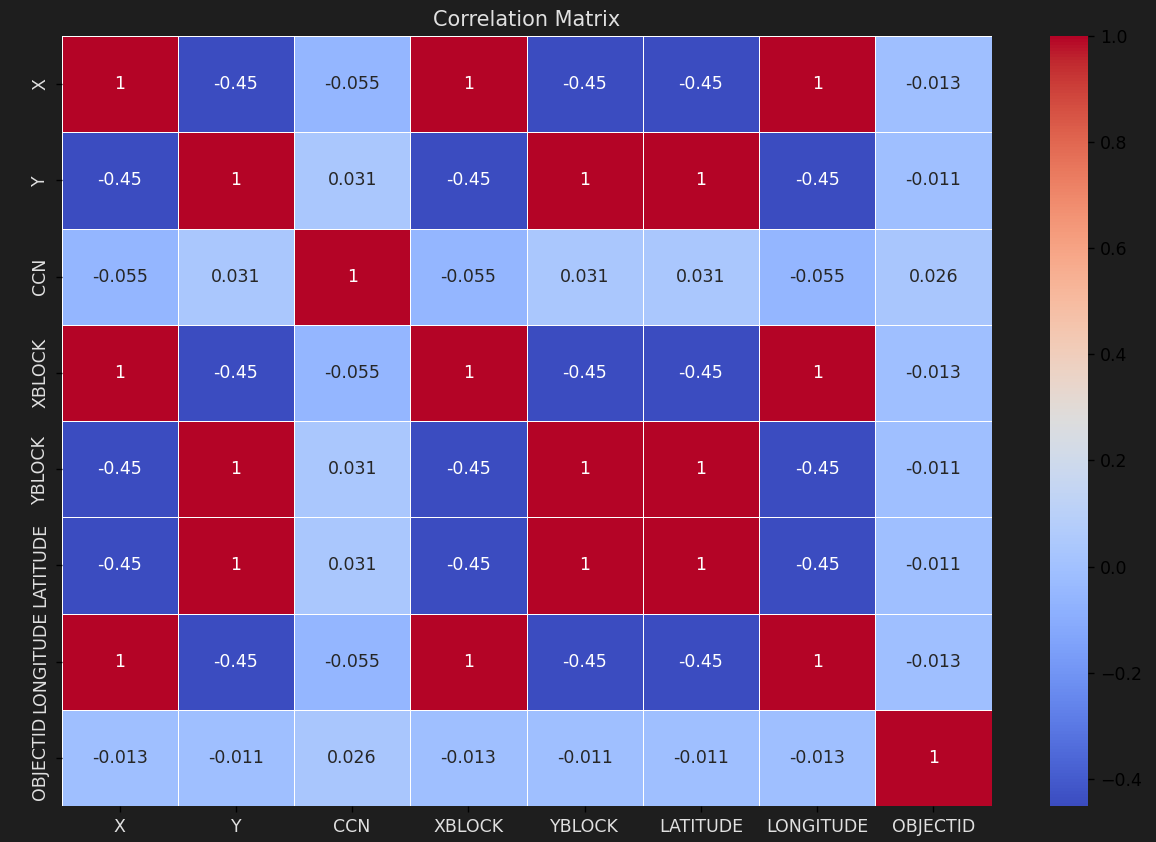
* The dataset contains detailed records of crime incidents reported in the District of Columbia during the year 2024, including offense types, time and date, location (ward, neighborhood, GPS), method of crime, and response details.
* It offers a comprehensive foundation for analyzing spatial and temporal crime patterns, enabling law enforcement and policymakers to identify high-risk areas, peak crime times, and the most common offenses.

