Submitted By Group X

Group Members and Participation

|  |  |  |
| --- | --- | --- |
| **Student Name** | **Student ID** | **Group Assignment Development Contribution Percentage(%)** |
| **Ansar Ahmad** | **K2403224** | **25%** |
| **Iryna Audzei** | **K24542246** | **25%** |
| **Seyda Tatlicioglu** | **K2451397** | **25%** |
| **Kulanika Gnanaratna** | **K2443219** | **25%** |

# **Abstract**

This coursework is based on a dataset of the USA’s crime statistics from Montgomery County. The initial data was derived from reported crimes classified according to the National Incident-Based Reporting System (NIBRS) of the Criminal Justice Information Services (CJIS) Division Uniform Crime Reporting (UCR) Program and documented by approved police incident reports between 2018 and 2022.

The initial dataset was wrangled to gain a comprehensive understanding of crime dynamics within Montgomery County by examining where, when, and what types of crimes occur most frequently, and which incidents have the greatest impact on victims and communities. This study analyses patterns across various dimensions, including location type, crime categories, and incident frequency over time. Key objectives include identifying high-risk areas, understanding peak times for certain crimes, examining changes over time and potentially suggesting the local government to mitigate crime rate in the county.

* The most notable change is the visible decrease in Drug/Narcotic Violations over year.
* Most crimes are reported in the afternoon and evening, with a notable decrease in criminal activity after midnight.
* The commonest substance for the “Drug/Narcotic Violations” crimes is marijuana and the majority of crimes for all substances are connected with possession.
* The probability of vehicle-related crimes in Montgomery County is overall relatively low (just above 0.12). Half of the street types show either insignificant or zero vehicle-related criminal activity.

Table of Contents

[Abstract 2](#_Toc149067225)

[1. Introduction 4](#_Toc149067226)

[2. Preliminary Data Analysis 4](#_Toc149067227)

[a. Dataset 4](#_Toc149067228)

[b. Data Quality Initial Assessment 4](#_Toc149067229)

[3. Exploratory Data Analysis 4](#_Toc149067230)

[a. Introduction to EDA 4](#_Toc149067231)

[b. Descriptive Statistics 4](#_Toc149067232)

[c. Data Visualisation 4](#_Toc149067233)

[4. Summary and Conclusion 4](#_Toc149067234)

[5. References 4](#_Toc149067235)

[6. Supplementary material 4](#_Toc149067236)

# Introduction

This report presents a comprehensive analysis of Montgomery County's crime data, focusing on critical stages of the data science process: Data Processing, Data Cleaning, Exploratory Data Analysis (EDA), and Data Visualization. These foundational areas transform raw data into accurate, structured, and insightful information that reveals underlying trends and relationships. By conducting a structured analysis of this dataset, we can uncover insights that contribute to public safety, resource allocation, and crime prevention strategies.

#### **Problem Statement**

The primary objective of this report is to achieve a clean, structured, and visually informative view of crime data in Montgomery County. By focusing on data processing, cleaning, EDA, and visualization, we aim to gain a comprehensive understanding of crime dynamics within Montgomery County by examining where, when, and what types of crimes occur most frequently, and which incidents have the greatest impact on victims and communities. This study will analyse patterns across various dimensions, including location type, crime categories, and incident frequency over time. Key objectives include identifying high-risk areas, understanding peak times for certain crimes, and examining changes over time.

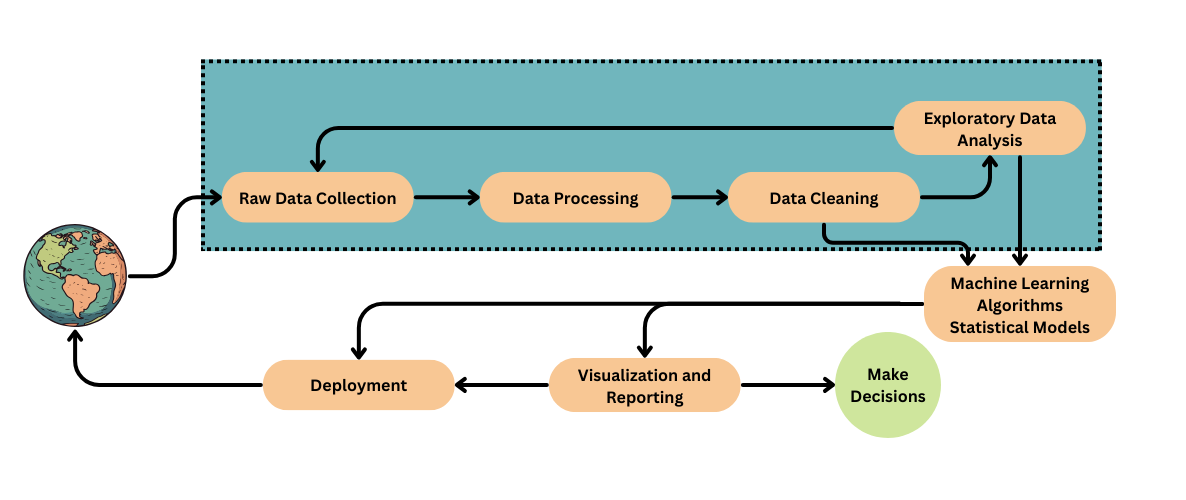
#### **Research Questions**

To comprehensively analyse the data, the following research questions will guide the study:

1. Which crime types are the highest in victim numbers?
2. What is the crime number pattern over the year?
3. What types of crime show the most meaningful change over the time?
4. How does the distribution of crime vary depending on the time of day?
5. How is the crime distributed according to the Crime Name 1 along the Police District Numbers and the cities?
6. What substance and what type of misuse are the commonest in the “Drug/Narcotic Violations” crimes?
7. Which street types/places are more prone to which types of crimes?
8. How are vehicle related crimes connected to the street types?
9. What type of crime frequently occurs in different geographic areas?
10. What geographic areas show the highest and lowest crime rate?

#### **Research Methodology and Expected Outcomes**

To achieve these objectives, the analysis will proceed through the following data science stages:



1. **Data Processing**
   1. Standardizing and organizing raw data for consistency, including reformatting columns and renaming variables.
2. **Data Cleaning**
   1. Addressing missing values, duplicates, and inconsistencies to ensure data accuracy and integrity.
3. **Exploratory Data Analysis (EDA)**
   1. Conducting preliminary analysis using descriptive statistics and visualizations to explore relationships and patterns within the data.
4. **Data Visualization**
   1. Developing informative visualizations, to provide a clear and interpretable overview of findings.

# Preliminary Data Analysis

# Dataset

The dataset includes reported crimes in Montgomery County, Maryland, USA, between 2018 and 2022. Available in CSV format, the dataset categorizes reported offenses following the National Incident-Based Reporting System (NIBRS) standards, which are part of the Uniform Crime Reporting (UCR) Program managed by the Criminal Justice Information Services (CJIS) Division. These crimes are documented through police incident reports that meet the required criteria (data.montgomerycountymd.gov, n.d.).

Data Montgomery, the county's open data platform, serves as the primary source for this dataset, providing the public with open access to crime statistics, including raw data and search capabilities. The reported crimes adhere to UCR guidelines, meaning multiple offenses can be recorded within a single incident, and each offense may involve more than one victim (data.montgomerycountymd.gov, n.d.).

The dataset comprises 306,094 rows, each containing 30 features. It also includes fields that may be either incomplete or contain erroneous data. A summary of the dataset’s features is presented in the table below.

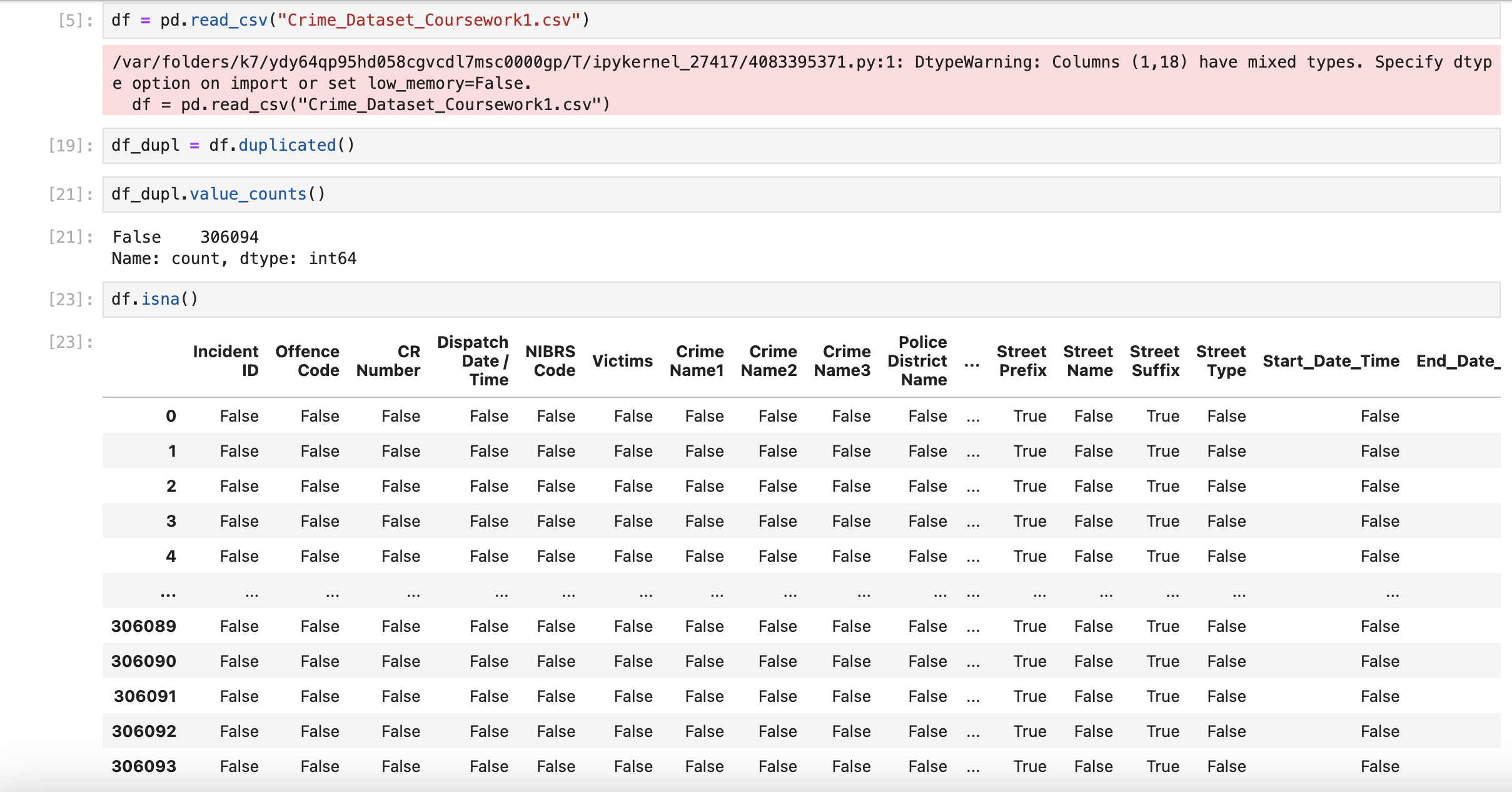
|  |  |  |  |
| --- | --- | --- | --- |
| **Column Name** | **Data Type** | **Description** | **Sample Values** |
| Incident ID | Integer | Police Incident Number | 201193163, 201202980 |
| Offence Code | String | Offense Code is the code for an offense committed inside the event, as established by the National Incident-Based Reporting System (NIBRS) of the CJIS Division  Uniform Crime Reporting (UCR) Program. | 5213, 2204 |
| CR Number | Integer | Police Report Number | 180042096, 16044222 |
| Dispatch Date / Time | String | The actual date and time an Officer was dispatched | 8/23/2018 9:52:08 PM,  12/18/2021 11:30:27 PM |
| NIBRS Code | String | FBI NIBRS codes | 35B, 90Z |
| Victims | Integer | Number of Victims | 12, 2 |
| Crime Name1 | String | Crime against Society/Person/Property or Other | Crime Against Society, Crime Against Person |
| Crime Name2 | String | Describes the NIBRS\_CODE | Motor Vehicle Theft, Simple Assault |
| Crime Name3 | String | Describes the OFFENCE\_CODE | WEAPON - FIRING, FORGERY OF CHECKS |
| Police District Name | String | Name of District (Rockville, Wheaton, etc.) | GERMANTOWN, BETHESDA |
| Block Address | String | Address in 100 block level | 12800 BLK MIDDLEBROOK RD, 3900 BLK BEL PRE RD |
| City | String | City | SILVER SPRING, GAITHERSBURG |
| State | String | State | MD, DC |
| Zip Code | Float | Zip code | 20874, 20902 |
| Agency | String | Assigned Police Department | MCPD, TPPD |
| Place | String | Place description | Residence -  Single Family, Street – In Vehicle |
| Sector | String | Police sector name, a subset of District | N, E |
| Beat | String | Police patrol area, a subset of Sector | 5N1, 4L1 |
| PRA | String | Police response area, a subset of Beat | 447, 701 |
| Address Number | Float | House or Business Number | 12800, 700 |
| Street Prefix | String | North, South, East, West | N, W |
| Street Name | String | Street name | RANDOLPH, GLENMONT |
| Street Suffix | String | Quadrant (NW, SW, NE, SW, etc.) | S, NW |
| Street Type | String | Type of street (Ave, Drive, Highway, etc.) | RD, AVE |
| Start\_Date\_Time | String | Occurred from date/time | 6/14/2018 10:26:00 PM, 04/20/2019 06:00:00 PM |
| End\_Date\_Time | String | Occurred to date/time | 03/09/2018 05:30:00  PM,04/21/2019 02:00:00 PM |
| Latitude | Float | Latitude | 38.9927, 39.148 |
| Longitude | Float | Longitude | -77.0971, -77.2182 |
| Police District Number | String | Major Police Boundary | 2D, 4D |
| Location | String | Location | (38.9927, -77.0971), (39.148, -77.2182) |

(Adapted from Source: data.montgomerycountymd.gov, n.d.)

# Data Quality Initial Assessment

To prepare the dataset for visualization, an initial assessment was conducted to determine the necessary cleaning steps, including identifying any missing values, outliers, duplicates, or other data quality issues that required attention.

