Kingston University London

CI7340 COURSEWORK 1
Crime Analysis in Montgomery County (2018 – 2022)

Submitted By Group 3:

Arya Pradeep Menon (K2275592)

Bhavana Koppula(K2278142)

Muskan Asmath (K2279003)

Shahesta Nazir Ahmed(K2283306)

Group Members and Participation:

Student Name	**Student ID**	**Group Assignment Development Contribution Percentage(%)**
Arya Pradeep Menon	K2275592	25%
Bhavana Koppula	K2278142	25%
Muskan Asmath	K2279003	25%
Shahesta Nazir Ahmed	K2283306	25%

Abstract

This study thoroughly analyses crime dynamics in Montgomery County using a diverse dataset. Through statistical and analytical methods, it identifies crime hotspots and temporal variations, offering crucial insights for law enforcement, policymakers, and community stakeholders. The analysis not only highlights high-crime areas but also assesses the impact of interventions and community programs on crime reduction. Ultimately, this research aims to improve public safety, guide resource allocation, and promote collaborative efforts for a safer Montgomery County.

Table of Contents

List of Figures

List of Tables

1. Introduction

Data science is a multidisciplinary field that leverages advanced techniques to extract meaningful insights from data. In the context of crime analysis, data science plays a crucial role in unravelling patterns, trends, and contributing factors within criminal activities. The Montgomery crime dataset represents the domain of law enforcement and public safety, where data-driven approaches aid in identifying high-crime areas in the Montgomery County of Maryland, USA. This application of data science contributes to evidence-based decision-making for law enforcement agencies, policymakers, and community stakeholders, ultimately fostering a safer environment in Montgomery County.

The increasing urban populations and complexities of modern societies makes it crucial to comprehend and address criminal activities. Montgomery County, with its diverse demographics and geography, reflects the complex dynamics of crime in today's urban settings. This study carefully examines crime in Montgomery County, using a detailed dataset on criminal incidents.

a. Problem Statement

This study addresses the need for a detailed analysis of crime dynamics in Montgomery County, Maryland, during the time period(2018 - 2022). This study seeks to unveil patterns, trends, and pivotal insights embedded in the dataset, with the ultimate goal of furnishing actionable intelligence for law enforcement agencies, policymakers, and the community at large. The core objectives encompass the identification of crime hotspots, an exploration of temporal variations, and an investigation into potential correlations among different crime types. By addressing these objectives, the analysis endeavours to contribute substantially to decision-making processes and the formulation of effective crime prevention strategies in Montgomery.

b. Objectives

Using advanced analytical and statistical methods, this study aims to offer a thorough insight into the crime situation in the County. Below are the major objectives of this analysis:

1. Identify patterns, trends and contributing factors underlying various criminal activities in Montgomery County.

2. Utilise advanced analytical and statistical methods to gain a comprehensive insight into the crime landscape in the County.

3. Pinpoint high-crime areas and analyse temporal patterns to better understand the distribution of criminal activities over the different states.

4. Evaluate the impact of interventions and community programs on crime reduction to guide future resource allocation and strategy development.

c. Research Questions:

After the initial assessment of the crime dataset, below is the list of different research questions that this study will investigate:

1	What are the predominant types of criminal activities, and how have their frequencies evolved from 2018 to 2022?
2	How responsive is police in each district?
3	Based on crime rates, which district necessitates increased police attention?
4	Which locations in districts have a higher likelihood of being a hotspot?
5	Which major drug cases have occurred across potential locations across cities?
6	How did the reforms and programs for the rising issue of illegal weapons in 2020 affect the crime rates related to weapons?
7	Which city where most people went missing?
8	What are the Seasonal trends in different fraud cases?
9	Which hours are most likely linked to accident prone cases on steet?
10	How does the frequency of larceny cases fluctuate over the week?

Table 1: List of Research questions

d. Workflow of the Research Methodology

Conducting a research methodology for the analysis of a crime dataset for Montgomery County involves several steps. Below is a general framework of the major tasks that we adapted based on the objectives and scope of the research:

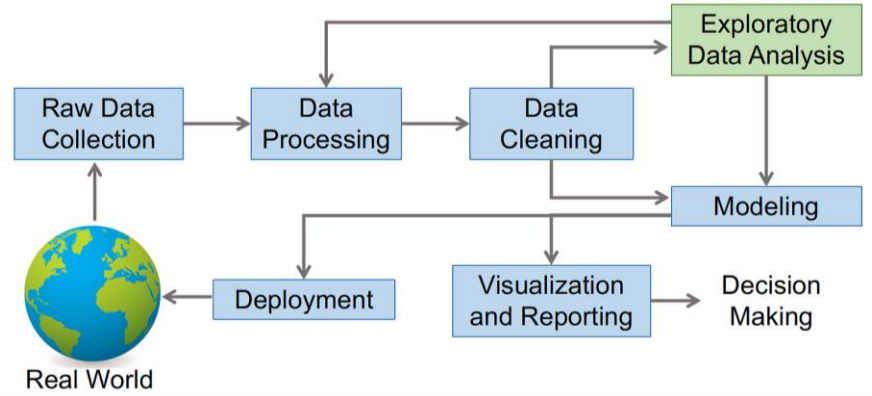

Figure 1:Data Science Processes Workflow [12]

This is a high-level overview, where each rectangular box represents a specific task or stage in the analysis process. The arrows indicate the flow of the analysis, with each task leading to the next. There could be iterations or feedback loops between stages.

1. Data Collection

A comprehensive crime dataset for Montgomery County from the official website is used. This dataset includes information on the type of crimes, locations, time of occurrence, and other relevant variables and ensures the data is reliable, up-to-date, and covers the required time period.

2. Outline Research Objectives

Define clearly the problem statement and objectives for the research. What specific aspects of crime dynamics would be under investigation and how to focus on spatial patterns, temporal trends, types of crimes, and the effectiveness of interventions and reforms.

3. IDA(Initial Data Analysis)

- Data Processing:

Involves handling missing values, and outliers, and ensuring data consistency. Converting the categorical variables into appropriate formats and standardise or normalise numerical variables as needed.

- Data cleaning:

Data cleaning for the crime dataset analysis includes tasks such as removing duplicates, correcting inaccuracies, and transforming variables for accurate and reliable analyses.

4. Exploratory Data Analysis (EDA)

EDA helps in gaining initial insights into the dataset. By using statistical and visual techniques, the analysis examines the distribution of crime types, identifies the trends, crime hotspots and explores spatial and temporal patterns which involve daily, monthly, or yearly trends.

- Statistical Analysis:

As part of EDA, apply appropriate statistical techniques to test hypotheses or relationships within the dataset. For example, for exploring the correlations between certain variables or to find the significance of temporal trends.

5. Visualisation and Reporting

A comprehensive report communicates the data sources, tools and platforms used, methodology and analysis techniques thoroughly with the visualisations and supporting evidence. Two visualisations for each research question provide insights, and identify the patterns and trends and contributing factors underlying various criminal activities.

6. Conclusion and Recommendations:

Summarise findings, draw conclusions, and offer recommendations. Discuss implications for law enforcement, policymakers, and stakeholders. Evaluate the effectiveness of interventions by comparing crime rates before and after implementation if applicable.

2. Preliminary Data Analysis

a. Dataset

The dataset contains the details of crimes reported in the Montgomery County of Maryland State, USA. The information displayed is based on reported crimes that have been categorised using the Criminal Justice Information Services (CJIS) Division Uniform Crime Reporting (UCR) Program's National Incident-Based Reporting System (NIBRS), and it has been verified by authorised police incident reports. The primary source of data is the database of found crimes maintained by the Montgomery County Police Department is accessible to the public through the website, dataMontgomery [1].

The data is available as a CSV file. The 306094 records (rows) and 30 features (columns) make up the structured format of the data. The table below shows the specifics of each feature.

Feature	Data Type	Description[1]	Sample Data
Incident ID	Integer	Police Incident Number	201202980
Offence Code	String	Offense_Code is the code for an offence committed within the incident as defined by the National Incident-Based Reporting System (NIBRS) of the Criminal Justice Information Services (CJIS) Division Uniform Crime Reporting (UCR) Program.	3550
CR Number	Integer	Police Report Number	180042096
Dispatch Date / Time	String	The actual date and time a Officer was dispatched	08/23/2018 09:52:08 PM
NIBRS Code	String	FBI NIBRS Codes	35B
Victims	Integer	Number of Victims	1
Crime Name1	String	Crime against Society/Person/Property or Other	Crime Against Society
Crime Name2	String	Describe NIBRS Code	Drug Equipment Violations
Crime Name3	String	Describe Offence Code	DRUGS - NARCOTIC EQUIP - POSSESS

Police District Name	String	Name of District	GERMANTOWN
Block Address	String	Address in 100 block level	12800 BLK MIDDLEBROOK RD
City	String	City	GERMANTOWN
State	String	State	MD
Zip Code	Float	Zip Code	20874
Agency	String	Assigned Police Department	MCPD
Place	String	Place Description	Street - In vehicle
Sector	String	Police sector name, a subset of District	N
Beat	String	Police patrol area, a subset of Sector	5N1
PRA	String	Police Response Area, a subset of Beat	447
Address Number	Float	House or Business Number	12800
Street Prefix	String	North, South, East, West	N
Street Name	String	Street Name	MIDDLEBROOK
Street Suffix	String	Quadrant (NW, SW, etc)	NE
Street Type	String	Ave, Drive, Road, etc	AVE
Start_Date_Time	String	Occurred from date/time	08/23/2018 09:52:00 PM
End_Date_Time	String	Occurred to date/time	06/15/2018 03:00:00 AM
Latitude	Float	Latitude	39.147954
Longitude	Float	Longitude	-77.218189
Police District Number	String	Major Police Boundary	6D
Location	String	Location	(39.148, -77.2182)

Table 2: Dataset Description

b. Data Quality Initial Assessment

The quality of data is the foundation of insightful outcomes. It directly influences the efficiency of further analysis and the accuracy of results. Thus, the data needs to be clean to guarantee the robustness of analysis. The data cleaning process includes:

1. Removing irrelevant data
2. Removing duplicates
3. Handling inconsistent values
4. Handling missing values
5. Converting data types
6. Removing outliers

The table below shows various methods for handling missing values, pre-processing, and transformation techniques.

