**Research Report for Crime Dataset**

**TABLE OF CONTENTS**

**FIGURE LIST.....................................................................................................................................3**

**ABSTRACT.........................................................................................................................................4**

1. **INTRODUCTION................................................................................................................................6**

**1.1** **OUR AREAS OF FOCUS FOR VISUALIZATION..................................................................6**

**1.2 UNDERSTANDING THE DATASET.........................................................................................7**

**1.2.1 DATA QUALITY ANALYSIS................................................................................................7**

* + 1. **INITIAL DATA CLEANING ............................................................................................7**

1. **EXPLORATORY DATA ANALYSIS................................................................................................9**
   1. **INTRODUCTION OF EDA.......................................................................................9**
      1. **FINDING NULL AND MISSING VALUES......................................................9**
      2. **CHARACTERISTICS OF DATA......................................................................9**
      3. **NORMALIZATION AND TRANSFORMATION...........................................9**

**3. VISUALIZATIONS AND RESEARCH QUESTIONS.....................................................................10**

**3.1** **ANALYSIS OF TRENDS WITHIN POLICE DISTRICTS...................................10**

**3.1.1 Identification of Police District with the greatest number of crimes...............10**

**3.1.2 What are the most common crimes per police district?...................................12**

**3.1.3 What districts would benefit the most from increased presence?...................13**

**3.1.4** **What time of day would benefit from increased police presence?..................14**

**3.1.5 Police district with success reacting to larceny over the years?......................17**

**3.1.6 What are the predominant type of Drugs Violations from 2018 to 2022........18**

**3.1.7 Comparative Analysis[ Drug Distribution Across Major Cities]....................19**

**3.1.8 What Annual Crime Trends are within the Data Set?.....................................20**

**3.1.9** **What Weekly Crime Trends are within the Data Set?.....................................20**

**3.1.10 Trends within larceny fluctuation (cost of living post covid).........................21**

* + 1. **ocational Trends of Larceny – focusing on crime and street types ............22**
    2. **rends on overall crime rate spanning years of interest................................24**

1. **CONCLUSION...................................................................................................................................25**
2. **BIBLIOGRAPHY...............................................................................................................................27**

*Figure list*

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 1 | Inputing python packages and datset information | 7 | Location of crimes on a map of Montgomery | 13 | Distribution of crimes per hour of day | 19 | Crimes per day of the week |
| 2 | Significance assessment of columns | 8 | Bar plot of different crime types in each district | 14 | Trends of different types of larceny crime | 20 | Types of crimes per day of the week |
| 3 | Checking how many null values | 9 | Most Common crimes per District | 15 | Pie chart of the types of drug crimes | 21 | Larceny types over Covid periods |
| 4 | Cleaning the dataset and formatting dates | 10 | Heatmap of crimes in each District | 16 | Distribution of drug crimes by city | 22 | Larceny types on Street types |
| 5 | No. of crimes per district | 11 | Heatmap of crimes per hour in each District | 17 | Drug crimes type over time | 23 | Overall crimes per month from 2018-22 |
| 6 | Boxplot for crimes per district | 12 | Top crimes per hour of the day | 18 | Seasonal plots by day/month | 24 | Distribution of crimes per month |

**ABSTRACT**

Analyzing crime records is crucial for developing strategies to reduce felony rates, improve community safety, and strengthen trust between police departments and the communities they serve. This research examines crime trends in Montgomery County and proposes strategies to address high-crime areas. Using a dataset derived from the National Incident-Based Reporting System (NIBRS) and sourced from Kaggle, this study analyzes crime data reported between 2018 and 2020. The analysis identifies prevalent crime types, temporal and spatial patterns, and incident correlations. Visualization techniques, such as heatmaps, bar charts, and line graphs, are used to present findings.Preliminary results indicate an overall decline in crime but show increased post-pandemic rates of larceny and drug-related offenses. These findings inform data-driven strategies for crime prevention and resource allocation. Additionally, this research emphasizes the potential of predictive frameworks, including machine learning models, to anticipate future crime trends and support long-term policy development.