Upon creating a data frame, it was observed that while the dataset contained no duplicates, some values were missing (Figure 1.1).



The assessment revealed that the “End\_Date\_Time” column had the highest proportion of missing values, exceeding 50%, which would prevent it from providing a statistically representative sample (Figure 1.2). Consequently, this metric, indicating crime duration, was excluded from the formulation of any research questions.

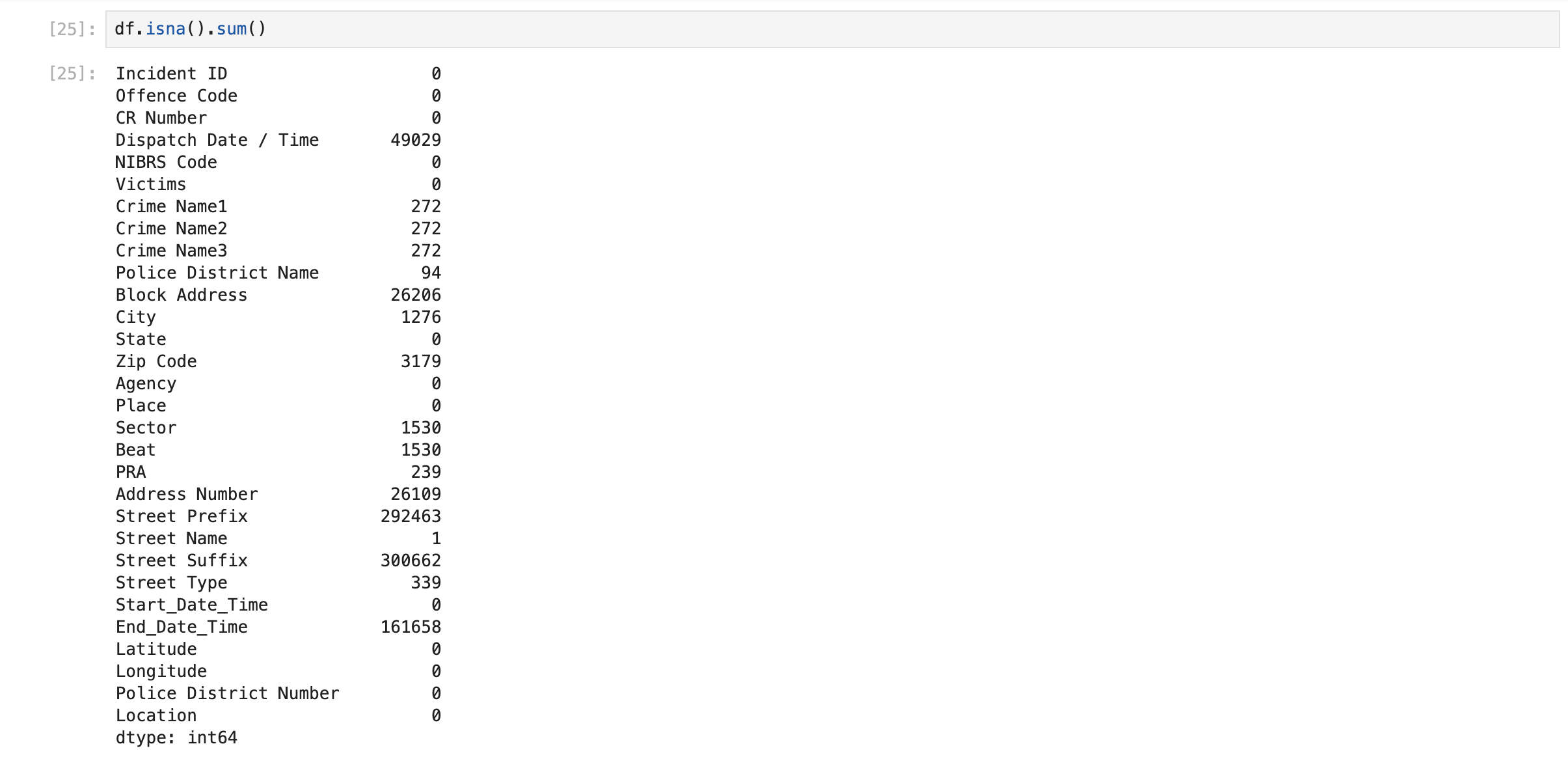
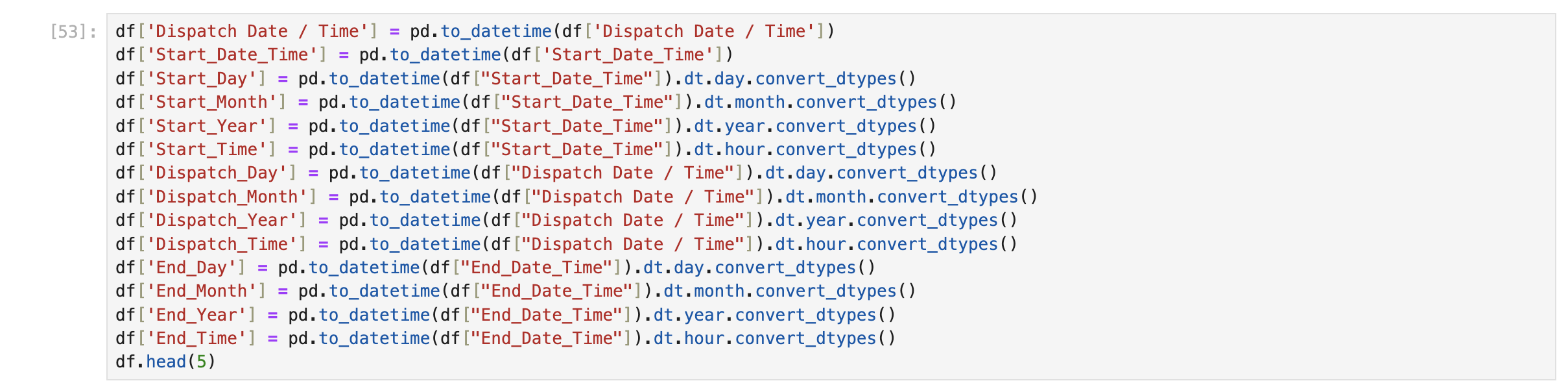


Figure 1.2

It was also observed that the date format was inconsistent across entries. To address this issue and to facilitate further visualization by separating day, month, and year components, several .to\_datetime() commands were applied (Figure 1.3).

Figure 1.3

McKinney (2022) notes that “all of the descriptive statistics on pandas objects exclude missing data by default” (p. 203). However, our goal was to retain as much data as possible. Given that the number of rows with missing crime names was statistically insignificant, these entries were filled with the value “Other” using the .fillna() function (Figure 1.4)

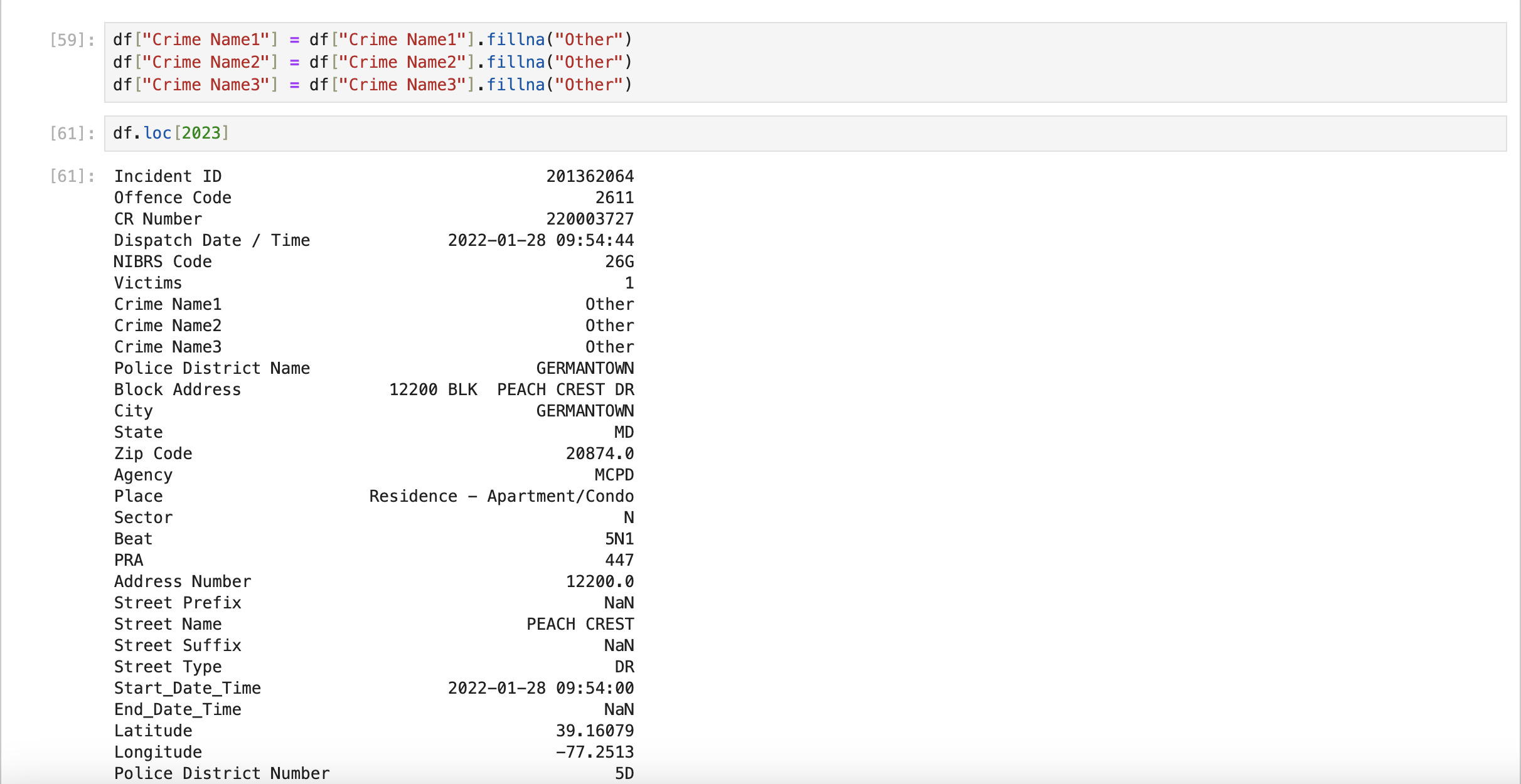


Figure 1.4

The assumption made to explain conflicting city and state entries is that these discrepancies arise due to the nature of interstate crimes. Such crimes often involve multiple jurisdictions, with different aspects—such as the crime itself, arrests, or incidents involving suspects and victims—occurring in various states. Local law enforcement agencies document the crime based on the jurisdiction where the specific aspect was reported. For example, an arrest in a state different from where the crime occurred would still be recorded under the arresting jurisdiction. This aligns with NIBRS reporting standards, where Group A offense codes capture incidents within a jurisdiction, and Group B codes record arrest details.

Finally, it also came to our attention that some values of the State column were state numbers rather than abbreviations, which we rectified by writing the replace\_no\_with\_abb function (Figure 1.5).

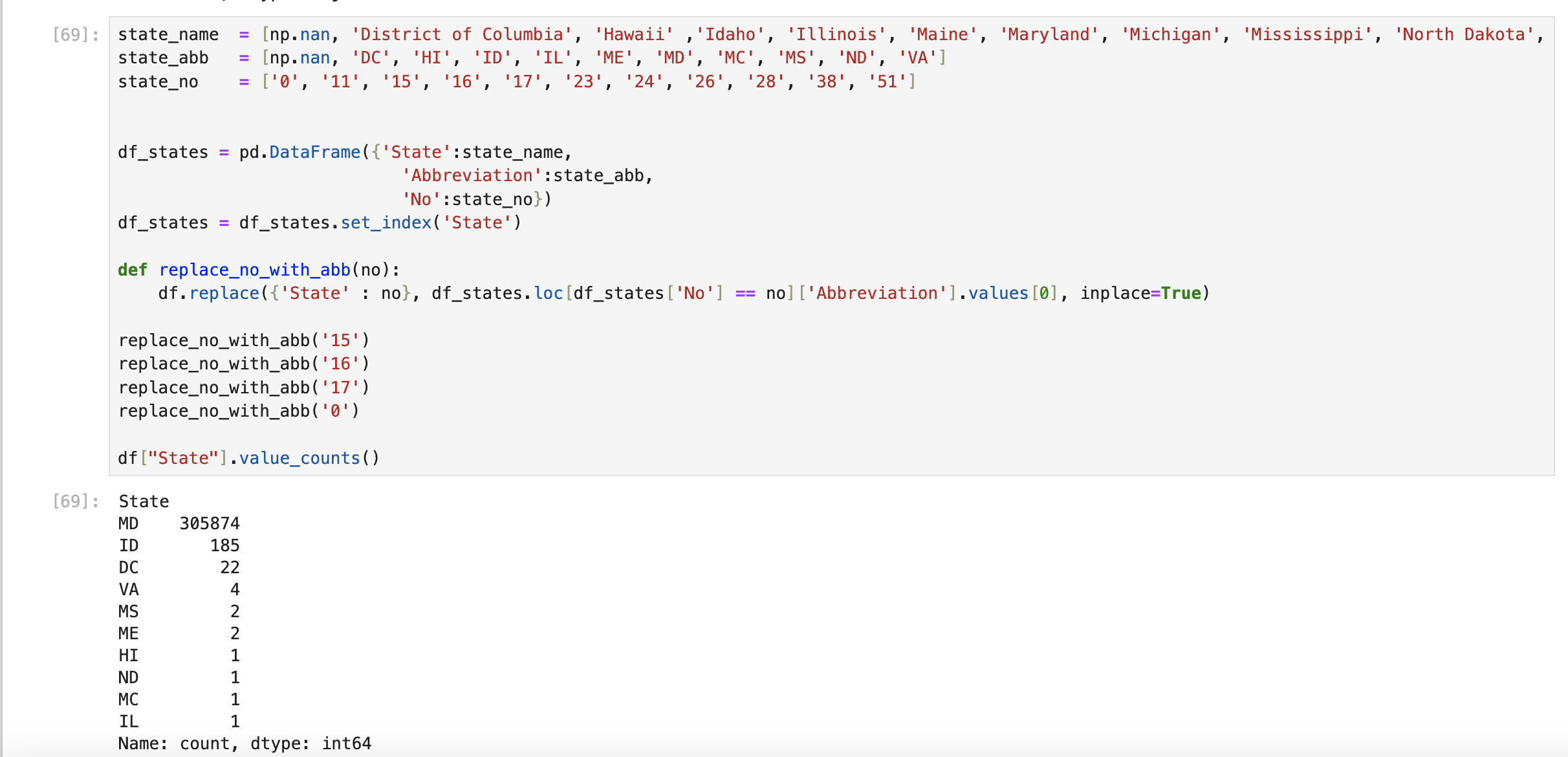


Figure 1.5

Following the above data manipulations, the dataset was now prepared for more in-depth exploration and modelling.