Process	Methods
Handling Missing Values	Remove records with missing value
	Impute missing values with constant values
	Impute missing values with appropriate values (Eg: Mean, Mean, Calculated values)
	Interpolation
	Use algorithms to predict missing values
Handling Noisy Data	Binning Method
	Regression
	Clustering
Data Transformation	Normalisation (z-score, min-max)
	Attribute Selection
	Discretization
	Encoding Categorical Variables
	Concept Hierarchy Generation
Data Reduction	Data Cube Aggregation
	Attribute Subset Selection
	Numerosity Reduction
	Dimensionality Reduction

Table 3: Data pre-processing steps [3]

Additionally, Data Wrangling techniques can be applied to the data to make it more suitable for data analysis. It transforms the raw data into the desired form for better analysis [4]. The different types of data-wrangling techniques which are applied to this dataset include:

1. Data Extraction which extracts relevant data from the dataset
2. Data Merging which combines data to create a more comprehensive dataset.
3. Construction of analytical dataset which creates a new dataset with summarised and/or aggregated data at different levels.

In the initial phase, the information about the dataset is taken using the df.info(). The result showed the presence of missing values and the necessity of amending the datatypes of certain features. There are no duplicate records and the percentage of null values for each feature is calculated to support further analysis.

```
def missing_percent():
  missing_percent = round(df.isna().sum()*100/df.shape[0],2)
  print(missing_percent.sort_values())

missing_percent()
```

```
Incident ID               0.00
Police District Number    0.00
Longitude                 0.00
Latitude                  0.00
Start_Date_Time           0.00
Street Name               0.00
Location                  0.00
Agency                    0.00
State                     0.00
Place                     0.00
Offence Code              0.00
CR Number                 0.00
NIBRS Code                0.00
Victims                   0.00
Zip Code                  0.02
Police District Name      0.04
Street Type               0.04
PRA                       0.10
Crime Name3               0.12
Crime Name2               0.12
Crime Name1               0.12
City                      0.42
Sector                    0.67
Beat                      0.67
Dispatch_Date_Time        3.09
Address Number            9.53
Block Address             9.56
End_Date_Time            54.62
Street Prefix            95.42
Street Suffix            98.28
```

Figure 2:Percentage of missing values per attribute

Thereafter extensive data cleaning is carried out which includes handling missing values and dealing with inconsistencies and discrepancies in data. The below table shows the various transformations or pre-processing applied to different features.

Feature	Transformation/Pre-processing Applied	Justification
Incident ID	Data type changed from int to string	
Offence Code	Data type changed from int to string	
CR Number	Data type changed from int to string	
Zip Code	Data type changed from float to string	
	Data is formatted to remove the decimal point	Example: 20783.0 is converted to 27083
Address Number	Data type changed from float to string	
Start_Date_Time	Data type changed from object to datetime	
	Records with Start_Date_Time between 2018 and 2022 are filtered out.	This research aims to analyse the crimes that occurred between 2018 and 2022
Dispatch Date / Time	Data type changed from object to datetime	
	Column renamed to Dispatch_Date_Time	To maintain uniformity of attribute names
State	Records with value 'MD' are filtered out	This research focuses on the crimes that happened in Montgomery County of Maryland. 'MD' stands for Maryland
Start_Date_Time	Records with Start_Date_Time between 2018 and 2022 are filtered out.	This research aims to analyse the crimes that occurred between 2018 and 2022.
Crime Name2	Dropped records with value 'NOT NIBRS CODE'	These are Traffic offences and are not collected by the FBI. This is 0.08% of the data.
	Dropped records with value 'Runaway'	The FBI stopped collecting those data from 2011 and this research focuses on data from 2018 to 2022. This makes up 1.15% of the data.
Police District Number	Typos are handled appropriately	Example: 1.0D is changed to 1D
City	Typos are handled appropriately	Example: BATHEDA is changed to BATHESDA

Street Prefix	Feature dropped	Feature not required for this analysis
Street Suffix	Feature dropped	Feature not required for this analysis
End_Date_Time	Feature dropped	Feature not required for this analysis
Location	Feature dropped	Feature not required for this analysis
Agency	Feature dropped	Feature not required for this analysis
PRA	Feature dropped	Feature not required for this analysis

Table 4: IDA Processes

The null values in various features are handles as follows:

1. The missing values in attributes 'Crime Name1' and 'Crime Name2' are handled using 'NIBRS Code' with respect to the official mapping provided by the UCR Program [13].

2. As 'Crime Name3' describes the 'Offense Code', 'Offense Code' is being used to handle the missing values in 'Crime Name3'. When investigated, the 'Offense Code' for those records with no 'Crime Name3' seems to be invalid according to the official documentation [14]. Hence the null values of 'Crime Name3' are filled with the value 'Incorrect Offence Code'.

3. As per the documentation [13], there exists a one-to-one mapping with 'NIBRS Code', 'Crime Name1', and 'Crime Name2'. Thus, all the records where there is a discrepancy in 'Crime Name1' and 'Crime Name2' in association with 'NIBRS Code' are handled accordingly.

4. To fill the null values in 'Police District Name', the value of 'Police District Name' for the corresponding 'Police District Number' is found from the other available records.

5. 'Sector' and 'Beat' attributes have null values for those records where 'Police District Number' is 'OTHER' and 'TPPD'. As Sector is a subset of Police District and Beat is a subset of Sector, the null values are filled with 'OTHER' and 'TPPD' for both the attributes corresponding to values in 'Police District Number'.

6. The attributes 'Police District Name', 'Street Name', and 'Zip Code' are considered to find the 'City' wherever possible. For other records, where 'Police District Name' is 'CITY OF TAKOMA PARK' and 'TAKOMA PARK', the 'City' is filled with the value 'TAKOMA PARK'. Similarly, if 'Police District Name' is 'ROCKVILLE' or 'SILVER SPRING', the 'City' value is 'ROCKVILLE' or 'SILVER SPRING' respectively.

4. Exploratory Data Analysis

a. Introduction to EDA

Exploratory Data Analysis (EDA) is a process that involves scrutinising an available dataset to identify patterns, detect anomalies, test hypotheses, and assess assumptions through the application of statistical methods. The primary objective is to glean insights from the data before embarking on formal modelling or hypothesis development. It serves as the initial phase where one comprehends the information encapsulated in the data, enabling the visualisation of data and the formulation of hypotheses for subsequent analysis. EDA plays a pivotal role in summarising data, conducting statistical analyses, and visually representing information. Unlike confirmatory data analysis, EDA embraces an "exploratory" approach, signifying a dynamic understanding of the problem at hand as insights evolve during the process.

The fundamental tools of Exploratory Data Analysis (EDA) encompass plots, graphs, and summary statistics. Broadly, EDA involves systematically navigating through the data, creating distributions for all variables (utilising box plots), plotting time series data, transforming variables, examining pairwise relationships through scatterplot matrices, and calculating essential statistics such as mean, minimum, maximum, upper and lower quartiles, and outlier identification.

While EDA involves substantial visualisation, it is distinct from data visualisation as the former occurs early in the analysis, whilst the latter is employed later to communicate findings.

As part of the research analysis, after the IDA process, below are the further data cleaning and transformations before visualising the crime analysis:

1. Outliers: Latitude and Longitude

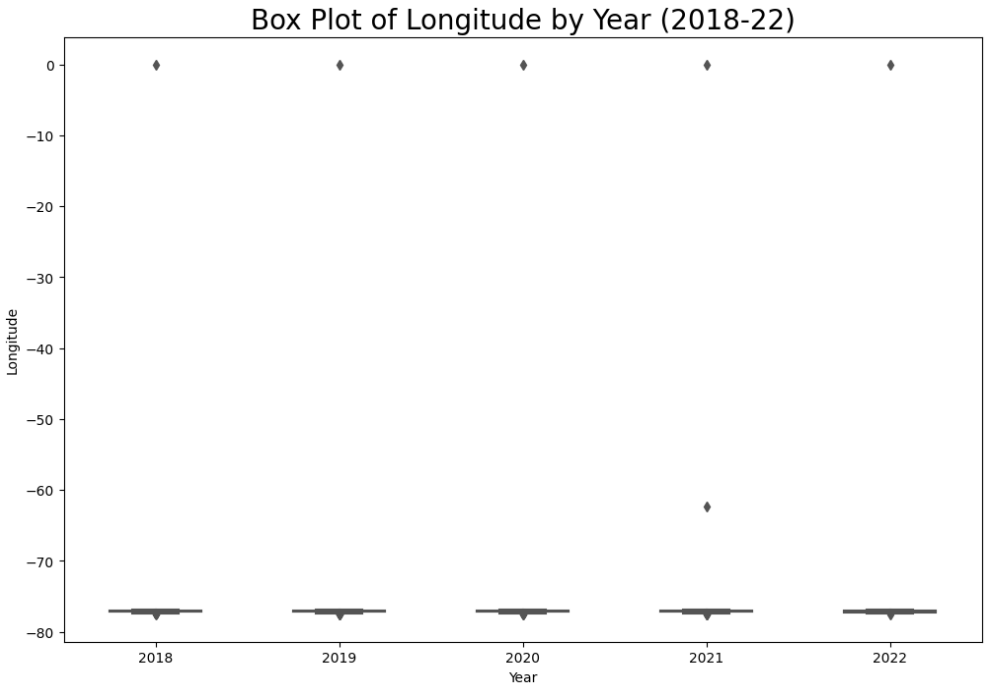

Figure 3:Box plot of Longitudes by Year

As seen in the graph above, the column of longitude range in the crime dataset exhibits outliers, notably at 0 and 60 degrees. These outliers may stem from manual data updation or missed information entry in the dataset, and it's crucial to investigate their impact on the overall analysis. While these anomalies could introduce bias, steps have been taken to remove these outliers as they lie outside the County. This ensures the robustness of the analysis.

2. Start Datetime and Dispatch Datetime:

In the dataset, certain entries showed a crime start datetime greater than the dispatch datetime, contrary to the expected sequence. Assuming a recording error, only the incorrect values were interchanged so that 'Start_Date_time' is consistently lower than 'Dispatch_Date_time.' This ensures a positive police response time.

3. Response Time:

Police response time in seconds was computed using the 'Start_Date_time' and 'Dispatch_Date_time.' These values were then utilised to visualise the distribution of police response times (refer visualisation of Q2).

4. Additional Attributes:

The 'Start_Date_time' attribute underwent further transformation into year, month, and days of the week before being employed for research visualisations (refer to Q1, Q2, Q6, Q8]).

An initial assessment of the data depicted in the graph below can be used to visualise the distribution of the different crime types, which facilitates the formation of research grounds.