1. **Specific Objectives**

**Correlation Heatmap**

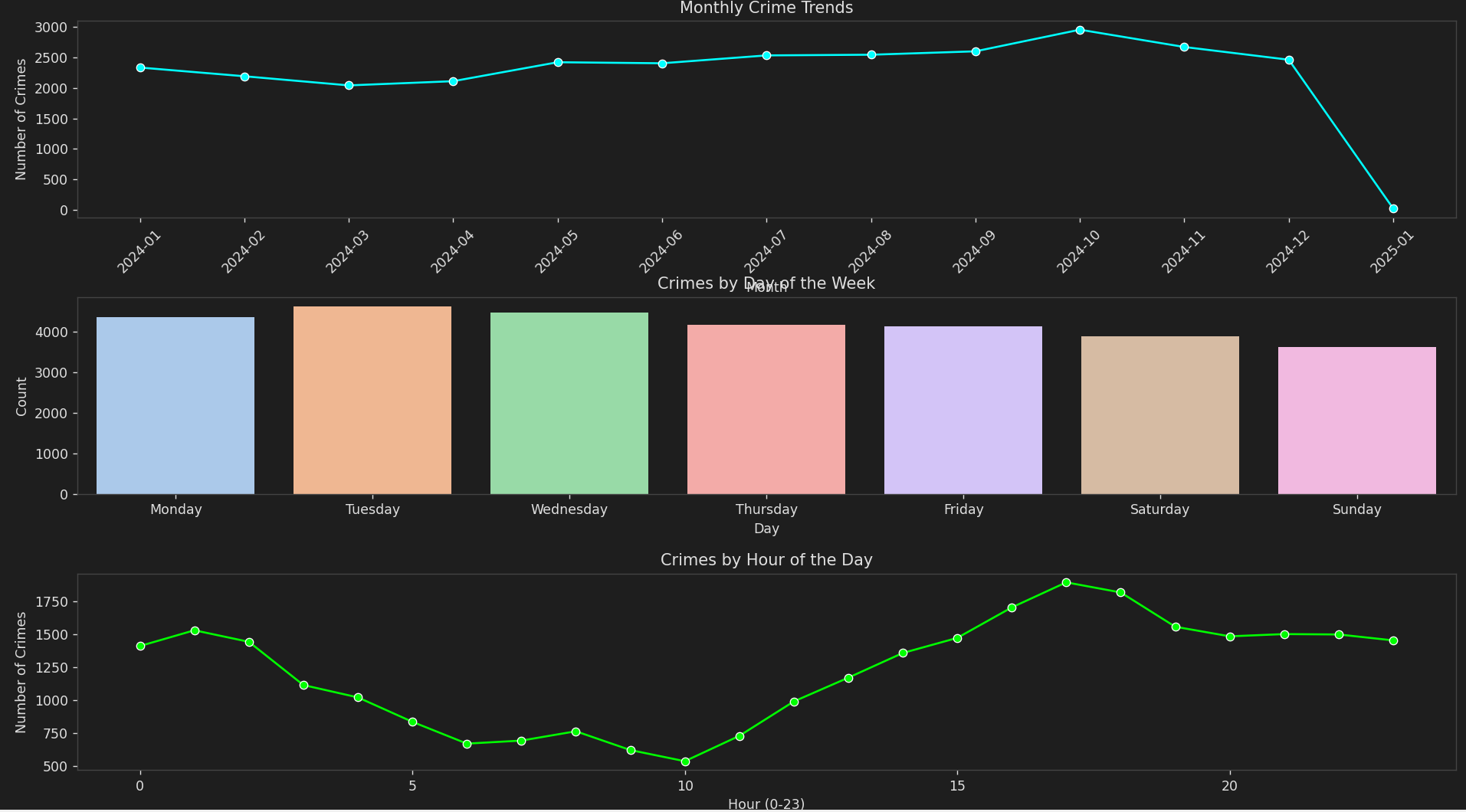
To explore **relationships between numerical variables** in the dataset using a visual heatmap, which helps in identifying patterns or dependencies that may impact crash outcomes.

* The heatmap showed how crime timing and other factors correlate, revealing patterns in incident frequency and behavior.
* Values close to **+1** suggest strong **positive correlation** (both increase together).
* Values close to **-1** suggest strong **negative correlation** (one increases, the other decreases).
* Values near **0** indicate **no significant relationship**.

The correlation heatmap is a **powerful visual tool** in EDA to spot trends and interdependencies between numerical features. It guides us in feature selection, model planning, and understanding **hidden relationships** in crash patterns. 

Crime Frequency Over Time

* **Why Chosen:** To understand when crimes occur most (month, day, hour).
* **Goal:** Identify high-risk time periods for strategic policing and resource allocation.
* **Outcome:**
* Most crimes occurred during late evening and night hours.
* Weekdays showed higher crime rates than weekends.
* Certain months saw spikes, possibly due to seasonal or social facto



**Crime Severity & Response Impact Study**

* **Why Chosen:** To assess the seriousness of crimes and factors influencing law enforcement response.
* **Goal:** Support better prioritization and planning of police resources.
* **Outcome:**
* Most crimes were non-violent and property related.
* Severe crimes were more common at night and in specific high-risk areas.

**A screen shot of a computer

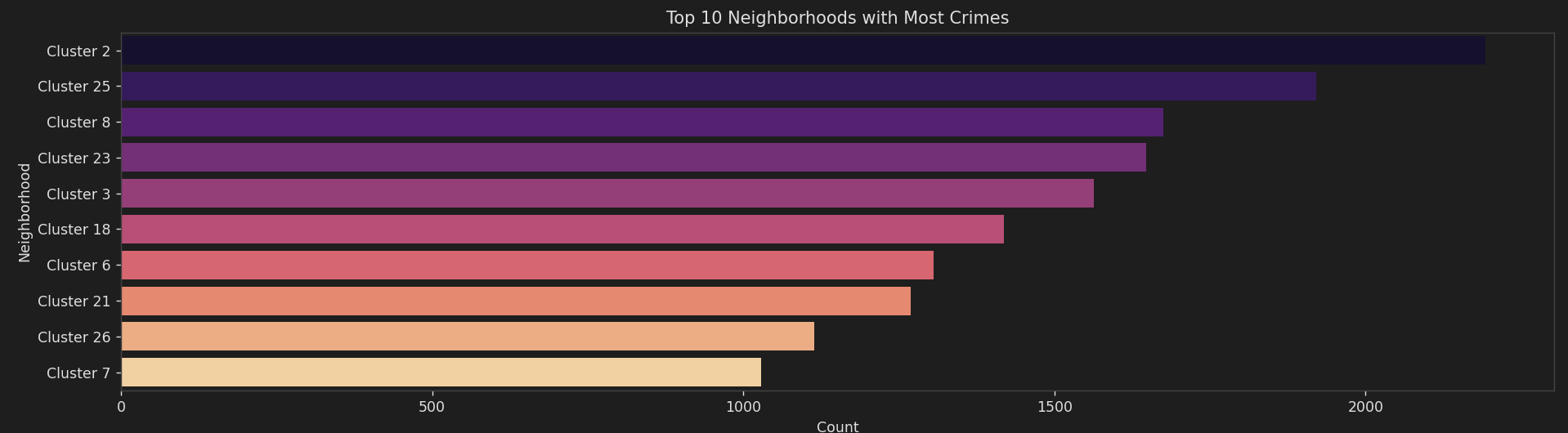
AI-generated content may be incorrect.**