# **Exploratory Data Analysis**

# **Introduction to EDA**

Exploratory Data Analysis (EDA) is a very significant part of data pre-processing in the field of Data Analysis which can provide users with an outlook of:

* Contents of the data
* relationship between different parameters/ variables of the data
* identify errors
* anomalies

EDA helps to conclude whether the methods used by the user are appropriate for analysing the data or not.

There are four primary types of EDA as follows [<https://www.ibm.com/topics/exploratory-data-analysis>]:

1. **Univariate non-graphical**: In this technique we use different methods such as mean, median and mode to determine the classification of the data as the data is in uni-variable. The data does not have more than one parameters therefore such techniques are used to deduce a result from the data.
2. **Univariate graphical**: We use plotting on graphs to deduce a result out of the data in this as we have a plottable number of figures in the data. With plotting , the data becomes self-explanatory and this method is useful to present. Some of the common univariate graphics include:

* Stem and Leaf plots
* Histograms
* Box Plots (it reflects the minimum, first quartile, median, third quartile, and maximum values)

1. **Multivariate non-graphical**: Multivariate data has more than one variable and more than one dimension. This method is used to show a relation between the parameters and the dimensions of the data which can be presented.   
   In this method we do not plot a graph but we do make a table which shows a relationship between the variables when examined using cross table values i.e. the values on top of the table and the values in the index column cross to the value which fits both parameters.
2. **Multivariate graphical**:In this method when we have complex valued parameters, we plot a graph to make it presentable. We can use many techniques to show a relationship between the axis on the graph. Different plots show different relationships such as correlation and common value frequency of the data.  
   Multivariate graphics include:

* Scatter plot
* Multivariate Chart
* Run Chart
* Heat Map

# **Descriptive Statistics**

Descriptive statistics provides a summary and description of the main features of the dataset. Various methods are available for descriptive statistics, including measures of central tendency (mean, median, mode), measures of spread (range, variance, standard deviation, interquartile range and frequency distribution) (Bowers, 1996).

Three types of analysis take a part of data analysis. The first is “Univariant Analysis” which examines one variable at a time by looking at key characteristics like mean, median, mode, distribution, etc. The second type is “Bivariate Analysis” which looks at relationships between two variables, using correlation and covariance. And last one is “Multivariate Analysis” which examines relationships among more than two variables. (Bhattacherjee, 2012).

* Univariant Analysis

Measures of Central Tendency

Mean: It gives the average of values. The average number of victims per incident is approximately 1.02.

Median: It gives the middle value after ordering the data. The median number of victims per incident is 1, indicating that most incidents involve a single victim.

Mode: The most frequently occurring value in a dataset. The most common value for Victims is 1.

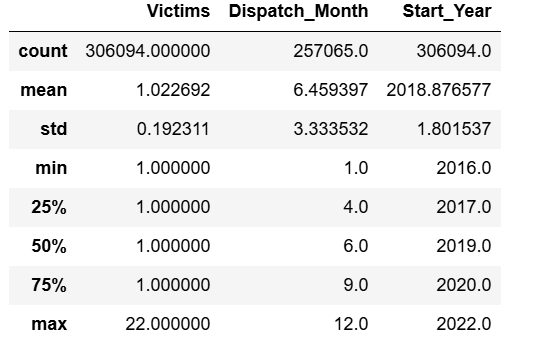
Measures of Spread

Range: It shows the spread of the data. It is calculated by differences between maximum and minimum values. The range of victims per incident is 21 by subtracting maximum and minimum values.

Variance and Standard Deviation: Variance shows how data points differ from the mean, while the standard deviation (square root of variance) expresses this spread in the dataset. The standard deviation for Victims is approximately 0.19.

* Bivariate Analysis

Correlation: It shows the relationship between the variables. For instance, analysing the relationship between Drug victims and Theft victims.



# **Data Visualisation**

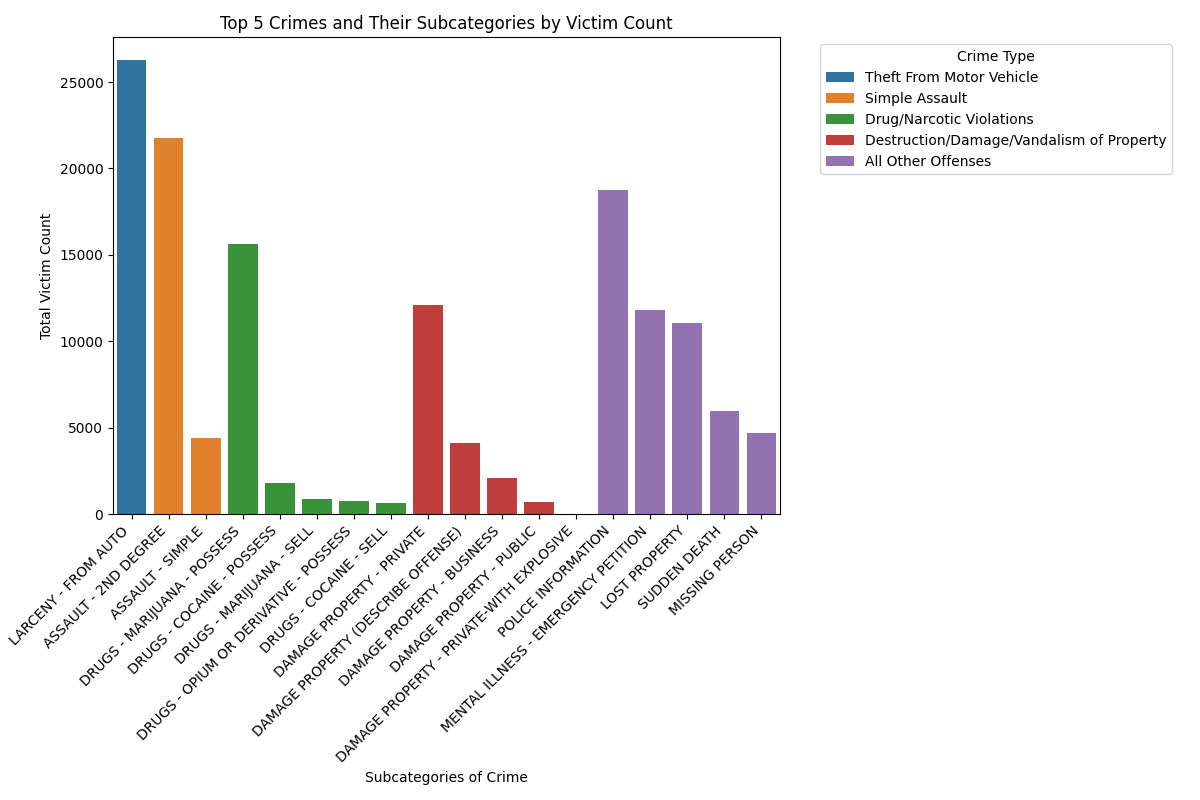
As Grus states: ‘’Although it is very easy to create visualizations, it’s much harder to produce *good* ones”. (Grus, 2019, p. 43). In this section, various visualizations are used to explore key patterns and insights within the dataset. Visual analysis provides an intuitive way to interpret complex data, uncovering trends, distributions, and relationships that inform our understanding of crime activity. These visualizations collectively aim to answer the research questions and reveal actionable insights that might support crime prevention and resource planning. Given below are some visualisation categories based on the analysis types (Hess, 2022):

|  |  |  |  |
| --- | --- | --- | --- |
| Comparison | Correlation | Distribution | Geospatial |
| Column chart  Bar chart  Line chart  Pie chart | Scatter chart  Heatmap  Bubble chart | Histogram  Scatter chart  Box chart | Geographic heatmap  Choropleth map |

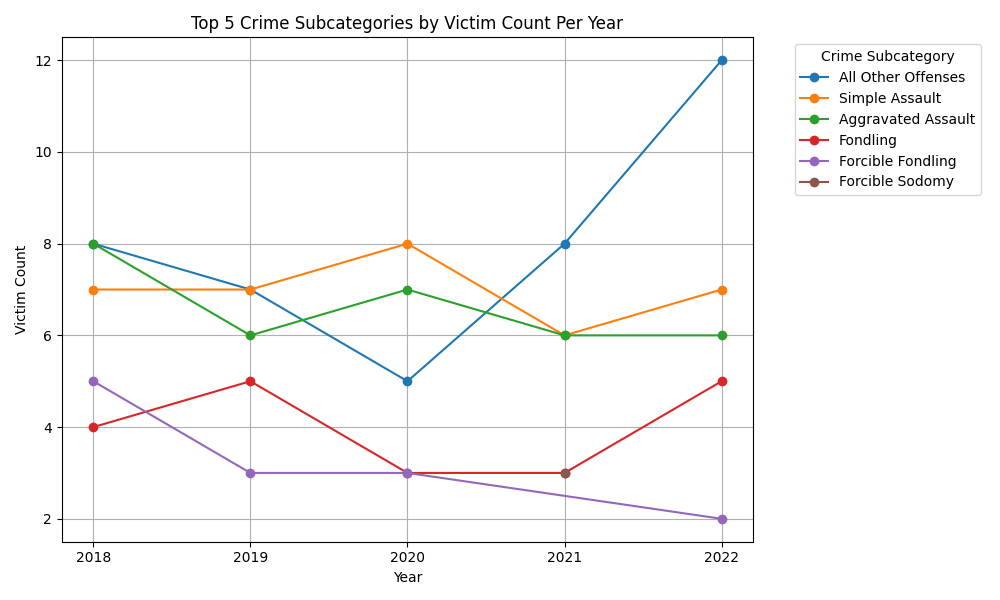
Research Questions

Q1: Which crime types are the highest in victim numbers?

Analysing crime types with the highest victim counts over a given period is crucial for understanding evolving patterns and identifying areas of significant impact on the community. For this purpose, two charts were analysed to better understand the impact of high-victim-count crime categories.

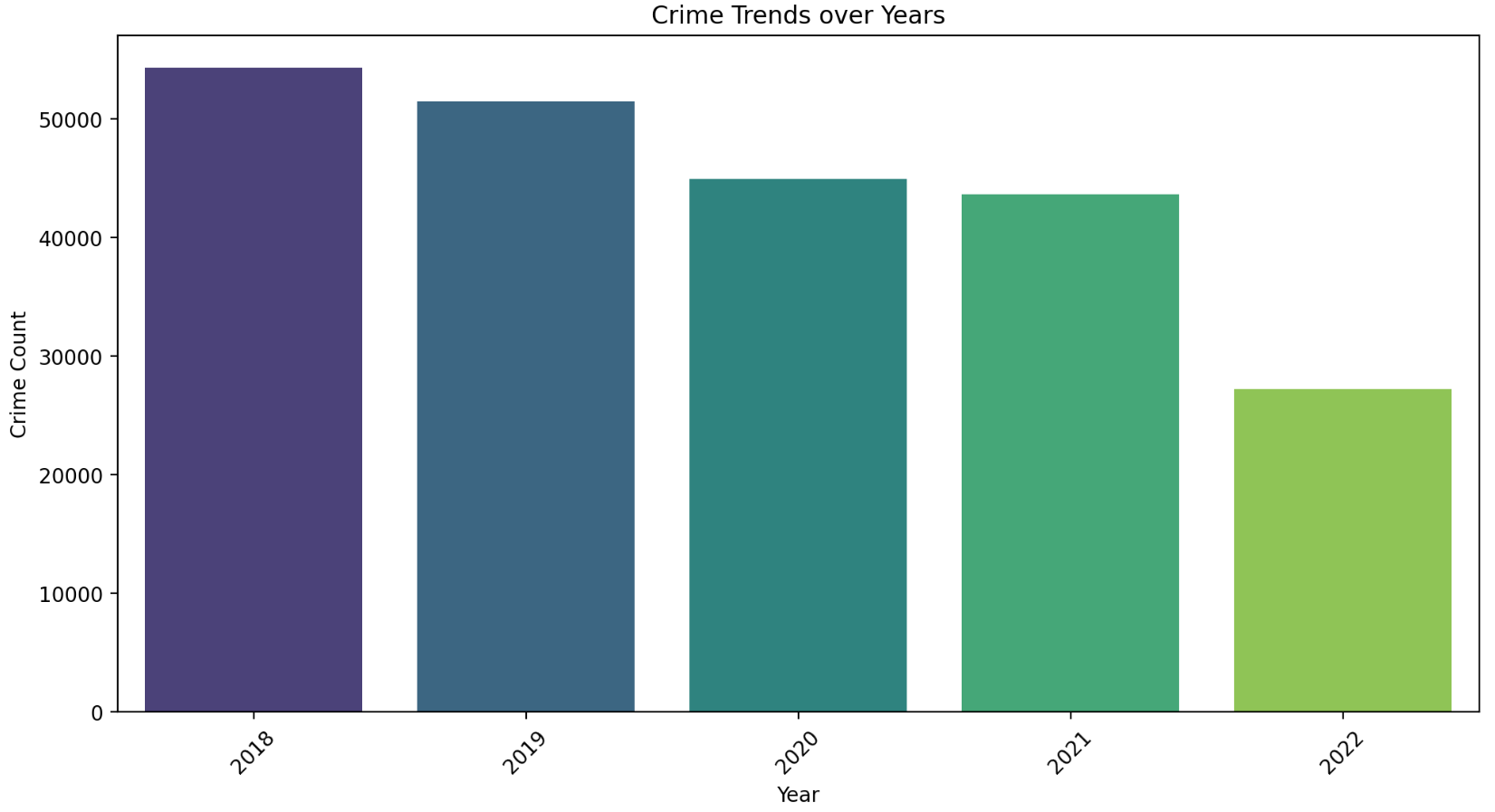


The bar chart allows for a clear comparison of victim counts within each subcategory, making it easy to identify the relative impact of each crime type.