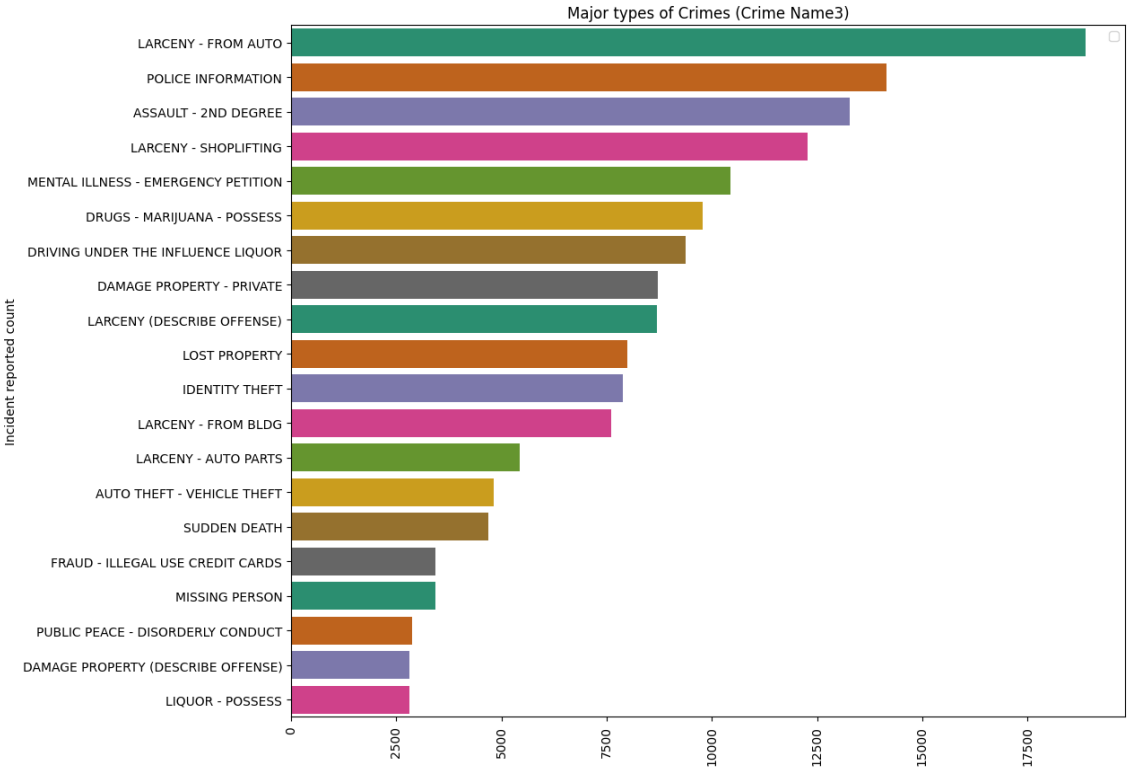

Figure 4: Major Types of Crimes

There are three main approaches of EDA to consider: univariate, bivariate and multivariate [11]:

Types	Description	Example Use Case
Univariate Non-Graphical EDA	Focuses on analysing a single variable's data to understand distribution and key statistical parameters, including central tendency, range, and dispersion.	No use case
Univariate Graphical EDA	Utilises visual methods for single-variable analysis, featuring stem-and-leaf plots and histograms for quick insights into central tendency, dispersion, and outliers. We could use different types of histograms: Simple Bar Charts, Multiple or Grouped, Percentage Bar Charts, Box Plots	Plotting 'Year' and their incident counts using grouped bar chart Plotting the 'Days of the Week' and their incident counts using swarm plot Refer to: [Q2, Q10]
Multivariate Non-Graphical EDA	Reveals relationships between two or more variables through cross-tabulation or statistics. Particularly valuable for categorical data, creating a two-way table.	No use case
Multivariate Graphical EDA	Depicts relationships between two or more variables through visual methods. Common types include: Scatter Plot, Multivariate Chart, Run Chart, Bubble Chart, Heat Map	A heat map showcasing a relationship between 'Street Name' and the 'Hours of the day'. A stacked bar chart showing a relationship between places and crime types. A grouped bar chart showing a relationship between the year and crime type. Refer to: [Q5, Q6, [Q9]

Table 5:Types of EDA and their use case

Dataset Assumptions:

1) Rows with Latitude and Longitude with (0,0), which map to Null Island and map to coordinates that are missing.
2) External factors that might influence the observed correlations, such as socioeconomic factors, population density, or law enforcement policies are not considered while considering cities as hotspots.

b. Descriptive Statistics

Descriptive statistics, in the context of Exploratory Data Analysis (EDA), involves the use of quantitative measures to summarise and describe key features of a dataset. These statistics provide a concise overview of the main characteristics, patterns, and trends within the data.

1. Mean (Average): The sum of all values divided by the number of observations. It provides a measure of central tendency. - latitude mean

2. Median: The middle value in a dataset when arranged in ascending or descending order. It is less sensitive to extreme values than the mean. - police response time

3. Range: The difference between the maximum and minimum values, indicating the spread of the data.

4. Variance: A measure of how much the values in a dataset deviate from the mean.

5. Quartiles: Values that divide the dataset into four equal parts. The first quartile (Q1) is the 25th percentile, the second quartile (Q2) is the median (50th percentile), and the third quartile (Q3) is the 75th percentile.

These descriptive statistics offer a summary of the dataset's central tendency, variability, and shape, allowing analysts to form a preliminary understanding before delving into more advanced analyses.

Below is a snapshot of df.describe() :

```
[367] df.describe()
```

	Victims	Latitude	Longitude	Response_Time
count	218753.000000	218753.000000	218753.000000	218753
mean	1.024041	37.836131	-74.658111	7 days 03:45:54.010614711
std	0.192932	6.865243	13.544423	46 days 20:24:19.925107173
min	1.000000	0.000000	-77.516753	0 days 00:00:00
25%	1.000000	39.017292	-77.196581	0 days 00:00:38
50%	1.000000	39.071405	-77.098969	0 days 00:15:01
75%	1.000000	39.141296	-77.028170	0 days 17:37:04
max	12.000000	90.000000	0.000000	1625 days 22:14:29

Figure 5: Descriptive Statistics

c. Data Visualisation

Data visualisation is the graphical representation of information and data. By using visual elements like charts, graphs and maps, data visualisation tools provide an accessible way to see and understand trends, outliers, and patterns in data. [10]

Chart Type	Sub Types	Description
Line Plots	Line Plot	Shows trends in data points over a continuous interval.
	Spline Chart	Smoothed line plot emphasising trends in data.
Bar Charts	Grouped Bar Chart	Compares categories or groups with distinct bars.
	Horizontal Bar CHart	Displays data using horizontal bars for comparison.
	Stacked Bar Chart	Illustrates parts of a whole with stacked bars.
	Comparative Bar Chart	Compares data across different categories.
Radial Charts	Pie Chart	Represents proportions of a whole as slices of a circle.
	Donut Chart	Similar to a pie chart but with a hole in the centre.
	Nightingale Chart	A specialised pie chart with multiple layers for additional details.
	Radial Chart	Displays data in a circular manner, often useful for proportions.
Scatter Plots	Scatter Plot	Shows relationships between two continuous variables.
	Bubble Chart	Enhances scatter plot with bubble sizes representing additional data.
Distribution Plots	Distribution Plot	Visualizes data distribution and statistical characteristics.
	Swarm Plot	Depicts categorical data points along a single axis.
	Box Plot	Illustrates statistical measures and outliers in a dataset.
Specialised Charts	Tree Map	Displays hierarchical data using nested rectangles.
	Bi-Directional Chart	Represents relationships between two variables in both directions.
	Heat Map	Depicts data values using colours in a matrix format.
	Count Plot	Displays the count of categorical variables.
	Area Plot	Shows data trends using filled areas on a graph.

Table 6: Types of Data Visualisations

Q1. What are the predominant types of criminal activities, and how have their frequencies evolved from 2018 to 2022?