**1. INTRODUCTION**

This dataset, sourced from Kaggle and covering Montgomery County, Maryland, provides reliable crime statistics for analysis. It aims to identify patterns and make recommendations to improve police district efficiency and safety. The study focuses on understanding crime trends to optimize resource allocation and inform policing strategies. The research includes data cleaning, exploratory analysis, and transformation for consistency. Key questions addressed include identifying the highest-crime districts, common crime types, and times requiring increased police presence. The study also examines trends in larceny, drug violations from 2018-2022, and crime distribution across major cities. Further analysis explores annual and weekly crime trends, and locational patterns for specific crimes like larceny. The goal is to provide actionable insights for targeted policing and enhancing public safety.

* 1. **OUR AREAS OF FOCUS FOR VISUALISATION**
* Identification of Police District with the greatest number of crimes
* What are the most common crimes per police district?
* What districts would benefit the most from increased presence (using specific incident examples)?
* What time of day would benefit the most from increased police presence?
* Which police district has been the most successful in reacting to larceny offences by reducing the number of incidents over the years?
* What are the Predominant type of Drugs Violations from 2018 to 2022
* Comparative Analysis of Drug Type Distribution Across Major Cities over the years.
* What Annual Crime Trends are within the Data Set?
* What Weekly Crime Trends are within the Data Set?
* Trends within larceny fluctuation (cost of living post covid)
* Trends on overall crime rate spanning years of interest

**1.2 UNDERSTANDING THE DATASET**

**1.2.1 DATA QUALITY ANALYSIS**

A screenshot of a computer

Description automatically generated*A screenshot of a computer

Description automatically generated*The dataset consists of a comprehensive set of variables representing aspects of police incident data within districts of Montgomery MD. It includes 30 columns, spanning approximately 306,094 rows (pre cleaning). These columns are representative of different data types, numerical, textual, and categorical, with data types falling into: float64, int64, or object for string-based entries. Most of the data in columns are complete with minimal missing values however, some columns do have substantial missing data. For the ‘Dispatch Date/ Time’ approximately 49,000 entries were missing which could impact time related analysis particularly related to analysing police response time. Similarly, for ‘End Date/ Time’, over 161,000 values were missing, which could impact the incident duration analysis. Around 26,000 entries were missing for ‘Block Address’ and ‘Address Number’, limiting the geographical analysis of crimes. Moreover, ‘Street Prefix’ and ‘Street Suffix’ contain over 90% of null values raising concerns for address readability.

Figure 1

**1.2.2 INITIAL DATA CLEANING**

As part of our initial data cleaning, we assigned each column a significance level between 1 and 5, with 5 being the most significant (see figure 1). This dataset includes approximately 865,131 missing values, notably within column ‘Dispatch Date/ Time’. Given the high significance level (5) assigned to this column, we removed rows with null values from this field to preserve integrity within our Exploratory Data Analysis (EDA). Conversely, we chose to omit columns of lower significance, such as ‘Block Address’ and ‘Street Name’ as they did not require significant analysis given their minimal impact on overall outcomes of our research.



Figure 2

**2. EXPLORATORY DATA ANALYSIS**

**2.1 INTRODUCTION OF EDA**

Exploratory Data Analysis (EDA) is a process that uses summary statistics and visualisations to explore and analyse a dataset.

**2.1.1 FINDING NULL AND MISSING VALUES**

To enhance data quality for analysis, various data cleaning and transformation methods were applied. Key columns, such as ‘Dispatch Date/Time,’ ‘Block Address,’ and ‘City,’ were reviewed for missing values. Categorical gaps were filled with ‘Unknown’. Standardising date and time fields to a consistent format supports effective time-based analysis. Text fields like ‘Street Name’ and ‘Crime Name’ were also standardised to prevent case-sensitive duplicates.

The figure below represents the total null values in the dataset prior to cleaning the dataset.