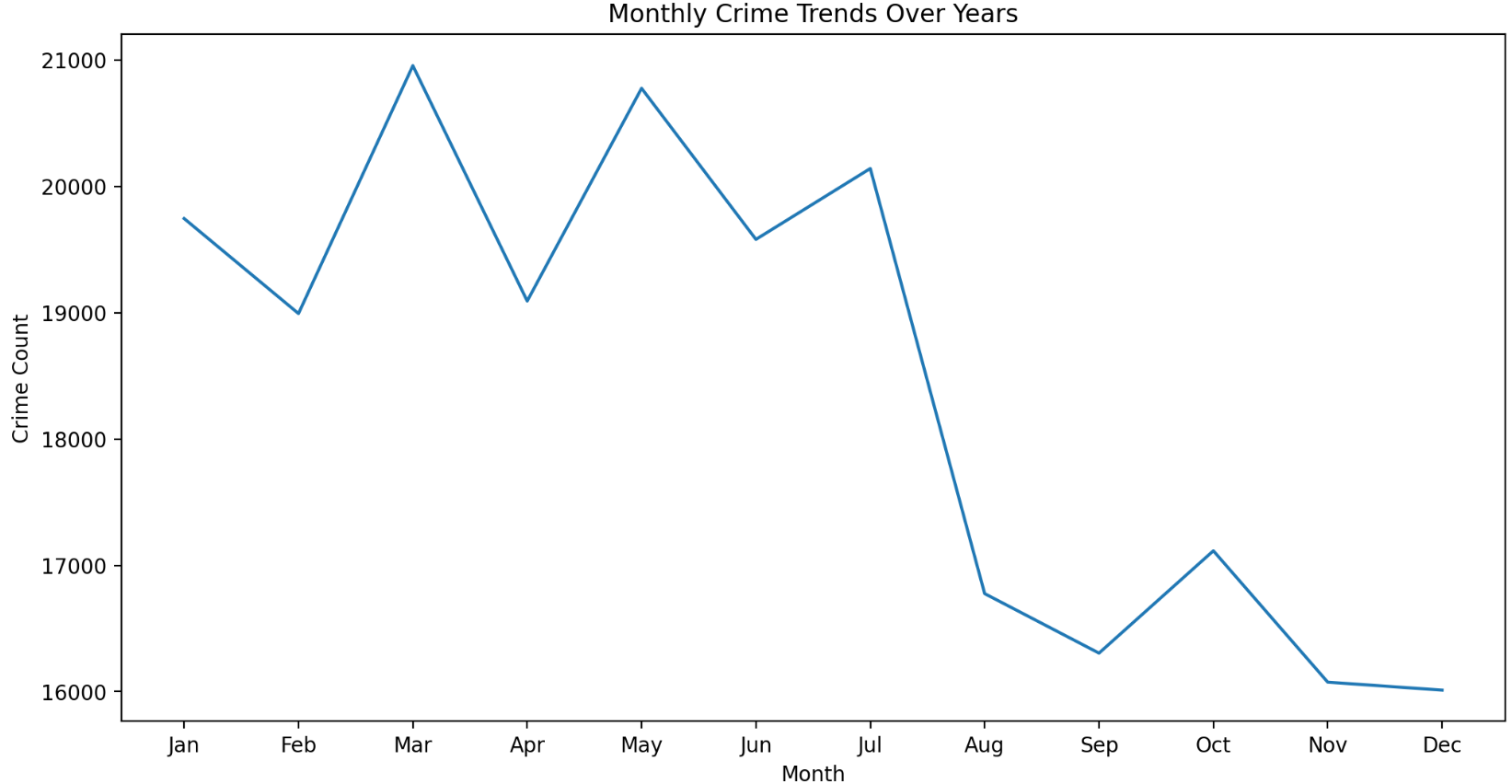
The line chart depicts clear visualization of multiple categories with distinct lines to enhance the understanding of how each subcategory evolves year-to-year, allowing for the identification of fluctuations and patterns, while the bar chart clearly compares the total victim count across subcategories, highlighting the relative impact of each crime type.

Q2: What is the crime number pattern over the year?

Analysis of the "Start\_Year" column reveals that crimes have been recorded consistently from 2018 through 2022. To better understand of crime trends by year, following bar chart was created, as bar charts highly effective for making comparison.

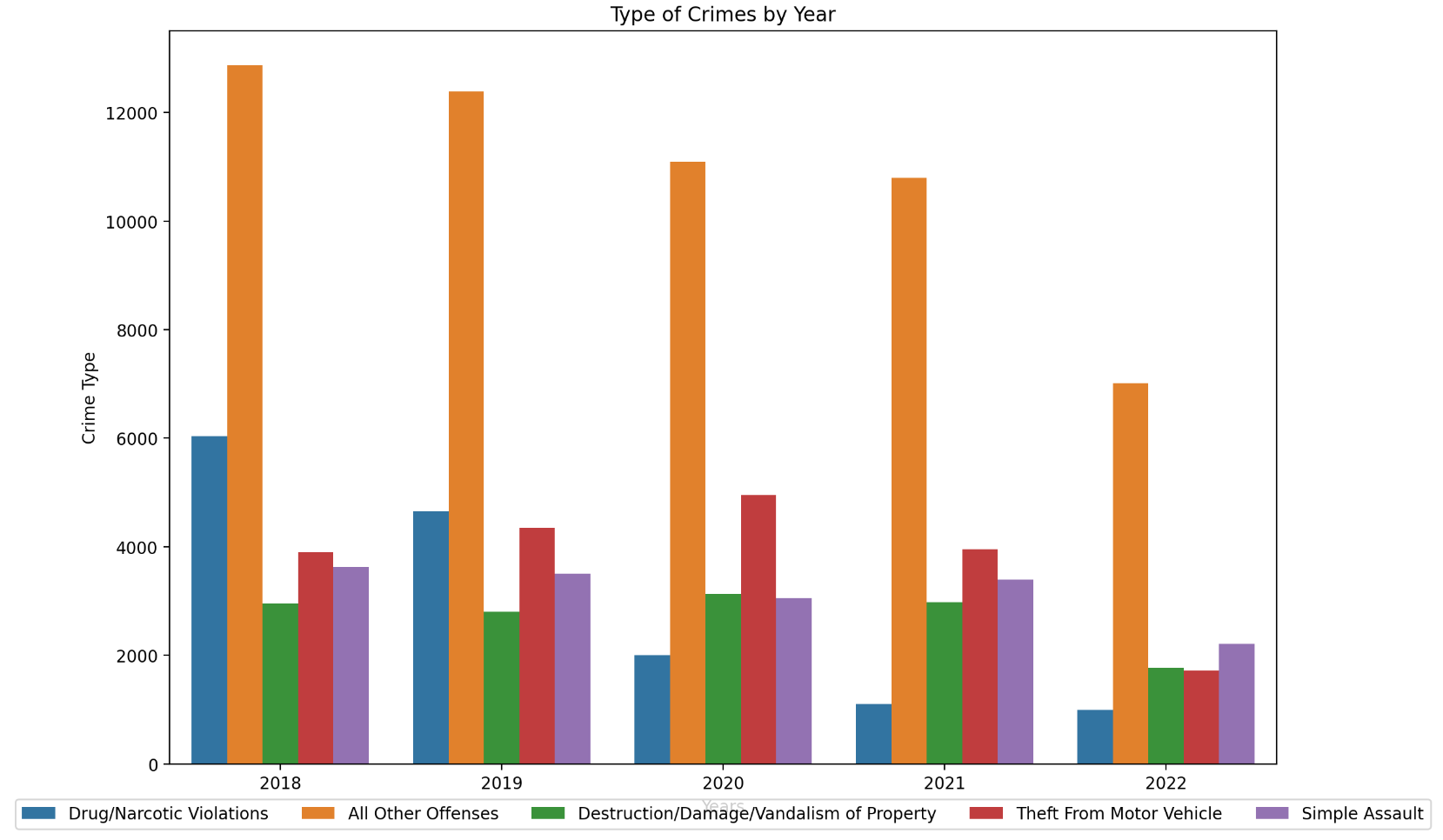


Line charts are often used to represent time series. To emphasize monthly crime changes over the years the line chart below was created. (Between the years 2018 - 2022 )

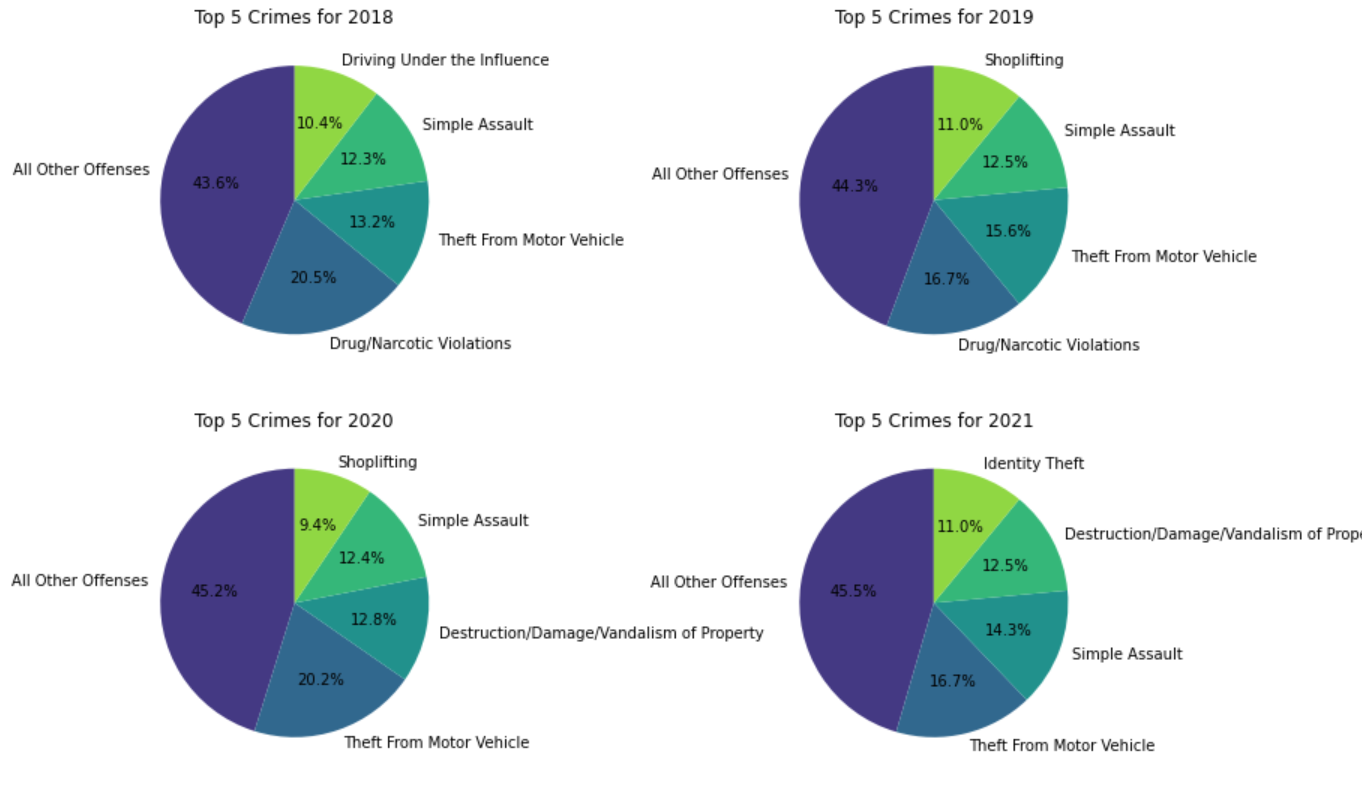


Q3: What types of crime show the most meaningful change over the time?

For this purpose, the subcategory column 'Crime Name 2' was examined to analyse the crime types by year. The bar chart below shows the ‘Crime Name2’ subcategory change over time.

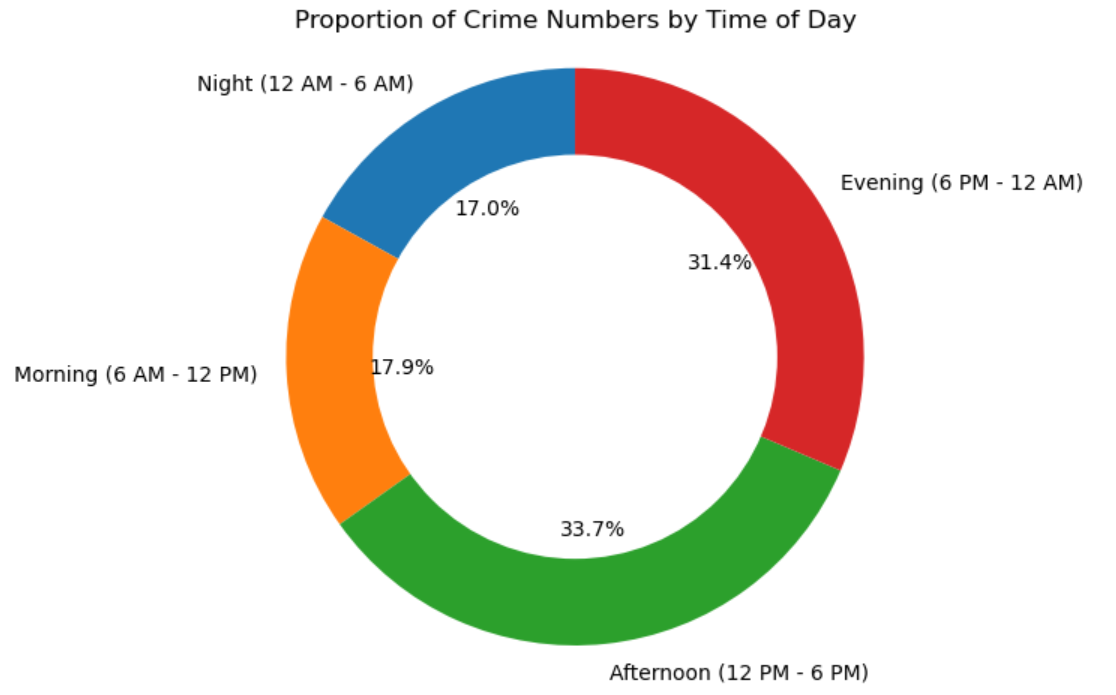


Following pie chart was used to show proportion of crime types based on each year. This visualization helps present the information in a clear and easy-to-read format.