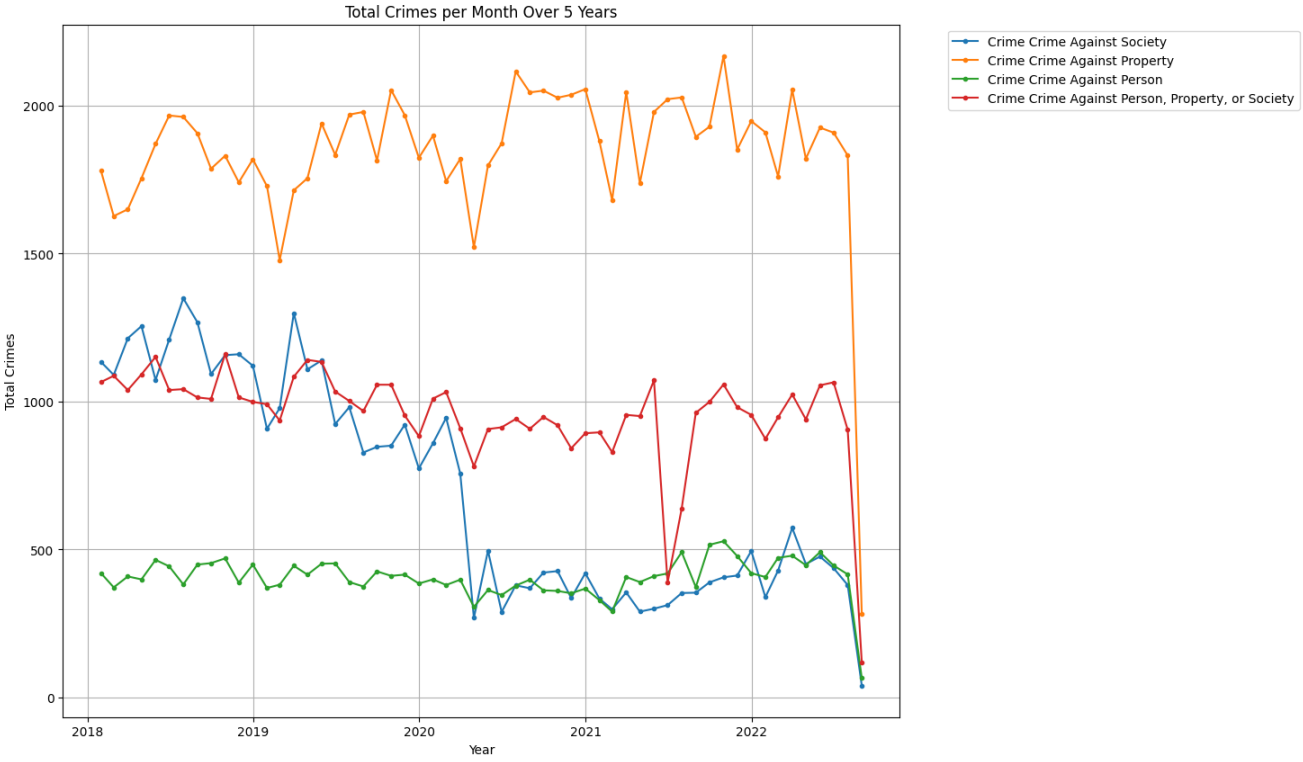

Figure 6: Line Plot - Total crimes over the years

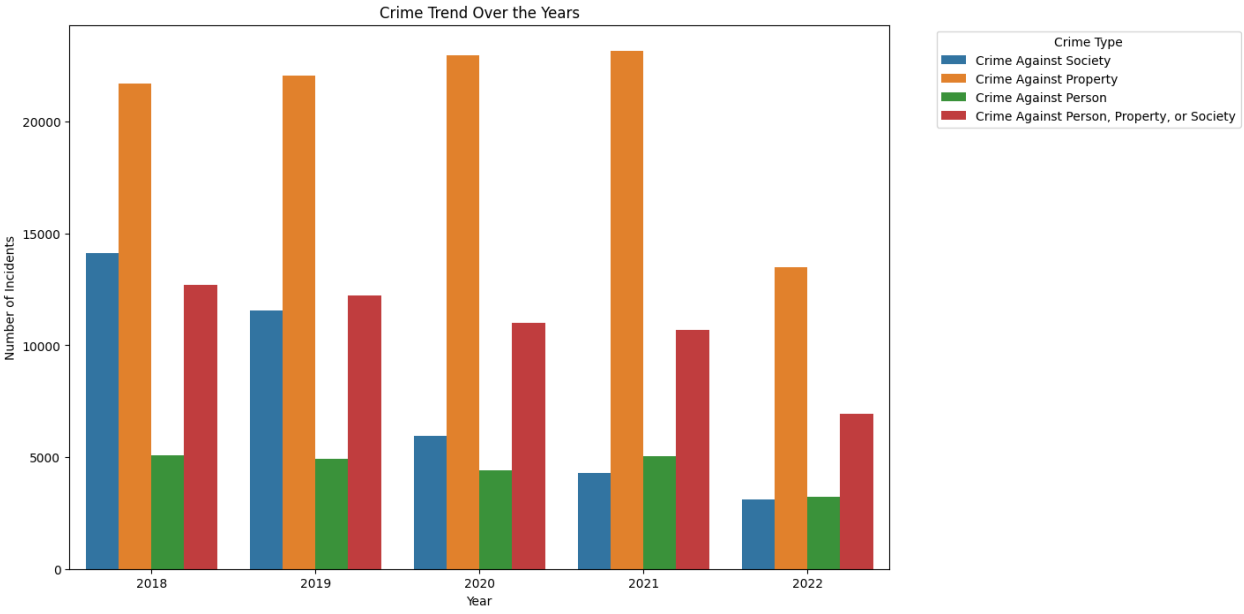

Figure 7: Count Plot - Total crimes over the years

INFERENCE:

Both graphs depict monthly crime patterns over the years, using distinct lines/bars for various crime types to highlight temporal fluctuations and suggest that Crime against property has the highest recorded incidents across all years.

Q2. How responsive are the police in each district?

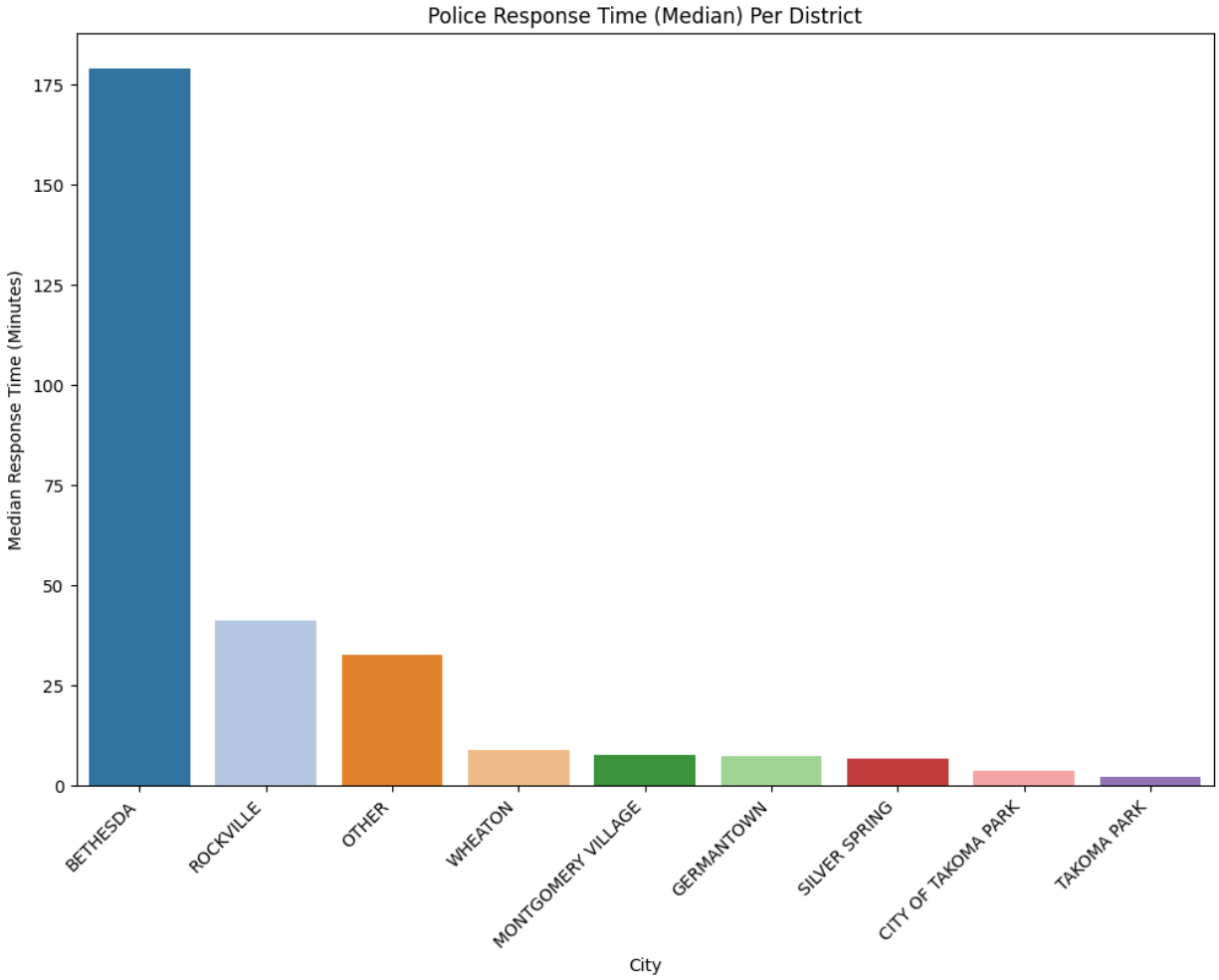

Figure 8: Bar Chart - Police response Time(median) per district

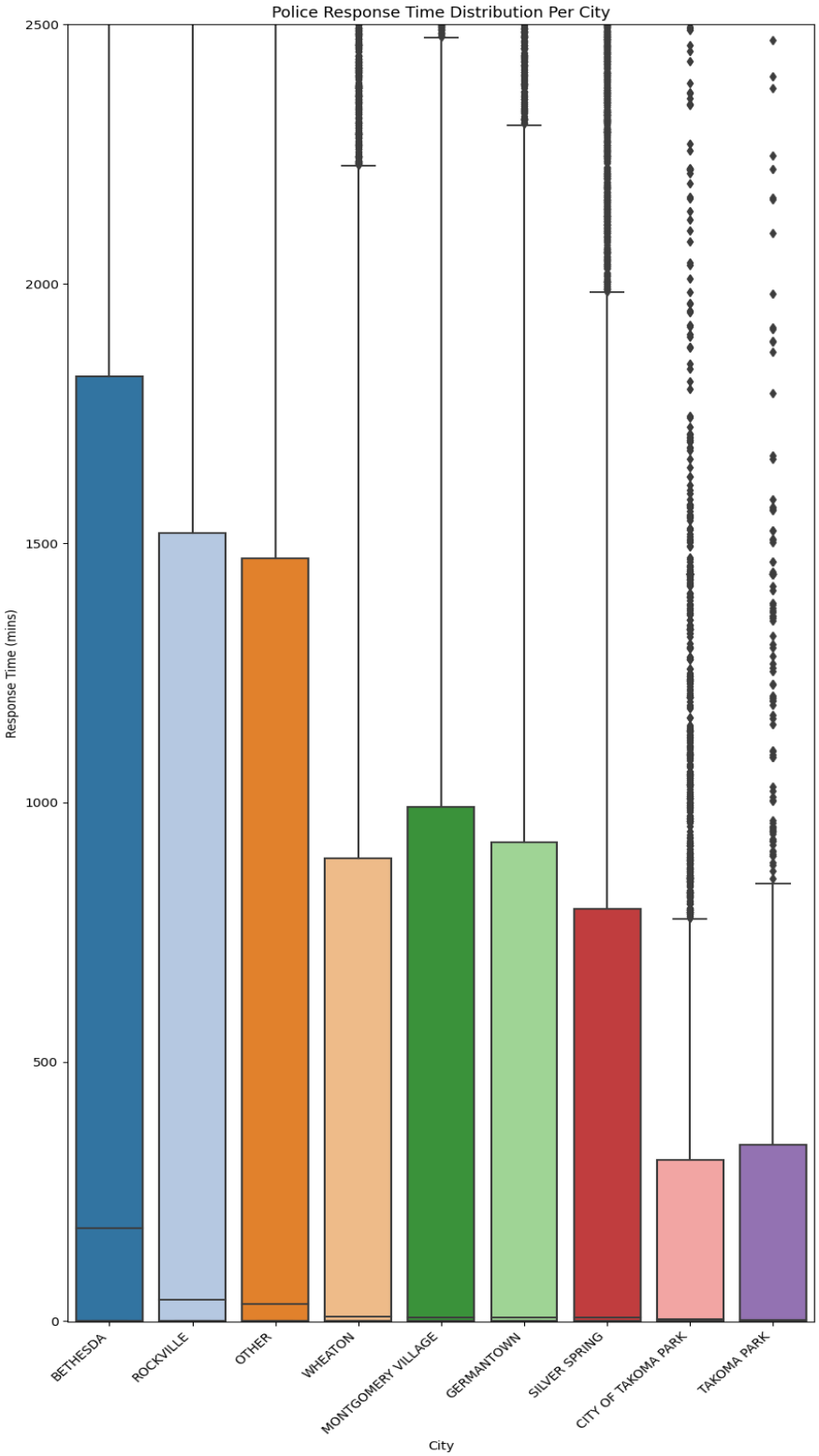

Figure 9: Box Plot - Police response Time(median) per district

INFERENCE:

The 2 graphs illustrate a comparison of median police response times across districts. Bethesda and Rockville had the highest response time medians and ranges. Thus, the authorities should try to deploy more forces and reforms within the police force to reduce the response time.

Q3. Based on crime rates, which district necessitates increased police attention?