A screenshot of a computer

Description automatically generated

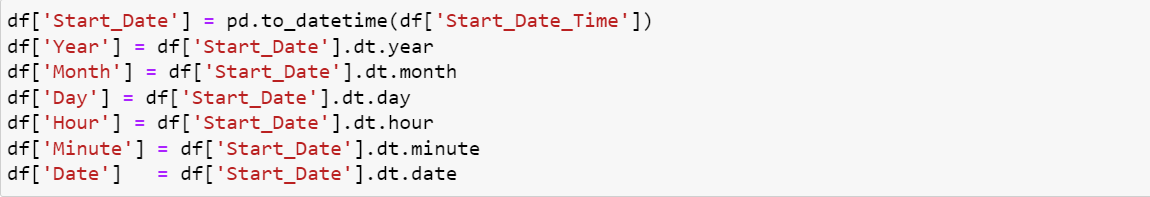
Figure

**2.1.2 CHARACTERISTICS OF DATA**

Duplicate entries present within fields such as ‘Incident ID’ and ‘Offence code’ were identified as we explored the dataset. From this, we were able to remove the duplicates within the ‘Offence code’ column to avoid redundancy within the data. However, the duplicates entries within ‘Incident ID’ were attributed to multiple offences being committed at the same time, and as such were left within the dataset to ensure it was representative of all crimes committed.

**2.1.3 NORMALIZATION AND TRANSFORMATION**

Column transformations and normalization were applied to scale crime data on a yearly basis, enabling trend analysis of response rates, repeat crimes, and police district response times. Key categorical variables, such as ‘Start Date/Time’, were split into separate columns for easier analysis. These variables were converted to numerical formats or dummy variables to improve model interpretability. Filtering focused on relevant rows, prioritizing columns like ‘Offense Code’, ‘Victims’, ‘Police District Number’, and ‘Crime Names’ to identify crime patterns. Missing data was excluded to preserve data integrity and avoid bias from null values. Data types were corrected for numerical accuracy, and date strings (DD/MM/YYYY) were split into components for time-based analysis. Data prior to 2018 and after 2022 were removed using pandas to reduce noise and ensure the dataset only included data from 2018 to 2022. This process helped ensure consistency and accuracy for meaningful analysis.







Figure

**3. VISUALISATIONS AND RESEARCH QUESTIONS**

**3.1.** **ANALYSIS OF TRENDS WITHIN POLICE DISTRICTS**

**3.1.1 Identification of Police District with the greatest number of crimes**

A screenshot of a computer

Description automatically generated

Figure 5

A screenshot of a computer

Description automatically generatedIdentification of police districts with the highest number of incidents is invaluable when considering the nature of the dataset presented. Represented by the above barplot, using Seaborn, data analysis for Montgomery County presents Silver Spring as having the highest number of reported crimes (22096) with a significant difference between Takoma Park (631) as the lowest. This visualisation aims to highlight areas with the highest crime rates, which is essential for understanding where law enforcement resources may be most effectively deployed, something that will be discussed further in this report. This data is visualised through using the pandas value counts tool in relation to police district name, ensuring all data is accounted for.

Figure 6

The interquartile range (IQR) is useful in EDA of crime data, showing the spread of crime rates and highlighting patterns and outliers. The 25th percentile (Q1) is 2,745 incidents, meaning 25% of districts have fewer incidents. The median (50th percentile) is 13,704, with half of districts above and below this value. At the 75th percentile (Q3), crime incidents rise to 17,191, indicating that 25% of districts exceed this number. The IQR of 14,446 shows significant variability in crime rates across districts, suggesting disparities and the need for tailored crime strategies, as shown in *figure 6*.

Mapping crime locations using Plotly visualises crime data by type, with each colour representing a different crime. An interactive scatter plot allows users to identify high or low crime areas, helping pinpoint crime hotspots and assess regional crime impact.

A map with many dots on it

Description automatically generated

Figure 7

**3.1.2 What are the most common crimes per police district?**

When considering the identification of crime types per police district data was visualised using a catplot. From this, comparisons can be drawn between the overall most common types of crime, in addition to average crime types per police district. Omitting the ‘other’ column from police district analysis ensures that an accurate and representative comparisons can be drawn as only districts with defined names and boarders can be considered. From this catplot, it is evident that the most common crime types across districts are larceny offences. For this visualisation the column name ‘Crime Name3’ was utilised to give a more accurate representation of the specific incidents that are occurring as the ‘All Other Offences’ label in ‘Crime Name2’ was too broad and did not provide enough information for specific analysis of crime incidents per district.