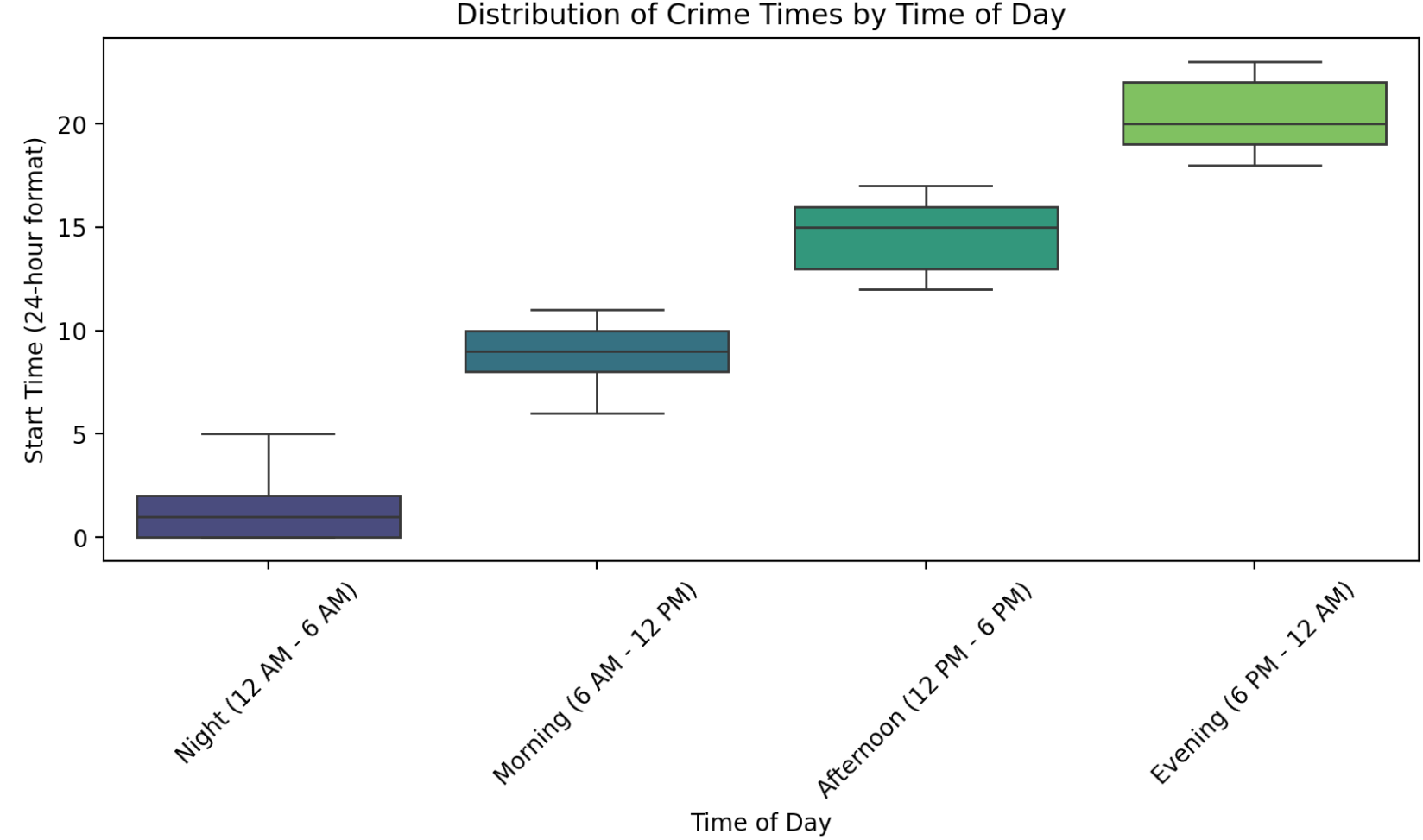


Q4: How does the distribution of crime vary depending on the time of day?

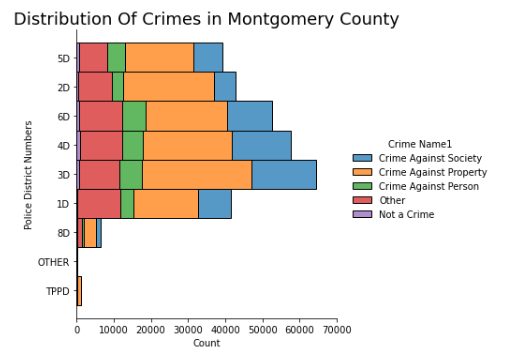
The pie chart shows the comparison of crimes based on the time of day, time-based bins are created and divided into four periods: Night, Morning, Afternoon, and Evening.



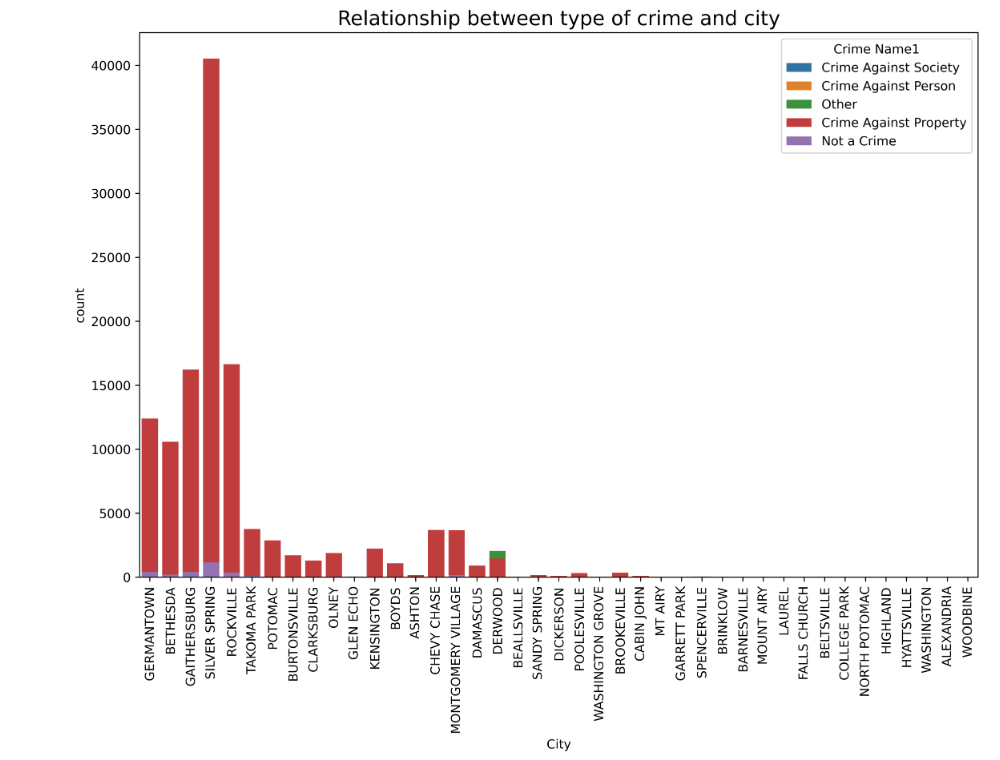
The following boxplot chart has been created to show the distribution of crime start times categorized by "Time of Day." Each category ("Morning," "Afternoon," etc.) presents a different period, with the boxplot revealing the range and median of crime incidents within those times.



Q5: How is the crime distributed according to the Crime Name 1 along the Police District Numbers and the cities?



In this question we wanted to find out that how the crime (crime name 1) count is distributed along the police districts in which the department has divided the county into. The plot shows that the crime against property is the major and most common type of crime in all police districts. The crime is highest in the Police district 3D as compared to other police districts but the crime against property is almost constant along all the police districts. Police district 8D has the lowest crime count which can be due to many reasons which include the area that the police district 8D covers. Police district 3 and 4 show the highest crime against the society as compared to the other police districts.

The second figure shows how the frequency of the crime has been different across the different cities in the area. It has been plotted as according to the Crime Name 1. The crime against property is a common major issue in all the cities in which the crime is reported the most. The plot also shows that the city of Silver Spring has recorded the most crime count than any other city in area.

Q6: What substance and what type of misuse are the commonest in the “Drug/Narcotic Violations” crimes?

To visualise the commonest substances and types of misuse in the “Drug/Narcotic Violations” crimes only we created a new dataset with 2 new columns “Substance” and “Misuse” via extracting the latter two from the “Crime Name3” column (Figures 2. and 2.)

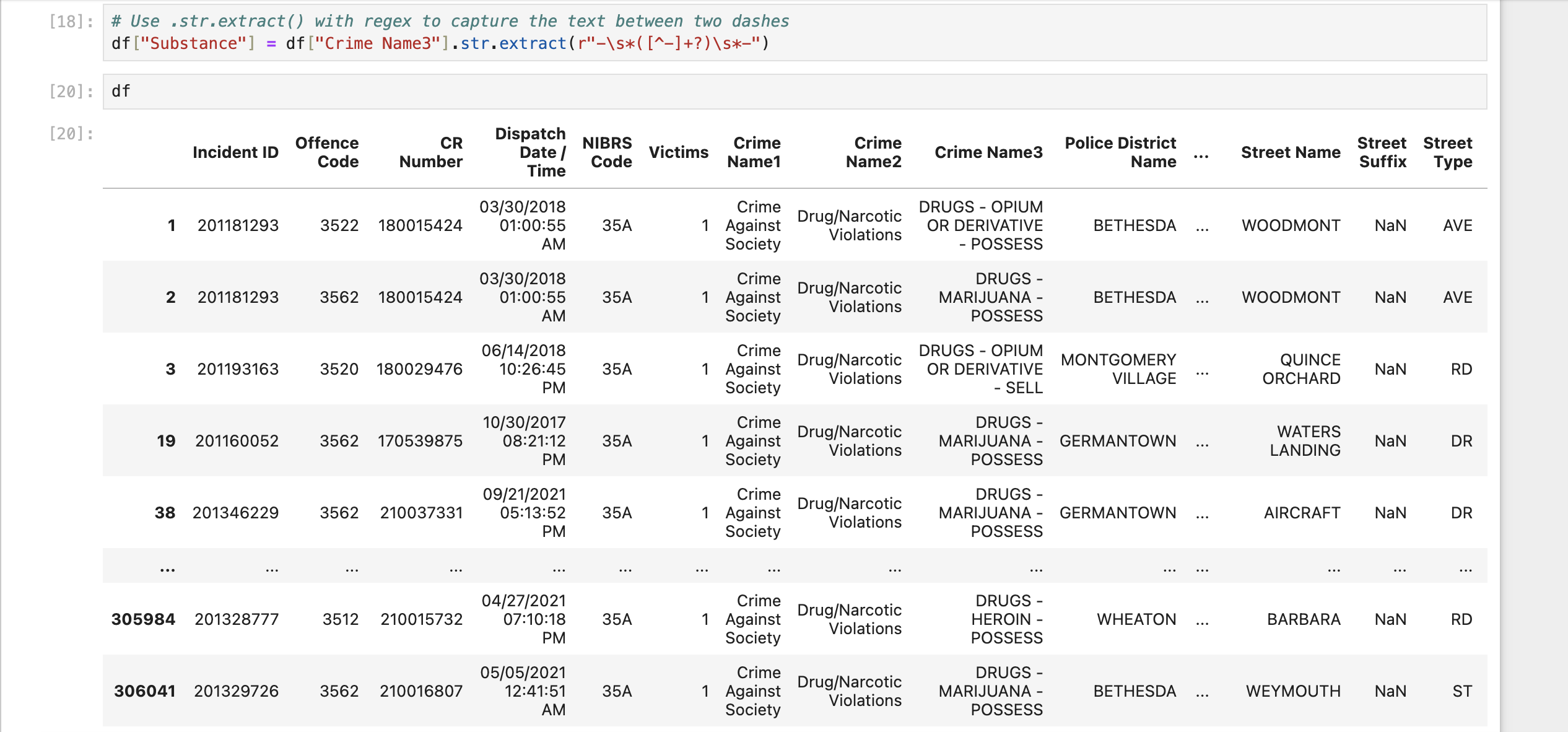
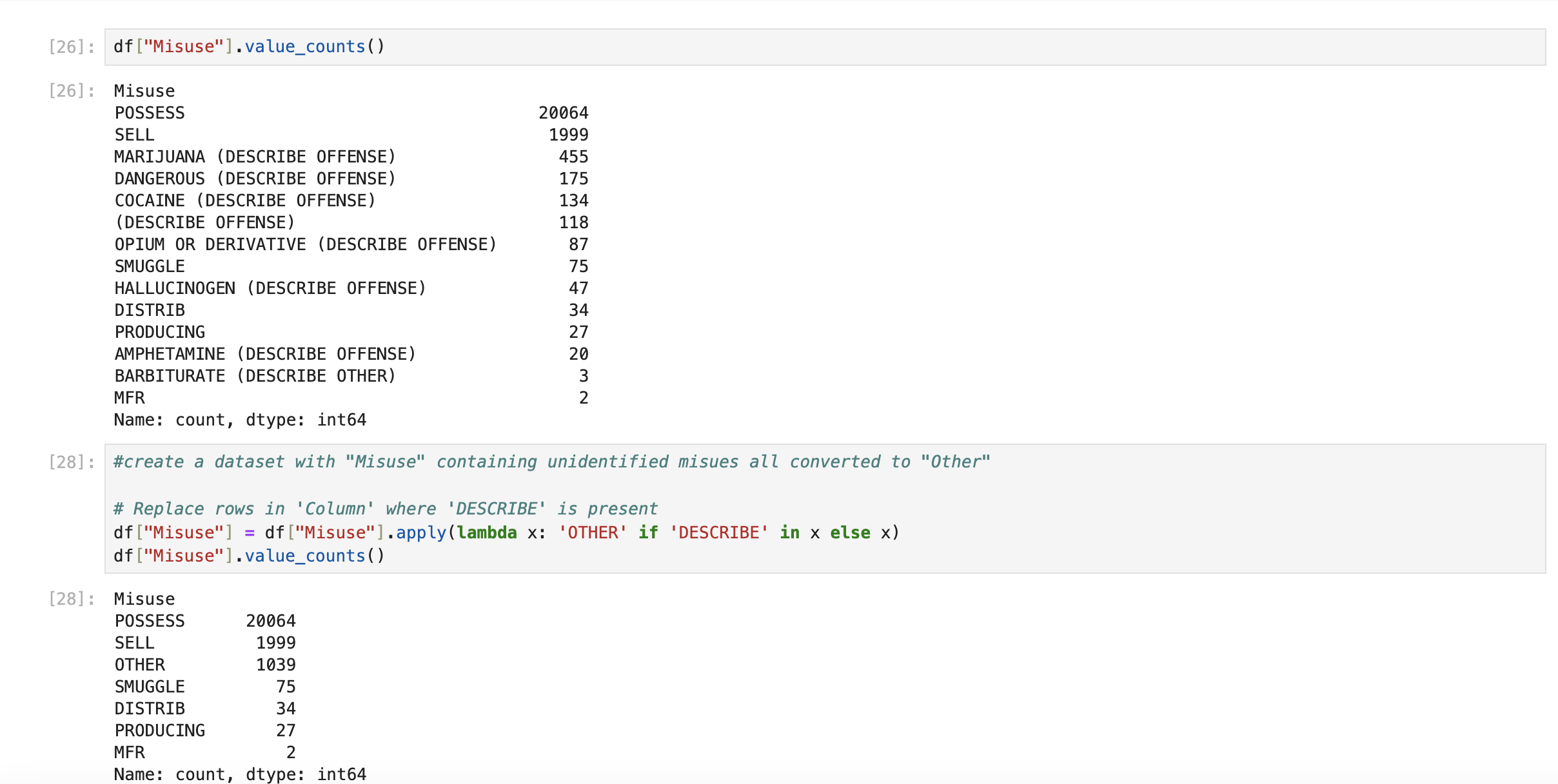


Figure 2.