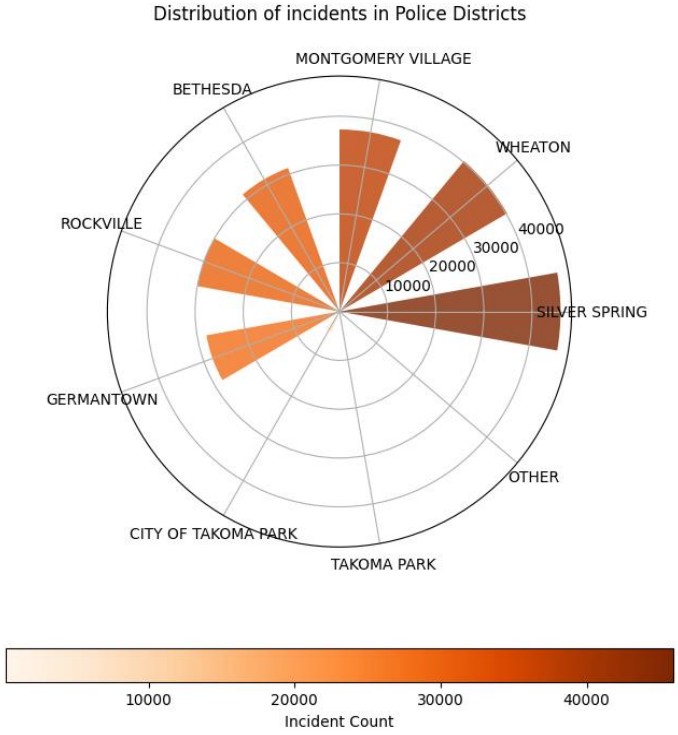

Figure 10: Radial Chart - Police response Time(median) per district

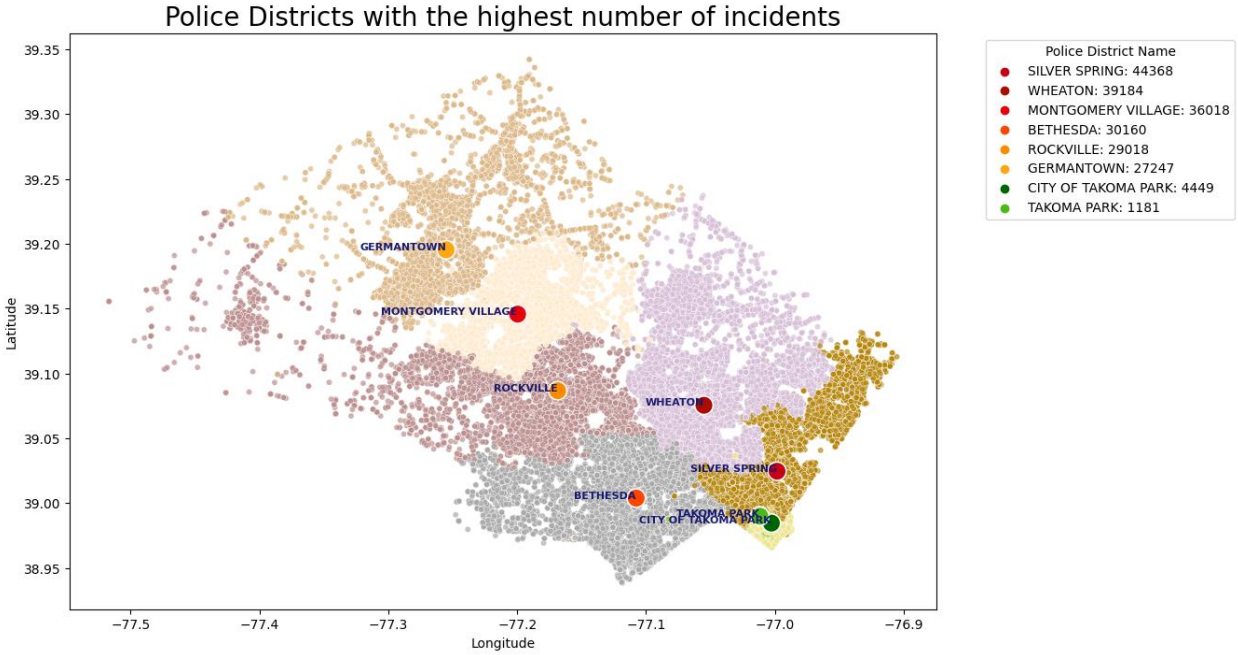

Figure 11: Scatter Plot - Police response Time(median) per district

INFERENCE:

From the above graphs, the cities - Silver Spring, Wheaton and Montgomery Village, have the highest recorded crime cases and require more police attention and stricter law enforcement.

Q4. Which locations in districts have a higher likelihood of being a hotspot?

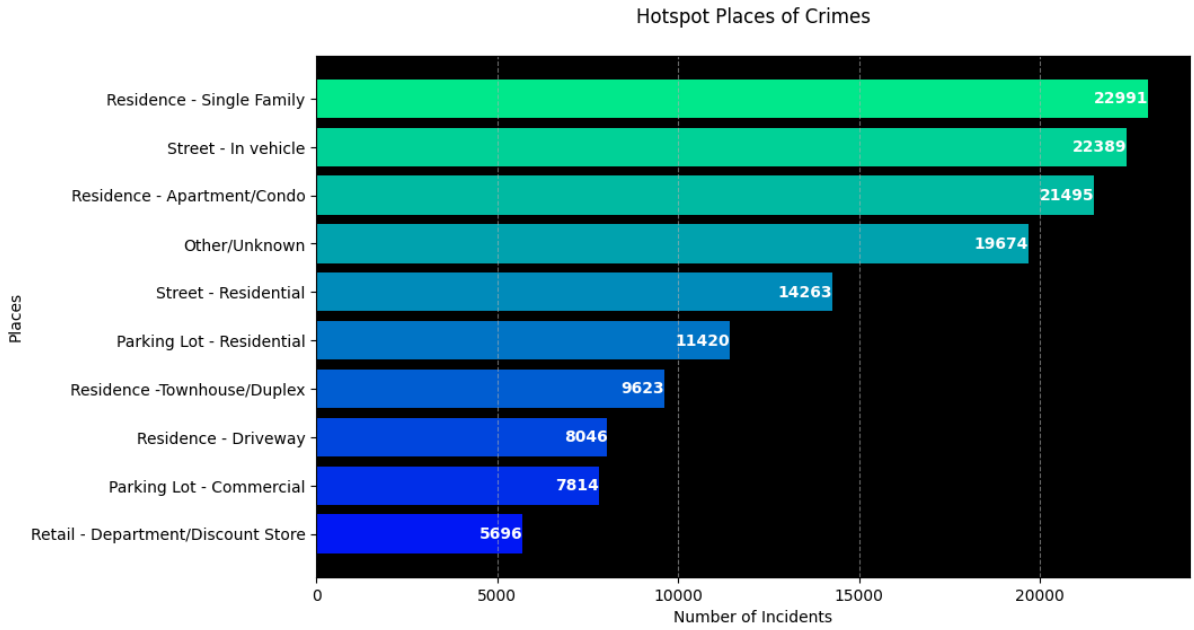

Figure 12: Horizontal Bar Chart - Hotspot of Crimes (Places)

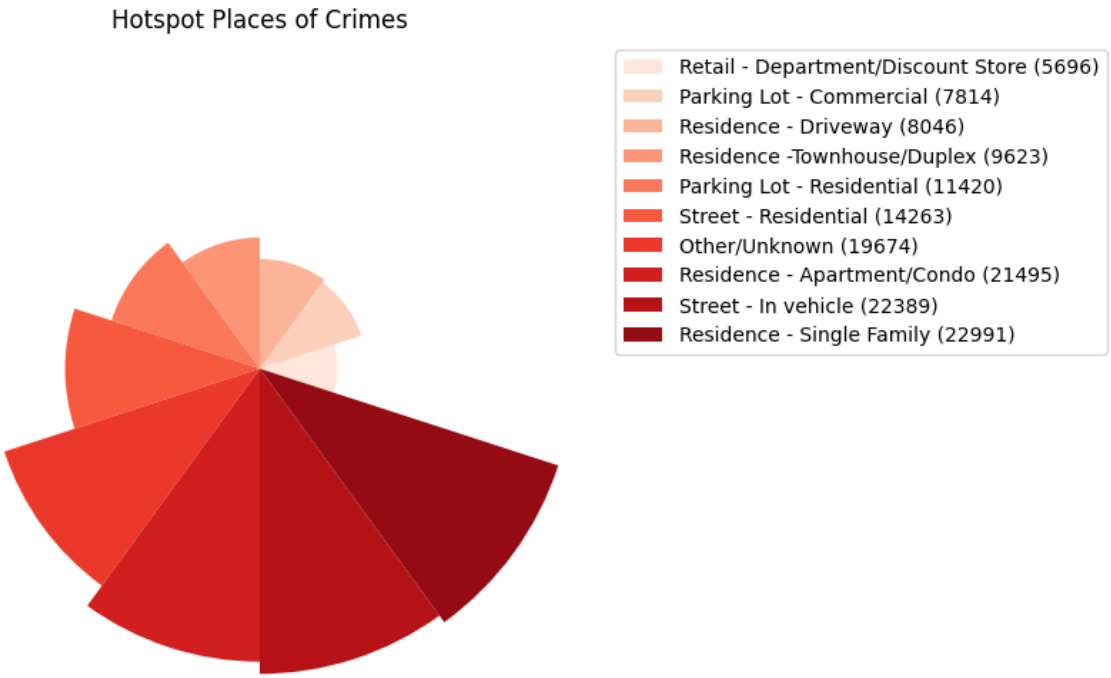

Figure 13: Nightingale Chart - Hotspot of Crimes (Places)

INFERENCE:

The graphs show hotspots of crimes as residence and streets of the major crime cities of Montgomery.

Q5. Which major drug cases have occurred across potential locations across cities?