A screenshot of a computer

Description automatically generated

Figure 9

This is further evidenced by *figure 9* which displays the top 10 most common crimes by police district. Using a bar plot with all data represented shows a comparison between districts and crime occurrence. With districts such as Rockville, Bethesda, City of Takoma Park, and Takoma Park consistently being presented as having lower crime rates for the most common crime types. Whilst this representation is indicative of most common crime types and their frequency, it does not include all crime types. Using figure 4 in conjunction with figure 1 allows us to compare the overall safety of police districts.

**3.1.3 What districts would benefit the most from increased presence (using specific incident examples)?**

Figure 10 presents the distribution of common crimes under ‘Crime Name2’ across Montgomery County’s police districts. Each row represents a district, while each column shows a crime type. The intensity of the colour reflects crime frequency, with darker blue indicating higher crime rates. Districts like ‘Montgomery Village’, ‘Rockville’, ‘Silver Spring’, and ‘Wheaton’ showed consistently high crime rates, especially for thefts from vehicles and DUI offenses. ‘City of Takoma’ and ‘Takoma Park’ had lower crime rates overall. Notably, Montgomery Village had the highest rates for property crimes such as vandalism and motor vehicle theft. Drug violations were most common in Silver Spring and Montgomery Village. Rare crimes, like liquor law violations and identity theft, were less frequent across all districts. These findings can help law enforcement allocate resources effectively and target interventions in high-crime areas to improve public safety.

A screenshot of a computer

Description automatically generated

Figure

**3.1.4 What time of day would benefit the most from increased police presence?**

A screenshot of a computer

Description automatically generated

Figure

A graph with lines and numbers

Description automatically generated

Figure

The heatmap (Figure 6) displays crime frequency by time of day across police districts, highlighting patterns not evident in raw data. Darker colours indicate hours with higher crime volumes, helping identify peak times for specific crimes. For example, Wheaton and Montgomery Village experience high crime rates between 12am and 3am, suggesting the need for increased patrols at night. Bethesda and Silver Spring show spikes in larceny offenses between 7am and 10am, guiding targeted resource allocation. Focusing on property crimes, Theft from Motor Vehicle is the most common, peaking at 4pm and 8pm. Other larceny and theft offenses show slight peaks around 12pm, with a decline after 9pm. This analysis aids in tailoring law enforcement efforts based on the timing and trends of crimes in each district.

**A graph with a line

Description automatically generated**

Figure

**3.1.5 Which police district has been the most successful in reacting to larceny offences by reducing the number of incidents over the years?**

A graph of different colored lines

Description automatically generated

A graph of different colored lines

Description automatically generated

A graph of different colored lines

Description automatically generated

Figure 14

Using lineplots to visualise these instances, a downward trend on all larceny offences is presented. A significant decrease in ‘Larceny – Describe Offence’ is displayed in Silver Spring between 2019 and 2022, with further decrease in larceny incidents across all figures (7,8,9). Police attention to this detail may have caused the significant decrease in 2022. Whilst Silver Spring remains the highest for Larceny – ‘Auto Parts’, the district of Bethesda has the highest number of larceny offences in the year 2022. Despite being on a downward trajectory, suggesting success in curbing larceny offences, a recommendation can be made to increase police attention on theft outside of Autoparts in Bethesda.

**3.1.6 What are the Predominant type of Drugs Violations from 2018 to 2022**

A pie chart with different colored circles

Description automatically generated

Figure

The pie chart illustrates the distribution of drug violations in Montgomery County from 2018 to 2022. Marijuana possession dominates, accounting for 69.2% of total violations, highlighting it as a key focus for drug enforcement agencies. Cocaine-related violations follow, with possession at 8.9% and sales at 3.1%, while Opium/Derivatives possession accounts for 3.4%. Other notable categories, including Amphetamines, Heroin, Hallucinogens, and Synthetic Narcotics, each comprise 2–2.2% of incidents. This distribution offers valuable insights for public health and enforcement strategies. The high rate of marijuana possession may align with broader societal trends, potentially informing future policy debates like decriminalization. Meanwhile, the presence of other drugs underscores the need for targeted enforcement and community education programs to address diverse drug-related challenges.