Figure 2.

We also filled in the missing values for misuse with “OTHER” (Figure 2.).

Figure 2.

As in the case of the “Drug/Narcotic Violations” dataset we were looking at qualitative (e.g. “nonnumerically valued”) values, we were trying to, as Weiss suggests, organize the data in graphs that give “the number of times each value occurs”, i.e. “frequency (or count)” (Weiss, 2017, p. 64). Countplots and pieplots visualise these data well.

Marijuana is legal in Montgomery County (Montgomery Country Government Webpage), so we assumed that most crimes will most likely be related to its possession.

As we can see from the pieplot and countplot below (Figures 2. and 2.), the commonest substance is marijuana indeed and the majority of crimes for all substances are connected with possession. The sell of cocaine counts are not too far from the sell of marijuana, which is probably due to the legalisation of marijuana.

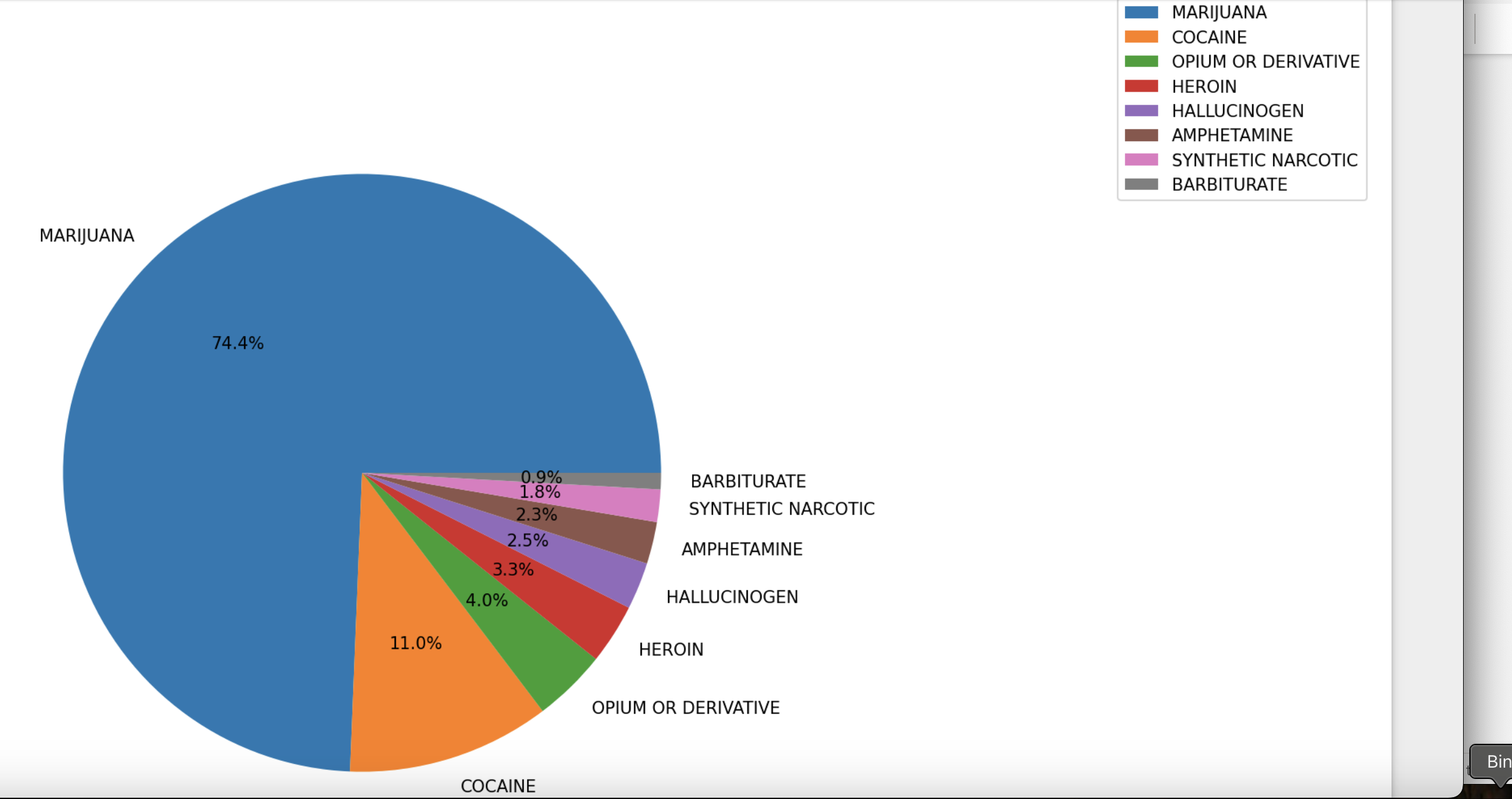


Figure 2.

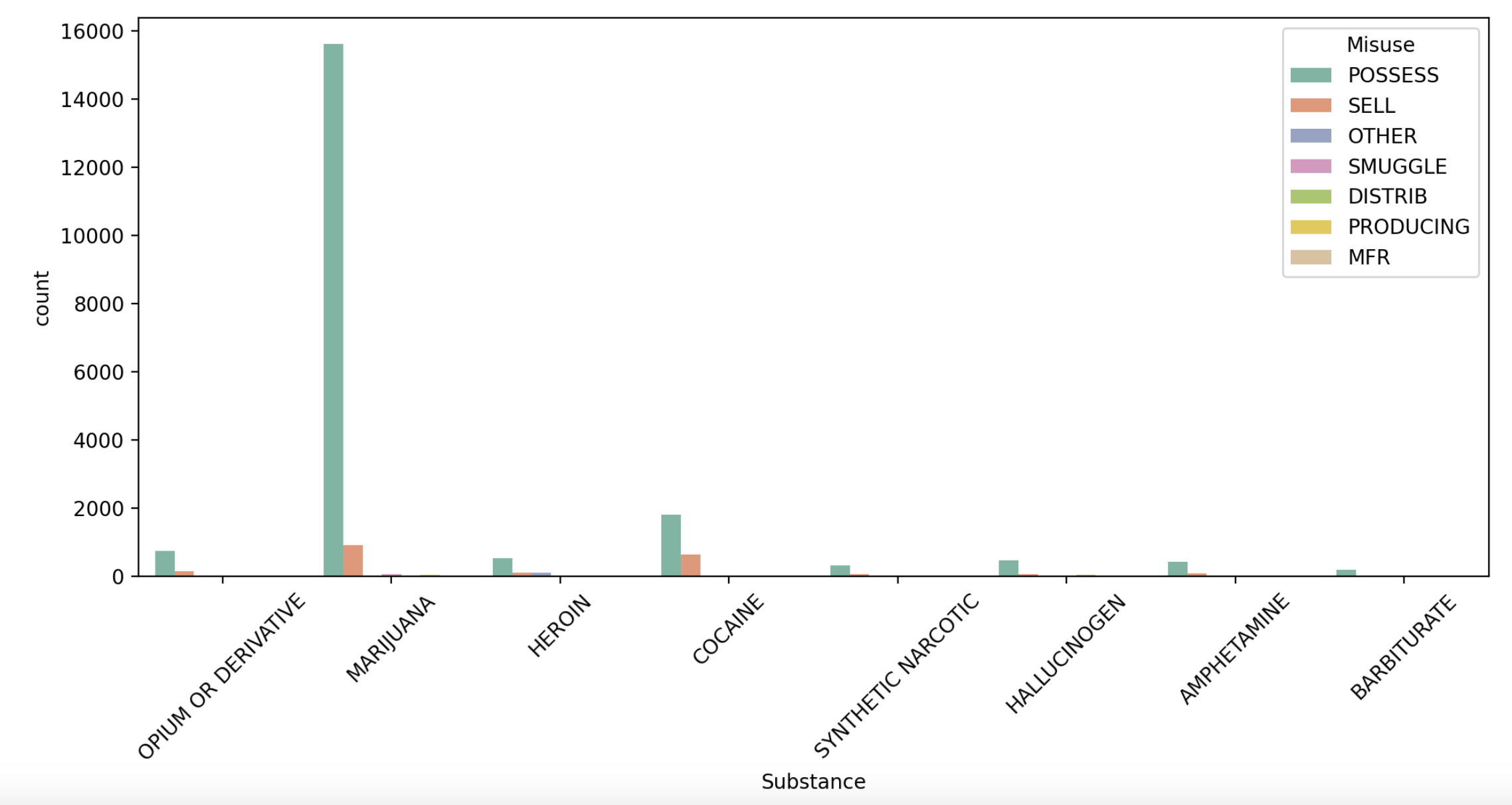
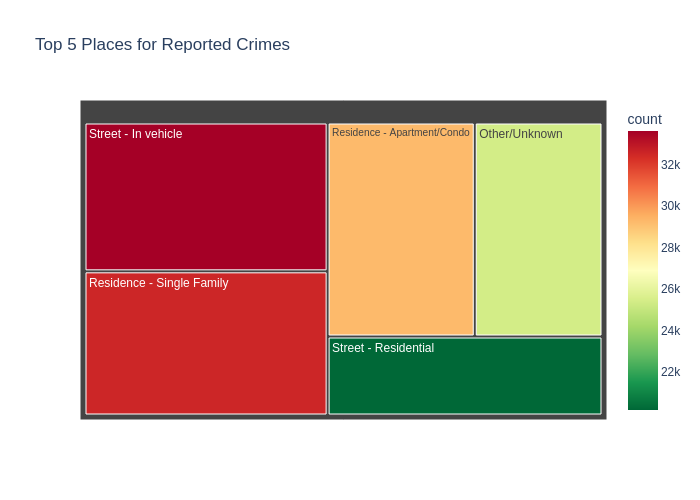


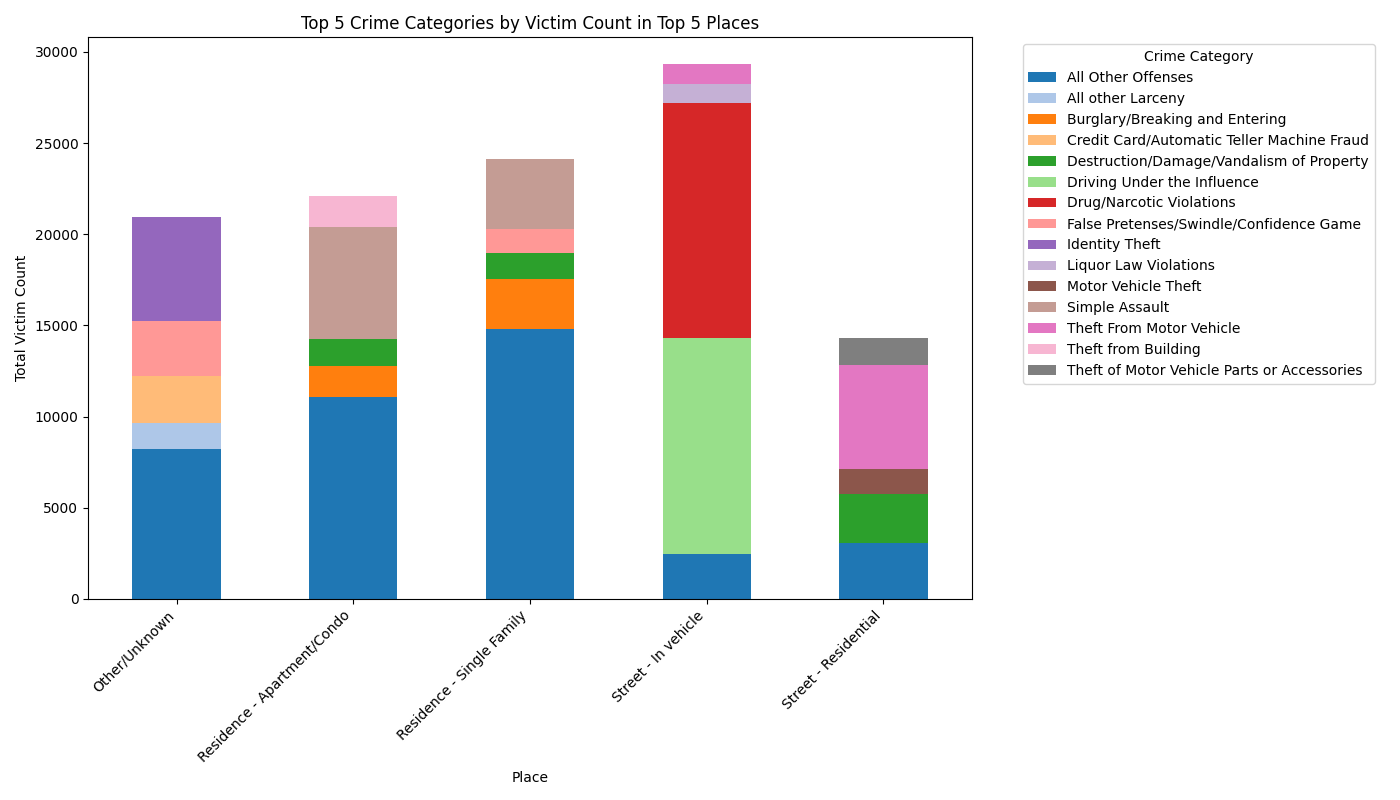
Figure 2.