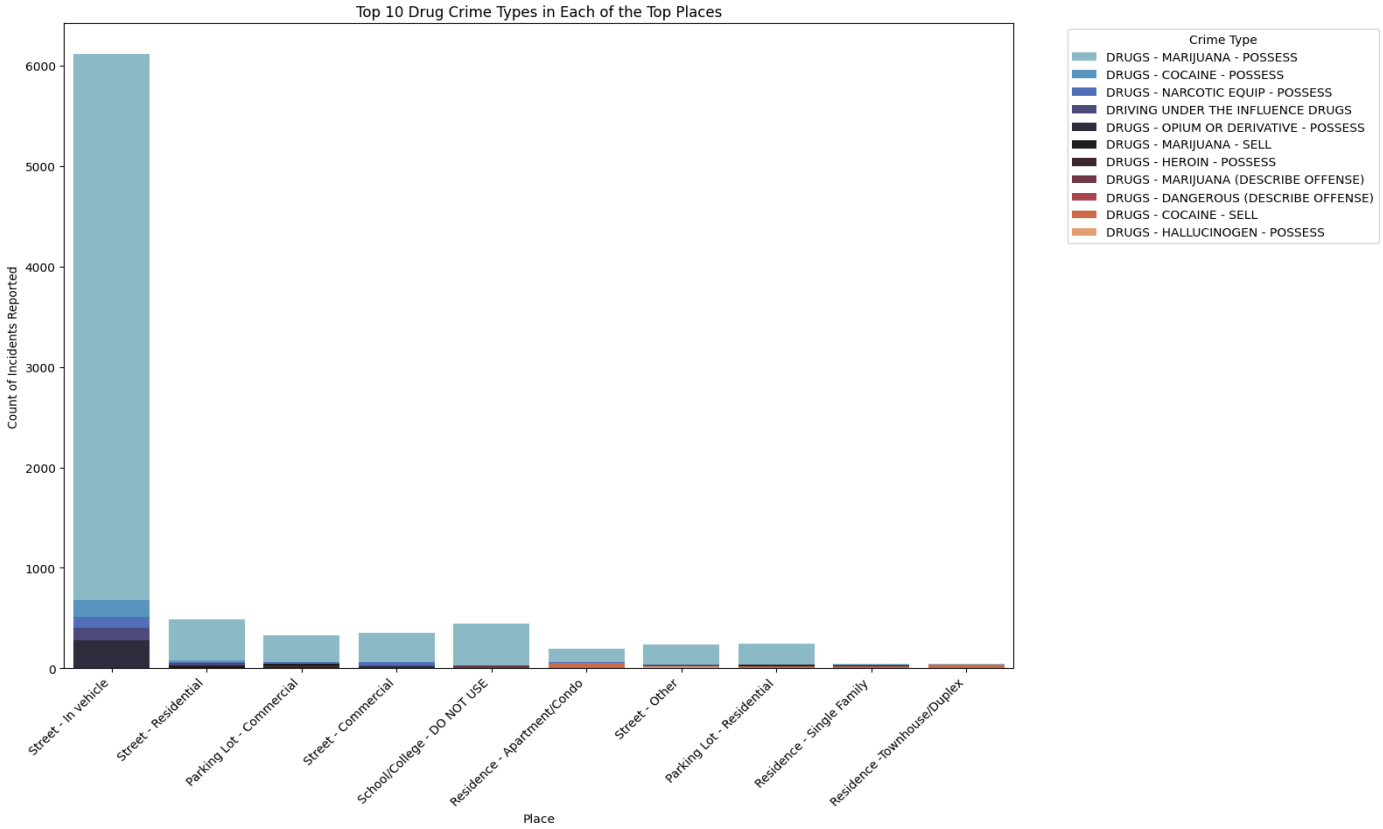

Figure 14: Stacked Bar Chart - Top 10 Drug Crime

Figure 15: Tree Map - Top 10 Drug Crime

INFERENCE:

The graphs illustrate the Incident count of the major places where the top cities where drugs were sold/bought/possessed. The possession of Marijuana and Cocaine are the major crimes reported mostly around streets (in vehicles), Streets in residential areas and Commercial parking lots.

6. How did the reforms and programs for the rising issue of illegal weapons in 2020 affect the crime rates related to weapons?

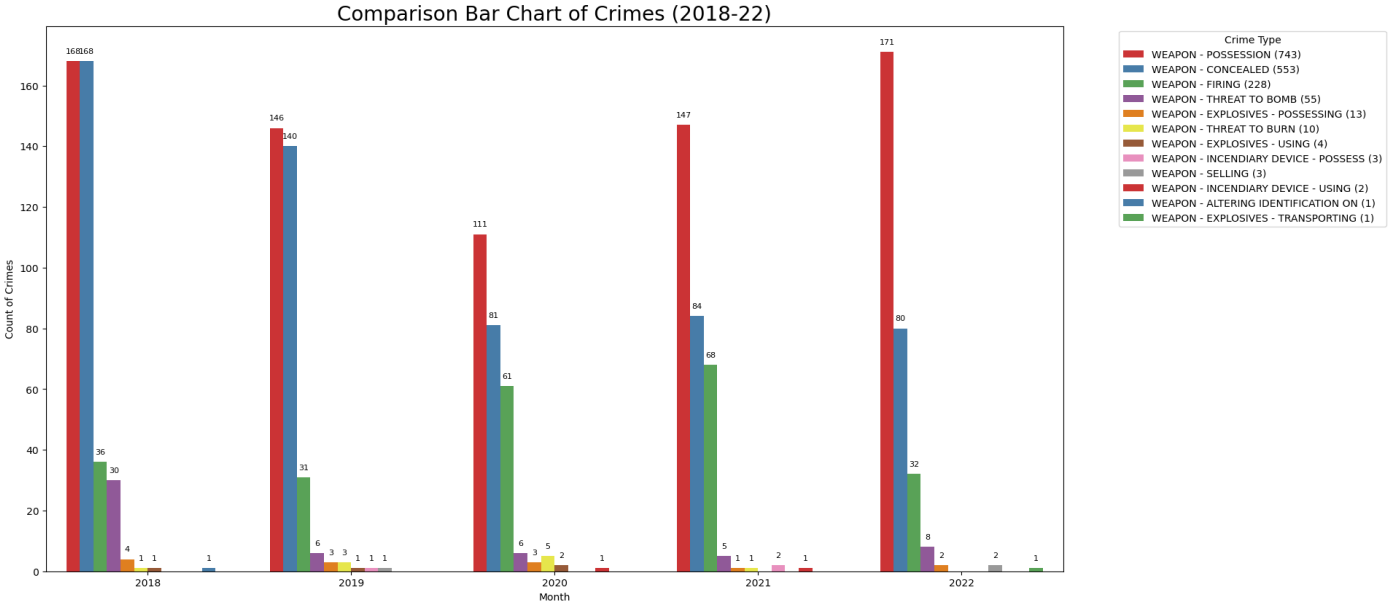

Figure 16: Comparative Bar Chart - Crime Rate related to Weapons

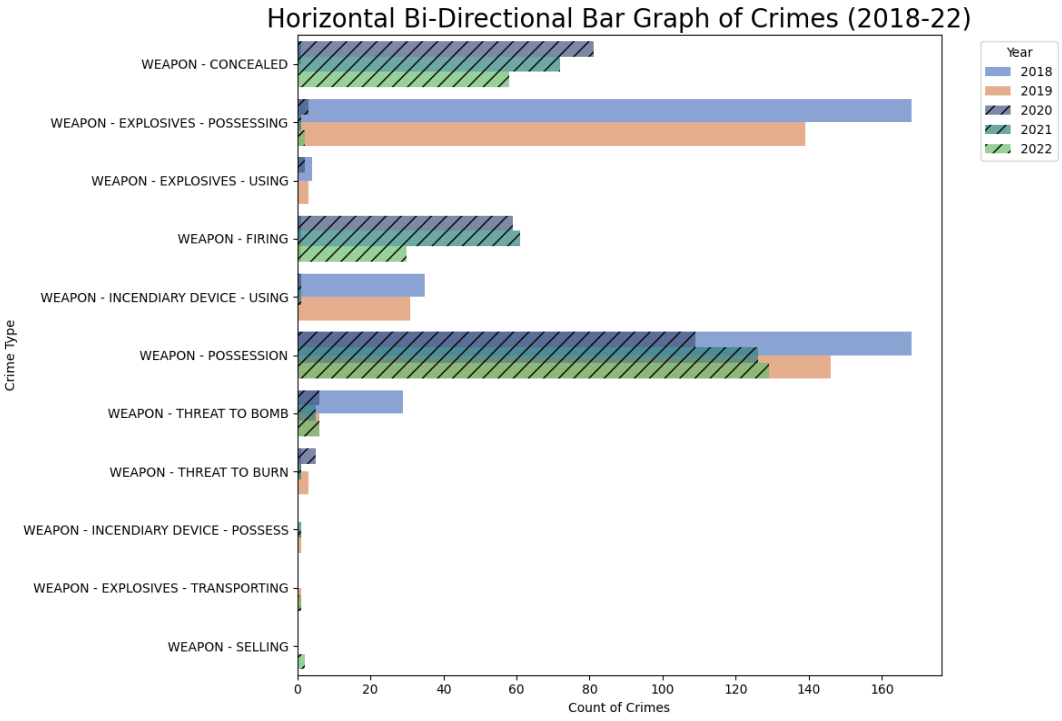

Figure 17: Horizontal Bi-Directional Chart - Crime Rate related to Weapons

INFERENCE:

In 2020, as per article, responding to an upsurge in public shootings, Montgomery County implemented reforms and legal measures against the use and carrying of weapons. Graphs

illustrate a significant decrease in weapon possession, concealed carry, and firing incidents in 2020. However, post-2020, the effectiveness of the intervention appears to wane, with rising rates of weapon possession and firing.

Q7. Which City where most people went missing?