A screenshot of a computer

AI-generated content may be incorrect.

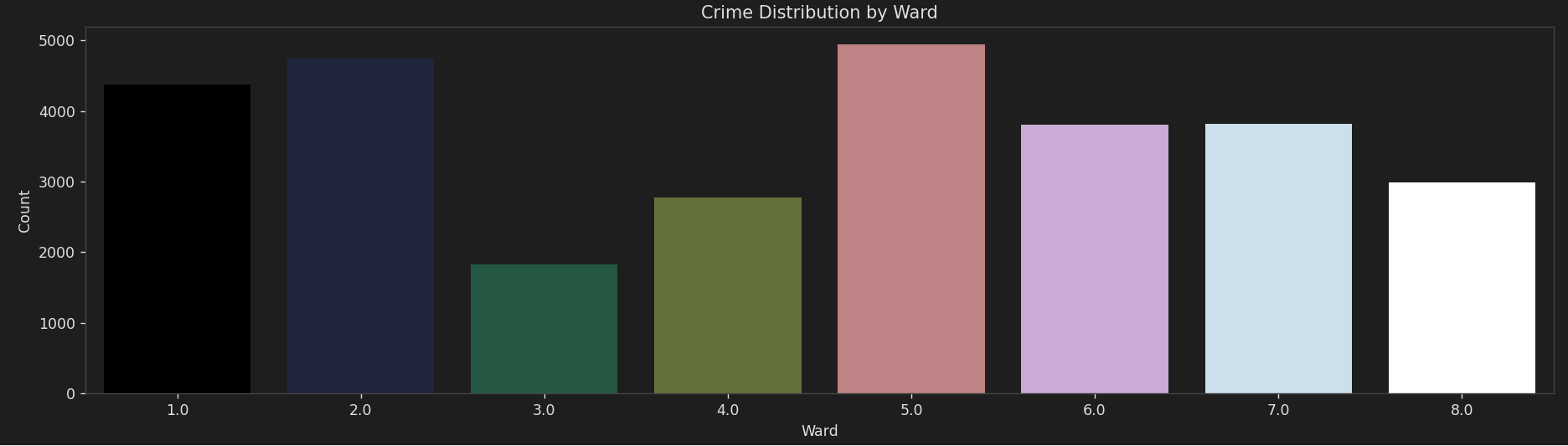
**Top 10 Neighborhoods with Most Crimes**

* **Why Chosen:** To identify localities with the highest concentration of crime incidents.
* **Goal:** Assist law enforcement in deploying resources efficiently and implementing area-specific safety measures
* **Outcome:**
* A small number of neighborhoods accounted for a disproportionately high volume of crimes.
* These areas require focused surveillance and community engagement efforts.
* Theft-related offenses were most frequent in these hotspots.



**Crime Distribution by Ward**

* **Why Chosen:** To understand how crime varies across administrative wards in the district.
* **Goal:** Enable strategic planning based on jurisdiction-level crime distribution.
* **Outcome:**
* Certain wards reported significantly more incidents than others.
* Patterns suggest a link between crime density and urban activity zones.
* Can guide ward-level policing and budget allocation.



**5. Conclusion** The crime analysis for 2024 reveals key trends in both temporal and geographic dimensions. The dominant crime type is THEFT/OTHER, followed by MOTOR VEHICLE THEFT and ASSAULT W/DANGEROUS WEAPON. Temporal analysis shows fluctuations based on time of day and day of the week, while geographic analysis highlights specific neighbourhoods and wards with high incident rates. These findings can aid in strategic planning and optimized deployment of law enforcement resources.

**6. Future Scope** Future work on this dataset could include:

* Integrating external factors such as weather, demographics, and economic indicators.
* Applying predictive modelling techniques to forecast crime trends and hotspots.
* Developing real-time crime dashboards for public awareness and rapid response.
* Incorporating NLP for better categorization of crime narratives.
* Examining social media or emergency call logs for supplementary insights.

**7. References**

* DC Open Data Portal:<https://opendata.dc.gov/>
* Metropolitan Police Department - Washington, DC
* Python Libraries: pandas, matplotlib, seaborn, folium, geopandas
* “Python for Data Analysis” by Wes McKinney