Q7: Which places are more prone to which types of crimes?

The analysis of crime-prone places uses two complementary charts to explore the most common locations for reported crimes and the nature of offenses in those areas.



The tree map provides an intuitive and visually appealing way to display the relative size of reported crimes across different places, with color-coding to emphasize higher crime counts.



The stacked bar chart effectively displays the distribution of victim counts across different crime categories within the top 5 places, allowing for a clear comparison of both the total and individual contributions of each crime type.

The tree map efficiently highlights the most significant locations, offering an immediate understanding of crime distribution while the visual structure of the stacked bar chart helps highlight trends and the relative impact of different crimes in each location.

Q8: How are vehicle related crimes connected to the street types?

To check if there was a link between crimes connected with vehicles and street types where these crimes were committed we created a new dataset containing only crimes connected with vehicles and excluding all missing values and missing Street Type values (Figure 2.).

Figure 2.

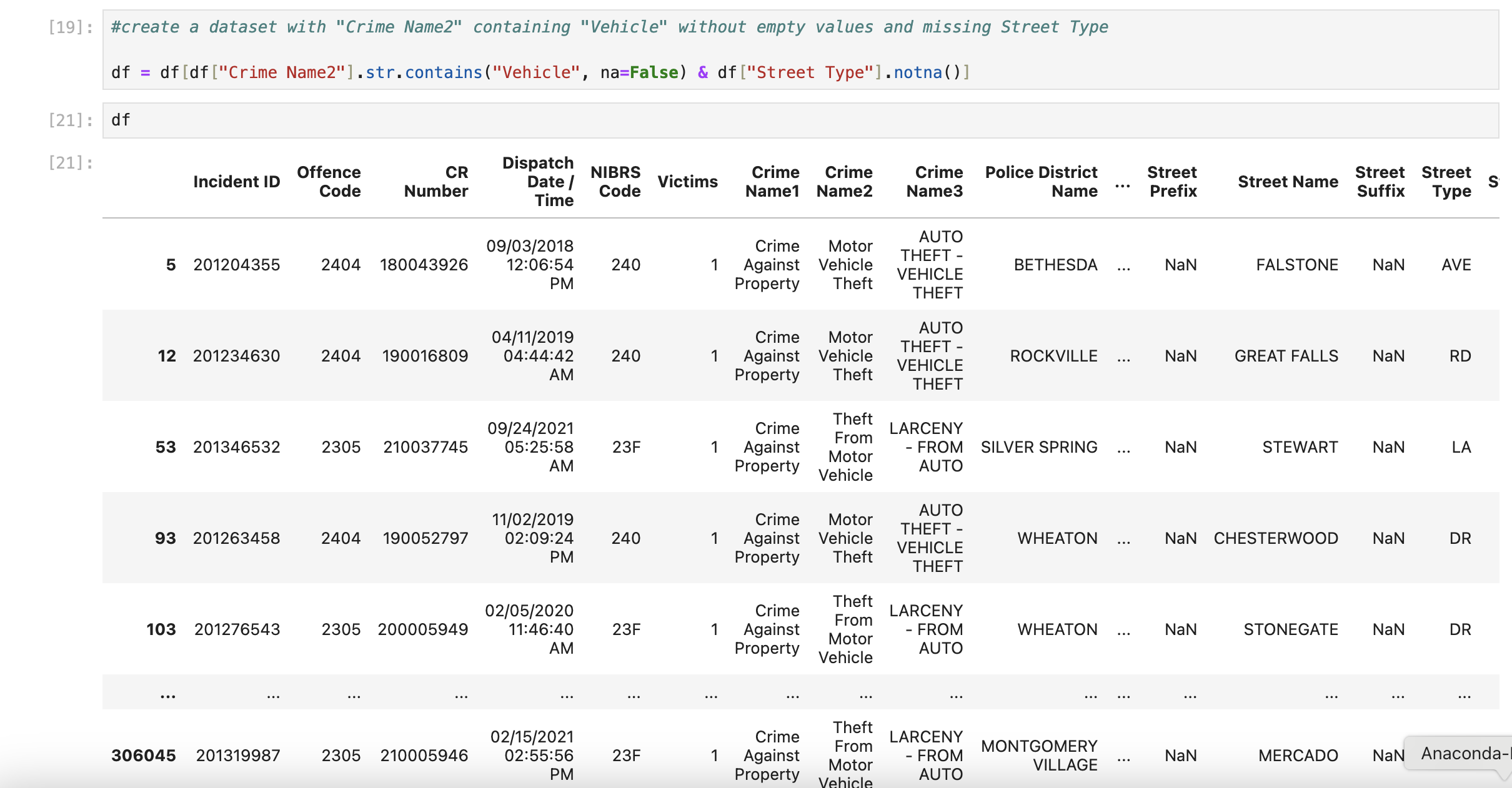


Figure 2.

To visualise the probability of vehicle-related crimes on different street type we selected the countplot graph with the step of probability rather than count for the y-axe as probability (“a generalization of the concept of percentage” (Weiss, 2017, p. 181)) is potentially more relevant for car insurance offers, whereas all cars need to be insured irrespective of the number of vehicle-related crimes. For better distinction between types of vehicle-related crimes we set the hue to Crime Name2.

As we could see from Figure 2., the highest probability of vehicle-related crimes in Montgomery County is just above 0.12, which is relatively low (i.e. much closer to 0 than 1). Half of the street types covered by the dataset show either insignificant or zero vehicle-related criminal activity and can potentially be offered better car insurance deals. The commonest crime for all street types is “Theft From Motor Vehicle” with “Motor Vehicle Theft” and “Theft of Motor Vehicle Parts or Accessories” being much less common and being approximately similar. Street types “DR”, “RD” and “AVE” show the highest probability of vehicle-related crimes.

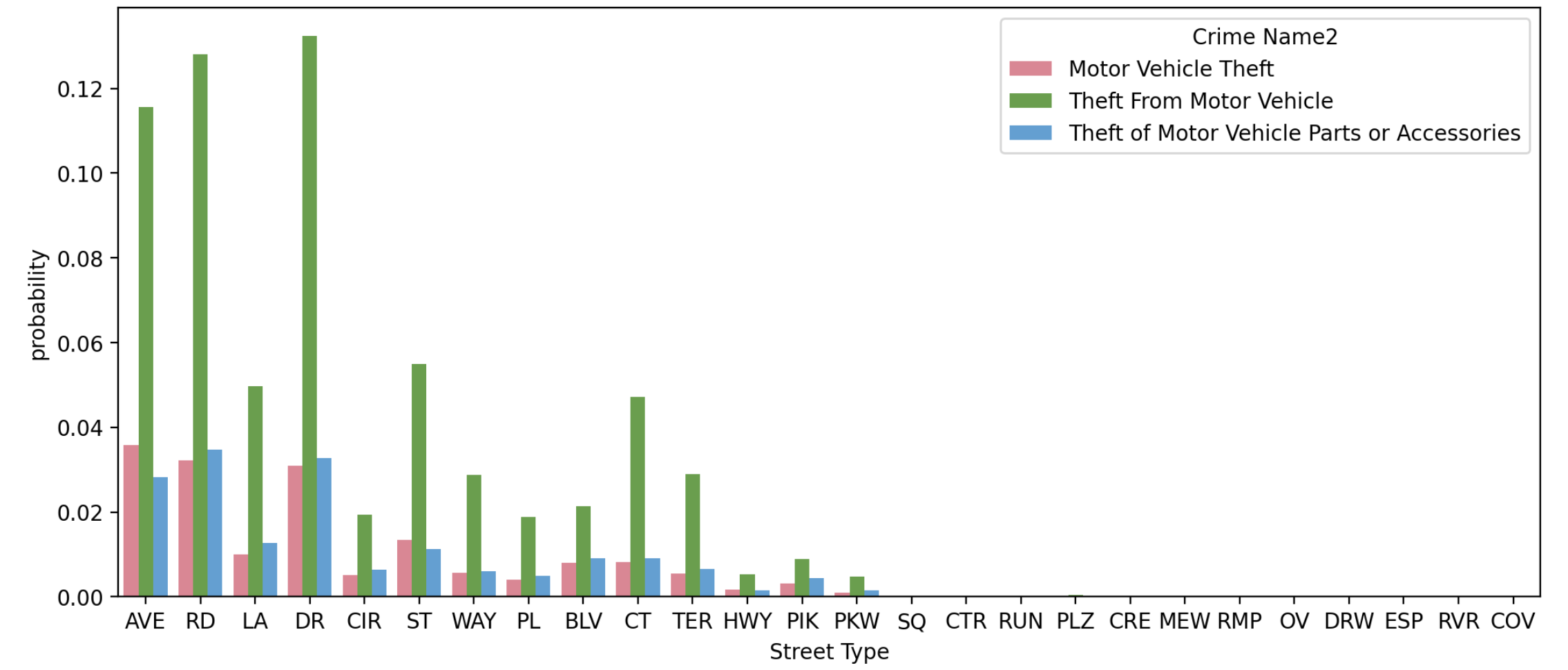
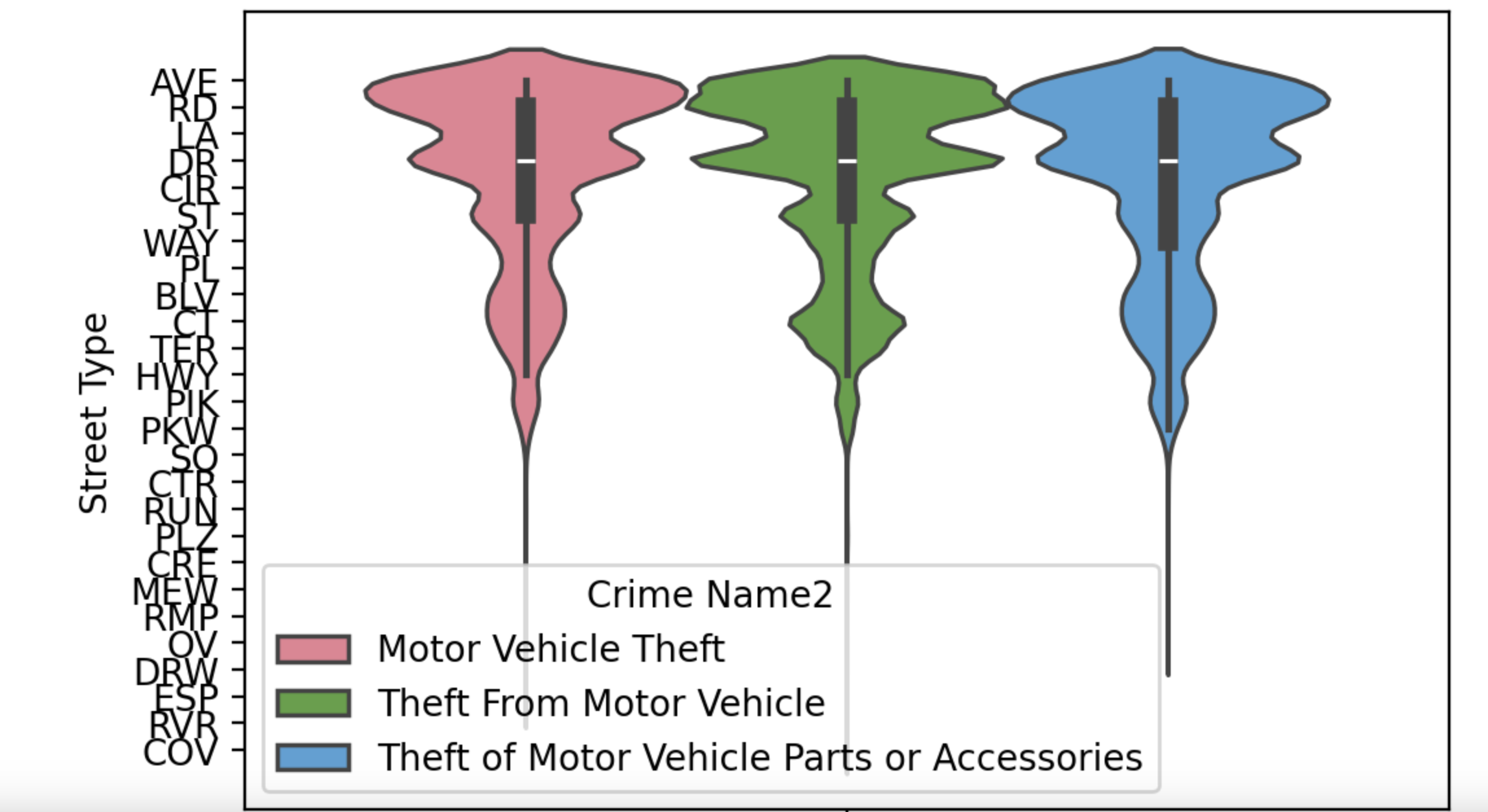
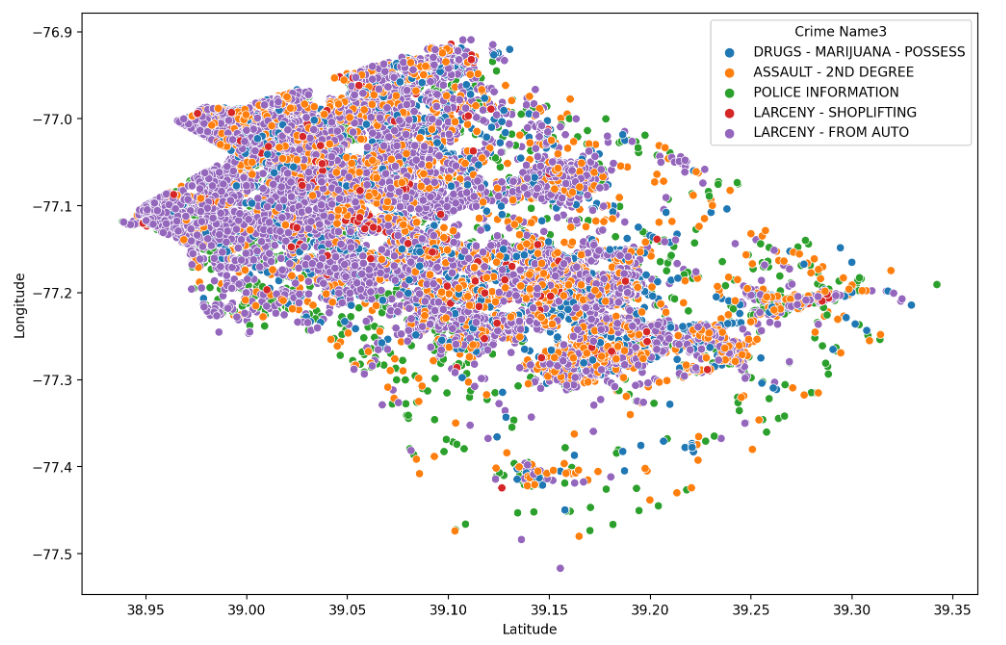


Figure 2.