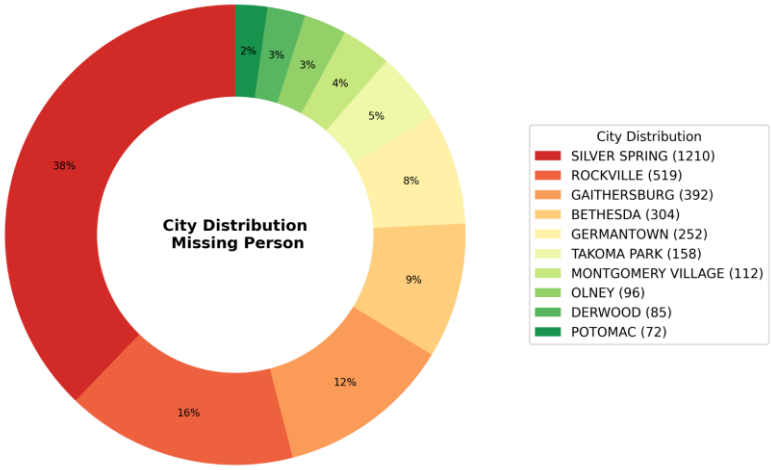

Figure 18: Donut Chart - Missing Persons

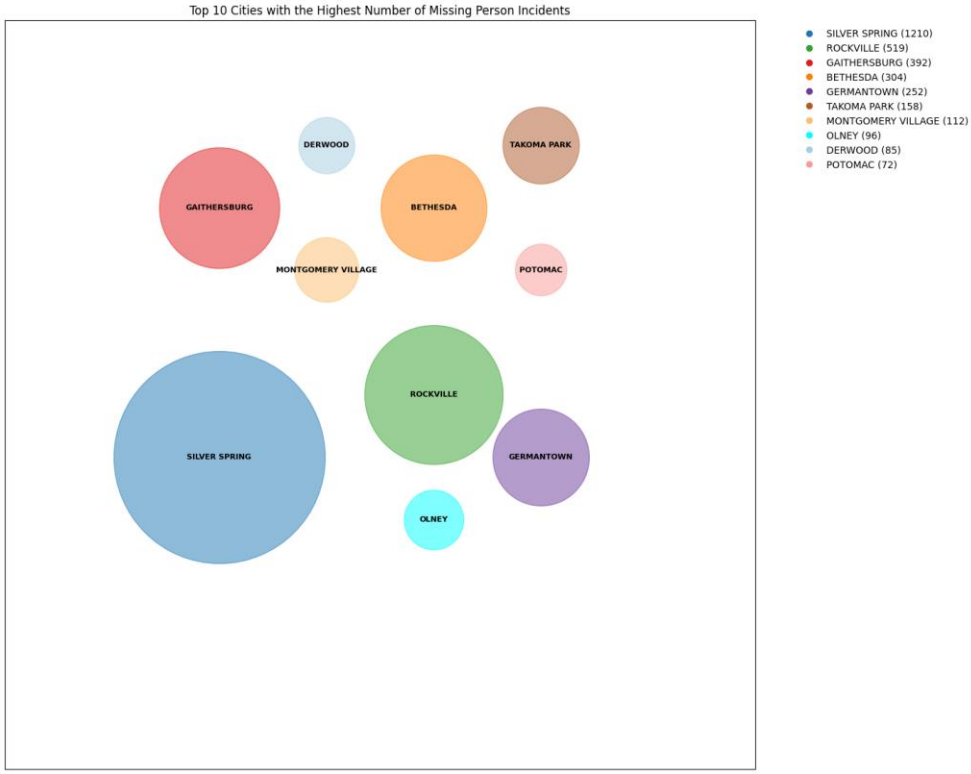

Figure 19: Bubble Chart - Missing Persons

INFERENCE:

From the above graphs, Silver Spring has the highest number of missing person cases, followed by Rockville and Gaithersburg

Q8. What are the seasonal trends in different fraud cases?

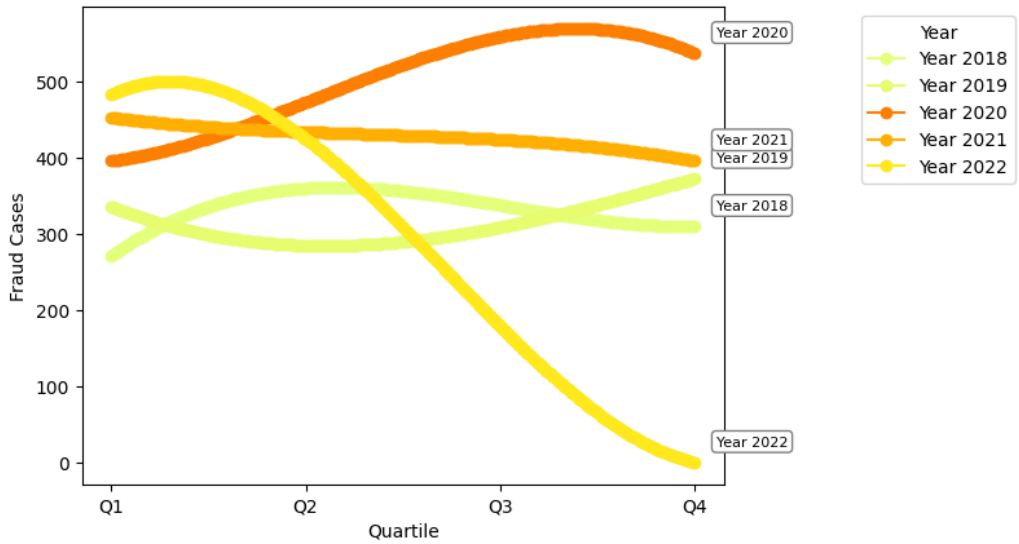

Figure 20: Spline Chart – Seasonal trends in fraud cases

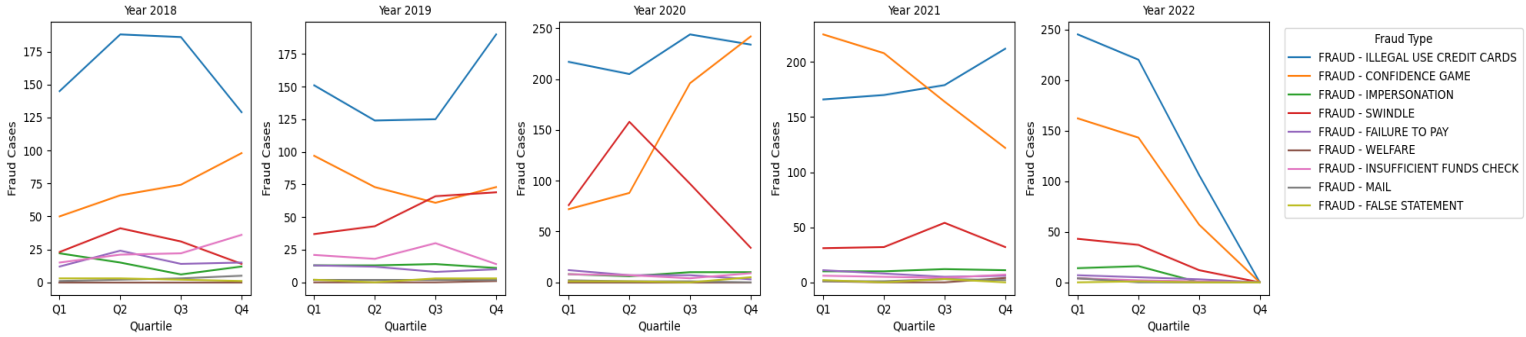

Figure 21: Line Chart – Seasonal trends in fraud cases

INFERENCE:

Illegal use of Credit cards and confidence game are major fraud types, occurring across all years.

Q9. Which hours are most likely linked to accident prone cases?