**3.1.7 Comparative Analysis of Drug Type Distribution Across Major Cities over the years.**

A screenshot of a graph

Description automatically generated

A screenshot of a computer screen

Description automatically generatedFigure

Figure 17

Figures 16 and 17 show drug violation distribution across the top five cities in Montgomery County, with each color representing a drug type. Marijuana possession is the most common violation, especially in Silver Spring and Gaithersburg, the cities with the highest counts. The line area plots display drug violation trends from 2018 to 2022, with a noticeable decline in offenses by 2020, likely due to changes in enforcement. Marijuana remains dominant, particularly in Silver Spring and Gaithersburg. Other drugs like cocaine and heroin show lower, ongoing violations, with Rockville, Germantown, and Bethesda reporting minimal activity by 2021.

**3.1.8 What Annual Crime Trends are within the Data Set?**

To identify patterns in the crime data, seasonal analysis (Hayes, 2022) will be used with line plots, box plots, and bar graphs for clearer visualization. This approach reveals a steady decline in crime each year since 2018, showing a negative correlation over time. Crime rates decreased by 11% from 2018 to 2020, 10% from 2020 to 2021, and 16% from 2021 to 2022. This trend may reflect improved data accuracy, enhanced policing and legislation post-COVID, or reforms following the 2020 protests after George Floyd’s death (Ash-har, Wender, 2024).

A graph of a graph of a graph

Description automatically generated with medium confidence

Figure 18

The line plot (*figure 18*) shows monthly crime fluctuations from 2018 to 2022 with peaks in May-August and a sharp decline in August 2022, the box plot on the bottom left reveals stable yearly medians with variations and 2022's lower range possibly linked to August's drop, and the bottom box plot on the right highlights consistent monthly medians with peaks in June, decreases in December, and notable outliers, including August.

**3.1.9 What Weekly Crime Trends are within the Data Set?**

A graph of blue rectangular bars

Description automatically generatedWeekly crime patterns showcase Friday, Monday, and Wednesday as the most frequent days for criminal activity *(see figure 19),* with Friday leading 15,342 incidents. Interestingly, crime levels dip on the weekends, with Saturday (13,484) and Sunday (12,664) showing a 5% decrease in crime. This finding suggests the county is aware of an increase in crime potential, however, have not found it viable reasoning to also increase police presence on a Friday

Figure 19

A screenshot of a graph

Description automatically generated

Figure

Types of crime being recorded are evidently higher when acted out against property. Figure 20 further supports the finding of Friday, Monday and Wednesday being the most proficient days of crime occurrence. In addition to this, we can dissect Crime Against Person to be the second highest offense taking place.

**3.1.10 Trends within larceny fluctuation (cost of living post covid)**

The COVID-19 pandemic brought a 25% decrease in overall crime in Montgomery County between March 30, 2020, and March 2021.

Pre-COVID (see Figure 20), crime was dominated by larceny, particularly involving automobiles, bank locations, and buildings. However, unique cases such as “Describe Offense” in the larceny category are not directly comparable to other data.

During COVID, larceny rates were cut in half, largely due to 2020 quarantine restrictions limiting public movement. A study on COVID-19's impact on crime in Montgomery County reported a 25.3% decrease in larceny (Ma & Wu, 2022).

Post-COVID, larceny rates rose again, but not to pre-pandemic levels. Larceny of auto parts saw its largest increase since 2018, potentially linked to the economic downturn caused by the pandemic (Grazado & Hunt, 2021; CBPP, 2024).