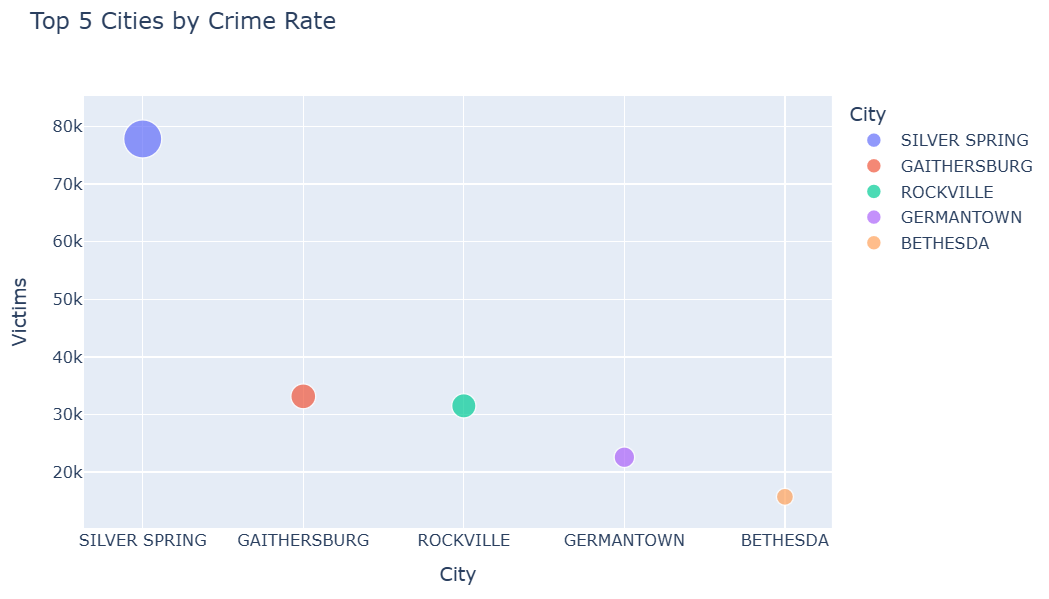
To confirm the pattern is correct and the vehicle-related crimes are the highest on drives, roads and avenues, we used the voilinplot graph with the same hue (Figure 2.). The graph showed the same trend, but also helped us see that crimes were committed on all street types, with the numbers on half of street types being statistically insignificant. These extremely low numbers on certain street types are not outliers, but rather “an indication of skewness” (Weiss, 2017, p. 152).



Q9: What geographic areas show the highest and lowest crime rate?

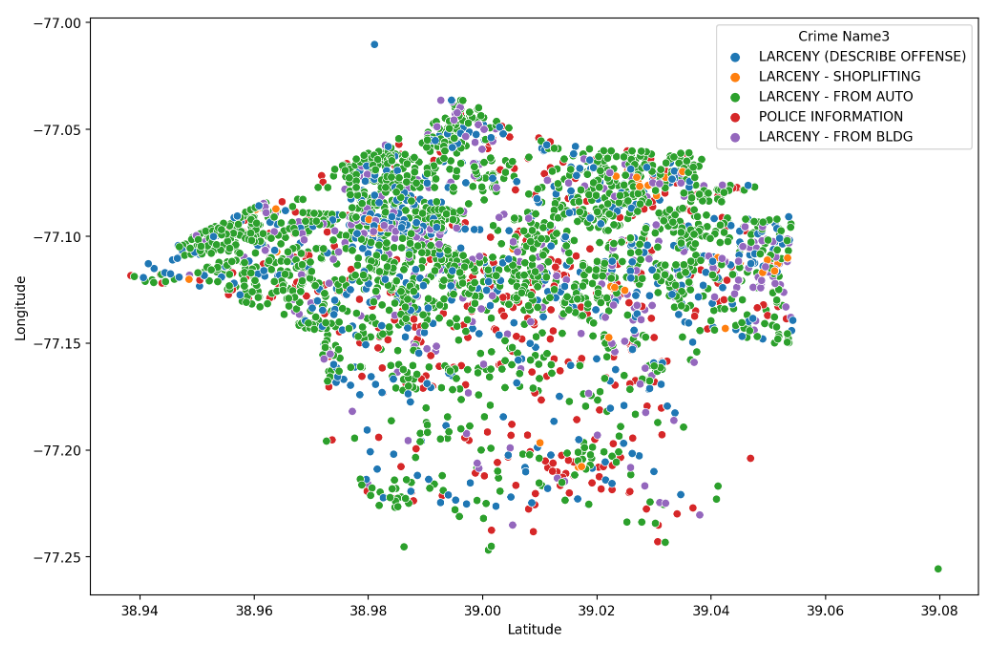


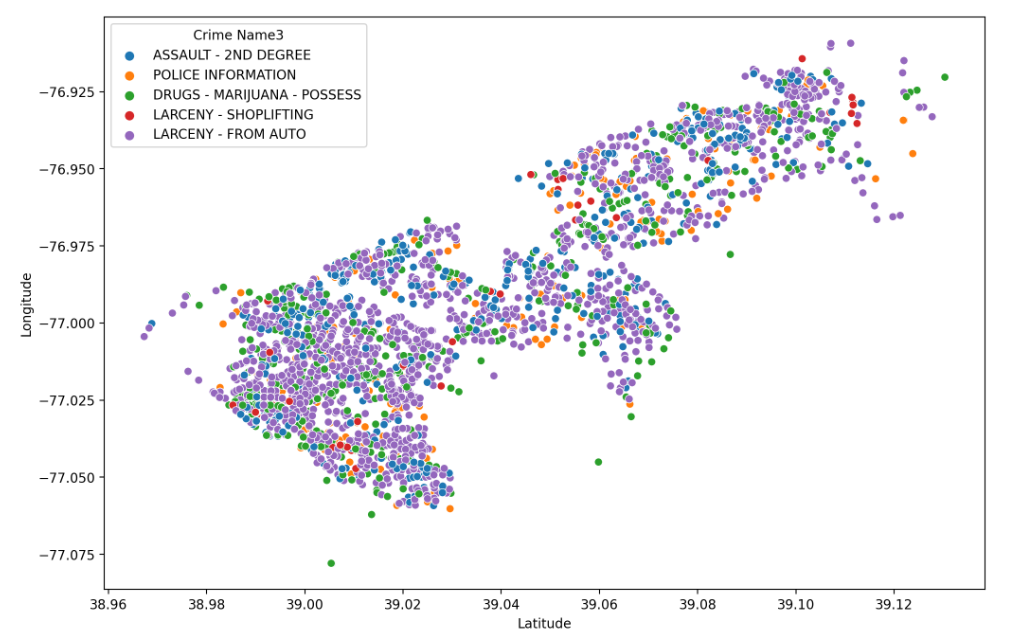
We wanted to see that which geographic areas show more crime which also reflects the fact that which areas are more dangerous than the others. To determine this we decided to plot the data on a scatter plot. To do this we plotted Latitude on the x-axis and took Longitude on the y-axis. In this way it helped us to get an outline of the county as well as the frequency of the crime happening in the regions. The scatter plot shows that the frequency of the crime is high in north west area of the county. This scatter plot shows top five most frequent crimes in the county and even it reflects that that the crime related to Drugs is very high in the west side of the county.

Another visualisation of the crime rates is shown by using the scatter plot. To highlight the total number of victims, varied sizes and colours of bubbles were assigned to each city.

Q10: What type of crime frequently occurs in different geographic areas?

We wanted to find out that which crimes are more frequent in different zones of the county. For this we took the division of the county by police districts. In the data we can see that there are 7 police districts, and we grouped the data according to the district numbers. Then we took the top 5 crimes (according to the ‘Crime Name 3’) in every district and chose the scatter plot to show the results. In the scatter plot we plotted the latitude in the x-axis and longitude in the y-axis. This showed us an outline of the police districts and the crimes happening within them as well. The results are shown for the two Police districts (as limited by the template to share two visualisations per question).

The first plot shows the data for the Police district 1D and top five crimes indicted in it. It shows that Larceny – from auto is the most indicted crime in the county 1D and other crimes follow it. The crime rate is high in the north and northwest region of the police district as compared to the south and south east region of the police district.



The second plot shows the crime happening in police district 3D. The crime that is common in this police district is Larceny – from Auto. It is spread all over the police district along with the other crimes.

1. Summary and Conclusion

* The analysis of crime types by total victim count and victim count per incident reveals that "All Other Offenses" and theft-related crimes, such as "Larceny from Auto," have the highest overall victim numbers. At the same time, crimes like human trafficking, which have a significant victim impact per incident, also emerge as critical. This combined perspective is essential for prioritizing resources, addressing both widespread crimes and those with severe impacts on victims.
* The analysis of crime distribution by location reveals that "Street – in vehicle" is the most frequent crime area, with significant variation in crime types across different places. Crimes like "Driving Under the Influence" and "Drug/Narcotic Violations" are more common in "Street – in vehicle," while "Simple Assault" is prevalent in residential areas, and "Theft from Vehicle" is concentrated in residential streets. The widespread occurrence of "All Other Offenses" across all top places suggests broad law enforcement reporting. This perspective helps identify key crime hotspots and their specific types, which is crucial for targeted crime prevention efforts.
* The distribution of recorded crimes over years generally shows decreasing trend over these years, especially after 2019. When examining crime trends monthly, we observe an increase during the spring and summer months. Most crimes occur during the Evening (6 PM - 12 AM), followed by the Afternoon (12 PM - 6 PM). This pattern suggests that crime incidents tend to increase later in the day. Police departments can be informed about this trend and encouraged to take precautions regarding staffing levels or enforcement measures.
* Among all crime categories, ‘All Other Offenses’ consistently show a high proportion, so having subcategories under this title would be beneficial to have knowledge of the crime types. The most notable change is the visible decrease in Drug/Narcotic Violations over year. We can conclude that the measures taken by police departments regarding drug related issues have been effective. Similar approaches could be discussed for addressing other types of crimes.
* The commonest substance for the “Drug/Narcotic Violations” crimes is marijuana and the majority of crimes for all substances are connected with possession.
* The probability of vehicle-related crimes in Montgomery County is overall relatively low (just above 0.12). Half of the street types covered by the dataset show either insignificant or zero vehicle-related criminal activity. The commonest crime for all street types is “Theft From Motor Vehicle”. Street types “DR”, “RD” and “AVE” show the highest probability of vehicle-related crimes.
* The data shows that Crime against Property is a major issue in the county as this is the most prevailed and consistent crime in all police districts in the county.
* Police District 3D has recorded the most crime count over the years as compared to other police districts. Crime against Property and Crime against Society are the major crimes in that Police District Number.
* When mapped according to the latitudes and longitude, it is evident that the North, North Western and the Western parts of the county are most disturbed by the crimes.
* When we analyse the data according to the Crime Name 3 and see Police District wise, we can see that the Larceny – from Auto is that specific crime which is prevalent in all police districts. This crime is the one crime which has been seen as the most occurring crime over the years in all Police Districts.

# References

* data.montgomerycountymd.gov. (n.d.). Crime | Open Data Portal. [online] Available at: <https://data.montgomerycountymd.gov/Public-Safety/Crime/icn6-v9z3/about_data>
* Bowers, D., 1996. *Statistics from Scratch*., John Wiley and Sons.
* Bhattacherjee, A., 2012. *Social Science Research: Principles, Methods, and Practices*. Tampa: University of South Florida, Digital Commons @ University of South Florida, Textbooks Collection.
* Hess, K., 2022. 80 types of charts & graphs for data visualization (with examples). [online] Available at: <https://www.datylon.com/blog/types-of-charts-graphs-examples-data-visualization>. [Accessed 7 November 2024].
* Grus, J., 2019. Data Science from Scratch, 2nd edition, O’Reilly Media.
* McKinney W., 2022. Python for Data Analysis, 3rd edition, O’Reilly Media.
* Weiss N., 2017. Introductory Statistics, 10th edition, Pearson Education Limited.

Montgomery County (MD) Government. [ttps://www.montgomerycountymd.gov/OPI/cannabis.html](https://www.montgomerycountymd.gov/OPI/cannabis.html) [Accessed 9 November 2024].

# Supplementary material

*Please note this section is only applicable if your group decide to upload the Jupyter notebook (.ipnyb ) or colab ( .ipnyb/.py) file either by merging to the report or creating a zip file. If you include the notebook file, you can* ***write* Supplementary material: Jupyter\_file\_name.ipnyb contains supplementary materials***. This section contents 0 marks for coursework 1*

Street\_Type.ipynb

Substance\_Misuse.ipynb