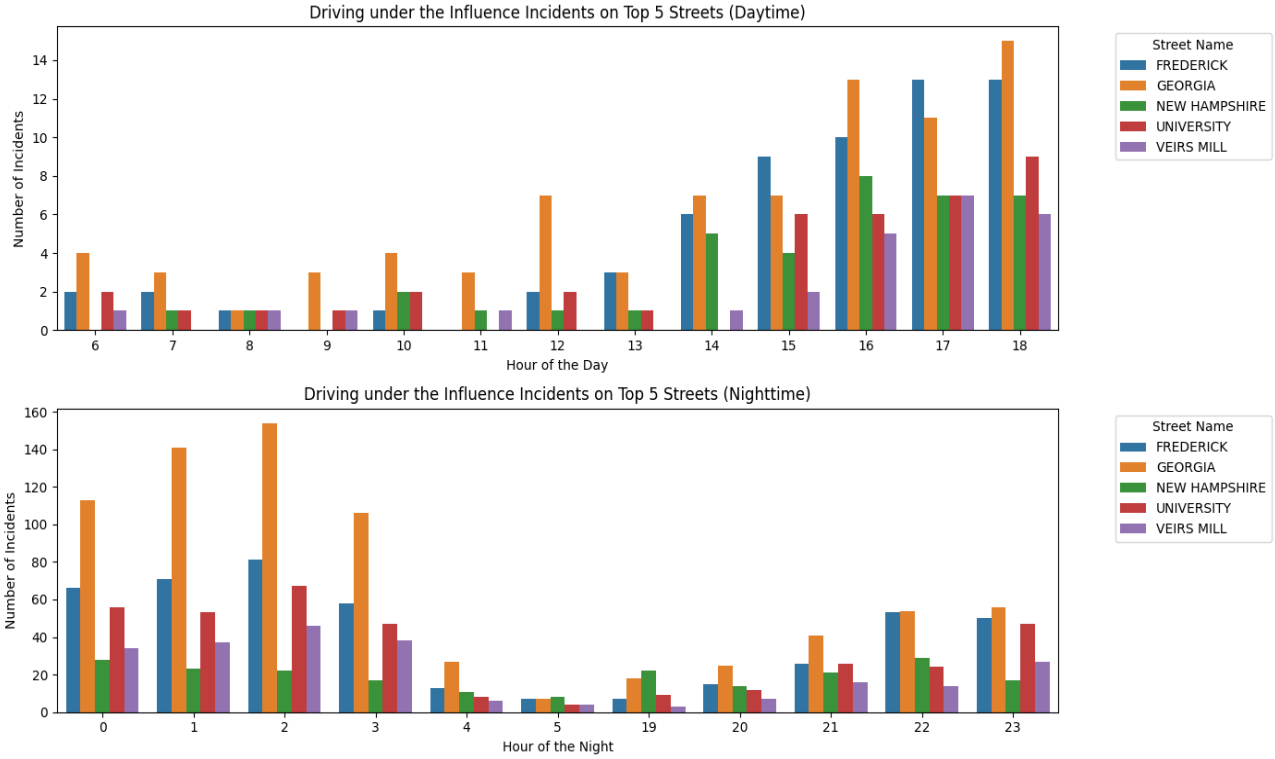

Figure 22: Grouped Bar Chart – Driving under the Influence

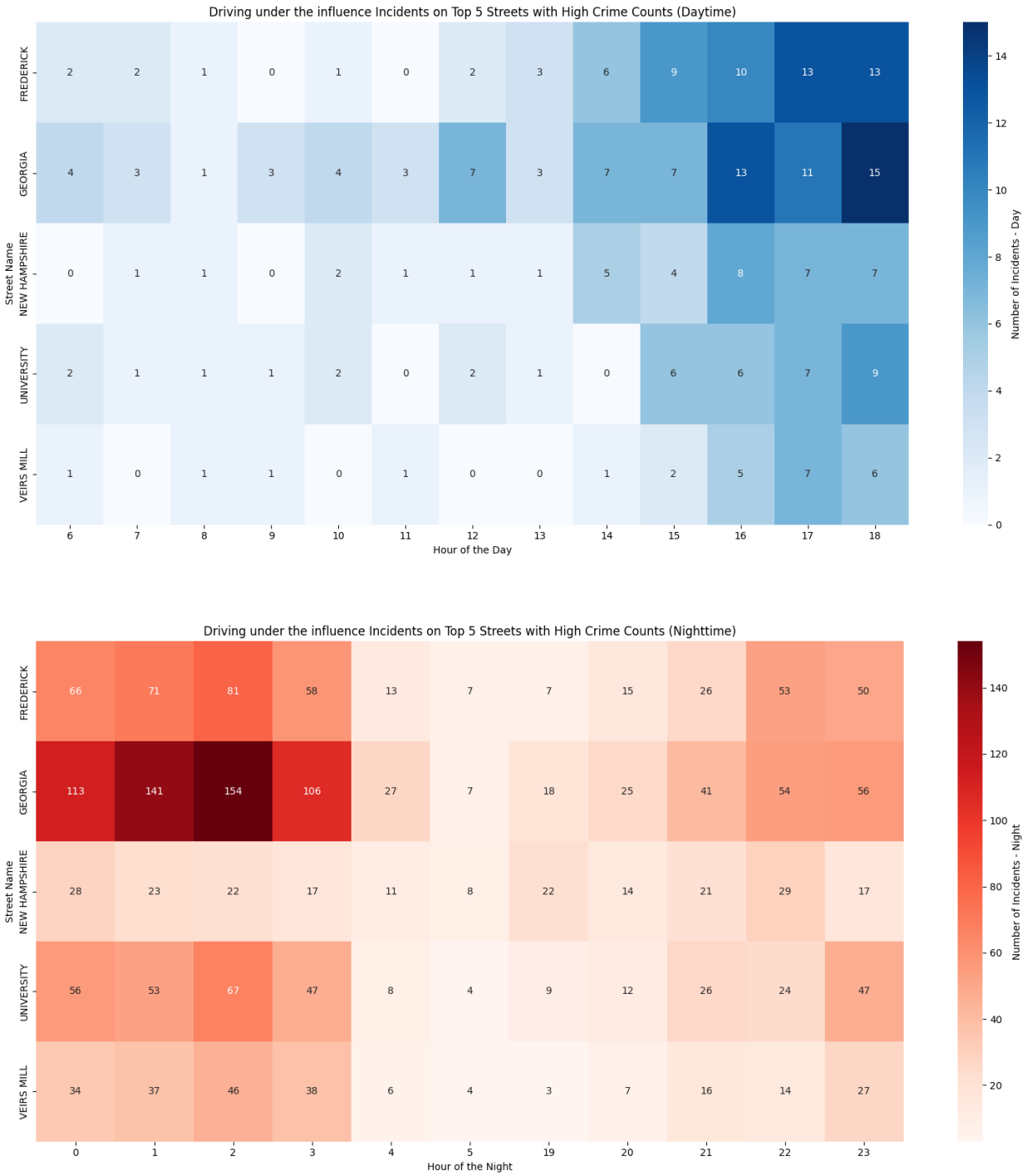

Figure 23: Heat Map – Driving under the Influence

INFERENCE:

Early morning hours and late evening hours reported the highest recorded incidents of driving under the influence, and they peaked for the streets of Georgia and Fredrick for both categories.

Q10. How does the frequency of larceny cases fluctuate over the week?

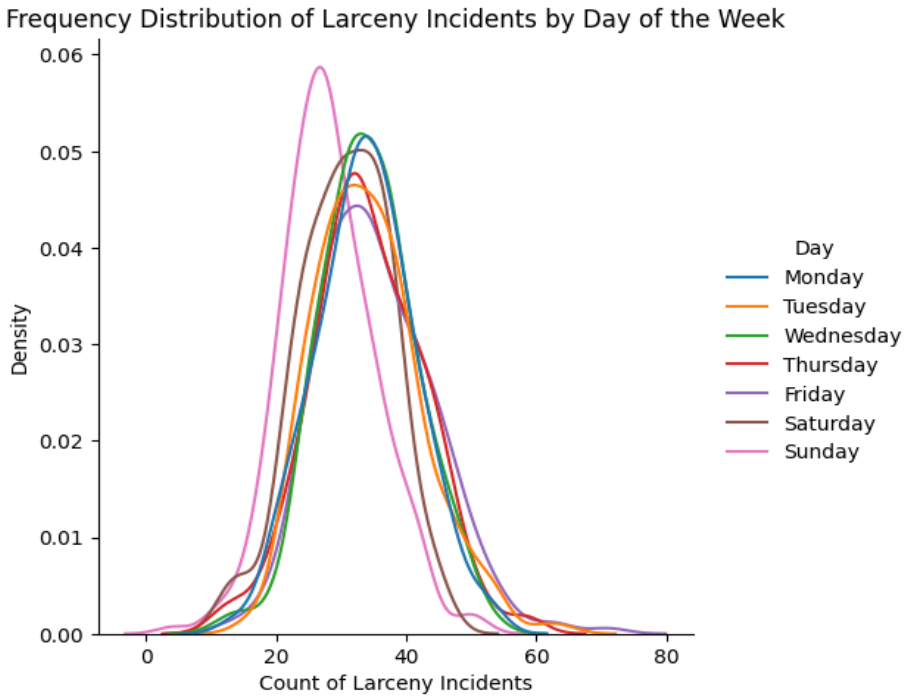

Figure 24: Displot – Frequency Distribution of Larceny Incidents

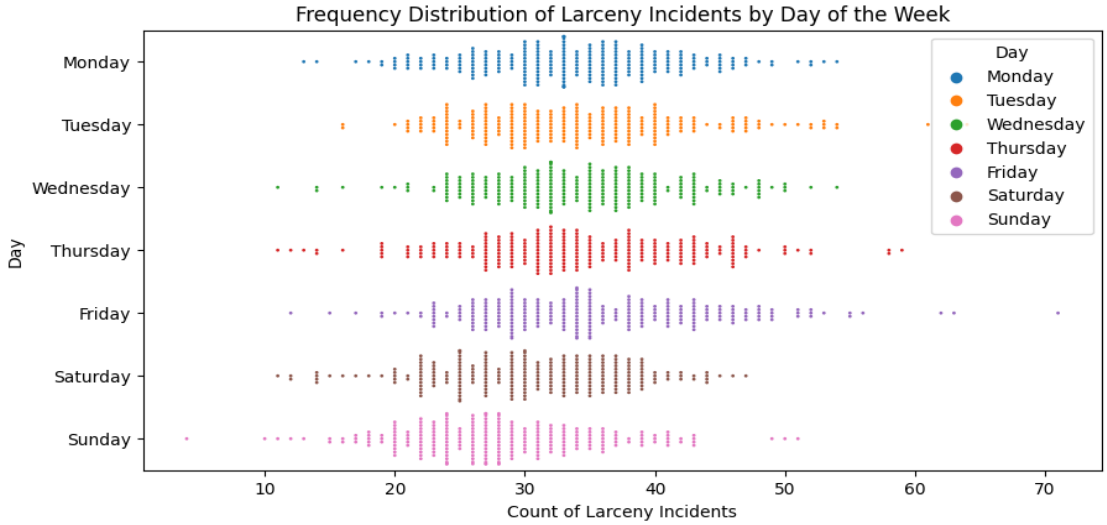

Figure 25: Swarm Plot- Frequency Distribution of Larceny Incidents

INFERENCE:

From the displot and swarm plot, it is evident that the majority of larceny incidents occurred during the weekdays rather than weekends. This observed trend could be because the routine of weekdays is more predictable. Weekdays offer a degree of predictability due to their regular schedule, which can be used for illegal activities.

7. Summary and Conclusion

This research compiles the IDA and EDA processes to ensure consistency and reliability of the data prior to the visualisation of the crime dataset and gaining insights such as crime patterns, hotspots, police response times, the impact of interventions, and specific crime trends over time. Based on the analysis and its insights, below are the suggested recommendations for crime reduction:

- Drug-related and missing person cases mostly occurred in vehicles on streets, residential areas or parking lots which can be reduced by implementing regular patrols, improving street surveillance, community members are encouraged to report suspicious activities and launching public awareness campaigns.
- Crime Rates related to Weapons can be addressed by the County authorities by considering implementing stricter legal actions, imposing higher penalties for possession and concealing, and enhancing community education on the associated consequences.
- Larceny during weekdays which is higher than weekends could be reduced by community policing, surveillance, public awareness, security measures, efficient law enforcement, legislation and youth programs.

8. References

[1] M. Montgomery County, "Crime: Open data portal," Crime | Open Data Portal, https://data.montgomerycountymd.gov/Public-Safety/Crime/icn6-v9z3 (accessed Nov. 21, 2023).
[2] P. Nabriya, "Exploratory data analysis and visualization techniques in data science," Analytics Vidhya, https://www.analyticsvidhya.com/blog/2021/08/exploratory-data-analysis-and-visualization-techniques-in-data-science/ (accessed Nov. 23, 2023).
[3] "Data preprocessing in Data Mining," GeeksforGeeks, https://www.geeksforgeeks.org/data-preprocessing-in-data-mining/ (accessed Nov. 23, 2023).
[4] "Data wrangling in python," GeeksforGeeks, https://www.geeksforgeeks.org/data-wrangling-in-python/ (accessed Nov. 23, 2023).
[5] R. Agrawal, "Exploratory data analysis using data visualization techniques!," Analytics Vidhya, https://www.analyticsvidhya.com/blog/2021/06/exploratory-data-analysis-using-data-visualization-techniques/ (accessed Nov. 24, 2023).
[6] "80 types of charts & graphs for Data Visualization (with examples)," Datylon, https://www.datylon.com/blog/types-of-charts-graphs-examples-data-visualization#waterfall-chart (accessed Nov. 24, 2023).
[7] Annual report 2020-2021 website File – Montgomerycountypa.gov, https://montgomerycountypa.gov/DocumentCenter/View/32469/Annual-Report-2020-2021_Website-File- (accessed Nov. 25, 2023).
[8] "Crime is up in 2023 compared to 2020," Annandale Today, https://annandaletoday.com/crime-is-up-in-2023-compared-to-2020/ (accessed Nov. 27, 2023).
[9] D. Morse, "Gun violence rises sharply in Montgomery County, police chief says," The Washington Post, https://www.washingtonpost.com/dc-md-va/2022/07/11/montgomery-county-gun-violence-double/ (accessed Nov. 30, 2023).
[10] Tableau. "Data Visualization: A Guide to Data Visualization Tools." Tableau, [Online]. Available: https://www.tableau.com/en-gb/learn/articles/data-visualization. [accessed Nov 26, 2023].
[11] KnowledgeHut. "A Comprehensive Guide to Exploratory Data Analysis (EDA) in Data Science." KnowledgeHut, [Online]. Available: https://www.knowledgehut.com/blog/data-science/eda-data-science. [accessed Nov 26, 2023].
[12] J. Weng, "Exploratory Data Analysis (EDA): A practical guide and template for Structured Data," Medium, https://towardsdatascience.com/exploratory-data-analysis-eda-a-practical-guide-and-template-for-structured-data-abfbf3ee3bd9 (accessed Nov. 30, 2023).
[13] FBI — Nibrs offense codes, https://ucr.fbi.gov/nibrs/2011/resources/nibrs-offense-codes (accessed Nov. 30, 2023).
[14] Uniform offense codes – city of Sacramento, https://www.cityofsacramento.org/-/media/Corporate/Files/Police/Crime/ucrcodes.pdf?la=en (accessed Nov. 30, 2023).

9. Supplementary Material

ADP_Coursework_1_Group_3.pdf