A graph of different colored squares

Description automatically generated

Figure 21

**3.1.11 Locational Trends of Larceny** – **focusing on crime and street types**

A graph with different colored bars

Description automatically generated

Figure 22

The street types "RD" (Road), "DR" (Drive), "AVE" (Avenue), and "ST" (Street) report the highest larceny counts (*figure 22)*, marking them as hotspots. The most frequent larceny types include “Larceny - Auto Parts” (orange), “Larceny - From Auto” (green), and “Larceny - Describe Offense” (blue), especially on “RD,” “DR,” and “AVE.” In contrast, less common street types like “TRL” (Trail), “WAY,” and “TER” (Terrace) show low larceny counts. Rare crimes, such as “Larceny - Postal” and “Larceny - Pocket Picking,” appear only occasionally, indicating they are less prevalent or more situational.

**3.1.12** **Trends on overall crime rate spanning years of interest**

A graph of a number of crime per month

Description automatically generated

Figure

Using the group by function in pandas, we visualized monthly crime totals, showing an overall decrease with notable fluctuations. A sharp decline from 1,837 to 1,384 crimes between March and April 2020 aligns with Governor Hogan's March 23rd closure of nonessential businesses and the March 30th statewide stay-at-home order (Shahzad, 2021)A blue line on a white surface

Description automatically generated

Figure

As we see here from the box plot of the monthly crimes, there are no great outliers in the dataset – to ensure the data wasn’t skewed at the end we have removed the crimes from August 2022 as we do not have the crimes from the entire month. If we had included August, where we have 211 crimes recorded, this would introduce an outlier in the boxplot, suggesting that there may be a reason for a huge decrease in crimes in Montgomery at the end of our data.

**4. Conclusion**

Montgomery County’s crime data highlights key trends for guiding policing strategies. Crime peaks during summer, with the highest activity on Fridays, Mondays, and Wednesdays, while weekends see a drop. Property crimes, especially vehicle-related, are prevalent, particularly in Silver Spring, Bethesda, and Rockville, indicating a need for increased patrols and public awareness on vehicle security. Drug-related offenses, mainly marijuana possession, are concentrated in Silver Spring and Gaithersburg, though other drugs contribute to crime. The pandemic shifted crime patterns, reducing traditional offenses but increasing fraud. Aligning patrol schedules with peak times (11 AM–1 PM, 8 PM–12 AM) could improve response and resource efficiency. By adapting policing strategies to these patterns, Montgomery County can better allocate resources, enhance crime prevention, and reduce criminal activity in the future.

**BIBLIOGRAPHY**

Ash-har, Q,. Samantha W.(2024) *Minneapolis poured millions into police reform after George Floyd’s death in 2020. where do things stand now?.* *CBS News*. Available at: <https://www.cbsnews.com/news/minneapolis-police-reform-george-floyd-death/#:~:text=Months%20after%20O’Hara%20took,the%20ones%20used%20on%20Floyd>

CBPP (2024) *Chart book: Tracking the recovery from the pandemic*, *Chart Book: Tracking the Recovery From the Pandemic Recession*. Available at: https://www.cbpp.org/research/economy/tracking-the-recovery-from-the-pandemic-recession

Gasset, G. (2022) *USCRIMESDATASET*, *USCrimesDataset*. Available at: <https://www.kaggle.com/datasets/jgiigii/uscrimesdataset>

Grazado, E. and Hunt, R. (2021) *Covid timeline*, *INTERACTIVE: A timeline of COVID-19 cases during phased reopening in Maryland*. Available at: https://cnsmaryland.org/interactives/spring-2021/covid-timeline/index.html

Hayes, S. (2022) *Finding seasonal trends in time-series data with python*, *Medium*. Available at: <https://towardsdatascience.com/finding-seasonal-trends-in-time-series-data-with-python-ce10c37aa861>

Ma, L. and Wu, C. (2022) ‘The impact of covid-19 on crime: A study from the spatial-temporal perspective in the Montgomery County, AL’, *Proceedings of the 14th International Conference on Agents and Artificial Intelligence*, pp. 532–539. doi:10.5220/0010856700003116

*Montgomery County of Maryland - Crime* (2024) *Data Catalog - Crime*. Available at: https://catalog.data.gov/dataset/crime

Shahzad, M. (2021) *One year later: A pandemic timeline in Montgomery County*, *Montgomery Community Media*. Available at: <https://www.mymcmedia.org/one-year-later-a-pandemic-timeline-in-montgomery-county/>