

**Understanding Louisville Metro Crime Through the Lens of Funding**

Preethi Mahankali

Richard Castaneda

Shailaja Bommena

Tamanna Fatima Syeda

University of North Texas  
Toulouse Graduate School of Advanced Data Analytics  
ADTA 5910 – Capstone  
Dr. Denise Philpot  
Fall 2023

**TABLE OF CONTENTS**

**CHAPTER 1- INTRODUCTION...............................................................................................4**

Background.....................................................................................................................................4

Research Questions.........................................................................................................................5

**CHAPTER 2- LITERATURE REVIEW...................................................................................6**

Crime History & Patterns...............................................................................................................6

Weather-Related Influences on Crime Rates.................................................................................9

Budget Constraints and the Erosion of Law Enforcement Efficacy.............................................11

The Economic Theory of Crime and Increased Police Force.......................................................16

Conclusions...................................................................................................................................18

**CHAPTER 3- METHODOLOGY: DATA PREPARATION..................................................21**

Software........................................................................................................................................20

Data Collection for Crime Rates...................................................................................................20

Data Wrangling Cleaning and Description for Crime Rates........................................................23

Data Collection for Funding.........................................................................................................25

Data Wrangling Cleaning and Description for Funding...............................................................27

**CHAPTER 4- METHODOLOGY: EXPLORATORY DATA ANALYSIS...........................28**

EDA and Methods for Crime Rates..............................................................................................28

Missing Values..............................................................................................................................29

Outliers..........................................................................................................................................30

Funding Methods and EDA..........................................................................................................31

Missing Data.................................................................................................................................31

Variables.......................................................................................................................................32

**CHAPTER 5- METHODOLOGY: MODELING....................................................................41**

Linear Regression.........................................................................................................................42

Random Forest..............................................................................................................................43

K-Means Clustering......................................................................................................................43

Chi-Square...........................................................................................................................................47

Count Plot......................................................................................................................................50

Logistic Regression...........................................................................................................................59

**CHAPTER 6-MODEL EVALUATION...................................................................................65**

**CHAPTER 7- CONCLUSION..................................................................................................71**

Discussion.........................................................................................................................................

Applications......................................................................................................................................

Limitations and Future Research......................................................................................................

**APPENDIX..................................................................................................................................75**

**REFERENCES............................................................................................................................76**

**CHAPTER 1: INTRODUCTION**

**Background**

This paper examines the relationship between crime rates and public safety funding in Louisville, focusing on two key datasets: the "Louisville Metro Crime Rates from 2017 – 2021" and "Louisville’s Funds Report to Compensate Crime Victims" datasets (https://data.louisvilleky.gov/). These datasets reveal detailed aspects of Louisville's crime patterns and financial efforts to aid crime victims. Louisville's crime rates and funding strategies are examined within the broader context of evolving national and local public safety challenges, offering insights into both historical trends and national benchmarks. The Louisville Metro Crime Rates dataset, containing 306,634 records across nine variables, provides a comprehensive overview of various crime types and incidents. The Louisville’s Funds Report dataset, though smaller with six rows and 19 variables, offers an in-depth look at fiscal allocations for victim support. Collected on August 28th and September 11th, these datasets are presented in their unaltered state to maintain data accuracy, providing a well-rounded perspective on Louisville's crime and financial response strategies.

**Theoretical Framework and Analysis**

In understanding these complex dynamics, the study employs criminological theories such as Routine Activity Theory and Rational Choice Theory to interpret the relationship between crime rates and public funding. These theories help elucidate how criminal behavior might be influenced by funding policies and their implementation. Additionally, the research is grounded in public policy analysis, focusing on the allocation of resources for crime prevention and victim support. This policy perspective provides a crucial lens for examining how funding decisions are made and their consequent impact, bridging the gap between theoretical criminology and practical policy application in the context of Louisville's public safety landscape.

**Research Questions**

Central to this study are several pivotal questions to find parallels between public safety funding and crime rates through several research questions:

* How does the allocation of various types of public safety funding relate to distinct categories of crime rates?
* How do temporal patterns influence the prevalence of crime, with a focus on the variation across seasons, the incidence during holidays, and the fluctuation throughout the days of the week?
* Is there a time lag between the allocation of funds and its effect on crime rates?
* Are there lower crime rates in areas receiving more funding?

To address these, the study will deploy a range of statistical methods, predictive modeling, and geospatial analysis to rigorously examine the Louisville crime and funding datasets. While acknowledging the inherent limitations and assumptions of our data and methods, our goal is to provide a detailed yet holistic analysis. The findings are expected to offer significant insights for policymakers in Louisville, potentially shaping more effective public safety strategies. Academically, this research contributes to data analytics, criminology, and public policy, deepening our understanding of the relationship between crime rates and public funding. This introduction paves the way for a comprehensive exploration between thorough data analysis and meaningful policy implementation, with the following chapters diving into the methodologies, analyses, and results, all intended to enhance public safety strategies in urban environments.

**CHAPTER 2 – LITERATURE REVIEW**

**Crime History & Patterns**

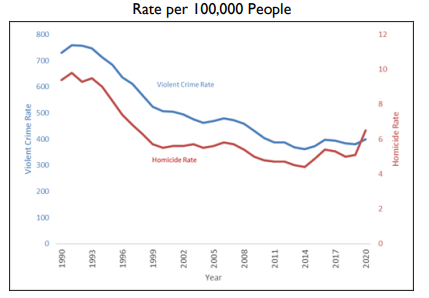
**Historical Context of Crimes**

Since the 1930s, the Federal Bureau of Investigation (FBI) and the Uniform Crime Reporting (UCR) program have collected and analyzed crime data from various law enforcement agencies at federal, state, local, and tribal levels. The UCR publishes this data, covering violent crimes such as homicide, rape, robbery, and aggravated assault. The program has been instrumental in tracking violent crime trends from 1990 to 2020(CRS Reports (2022, December 12)).

**Violent Crime Trends**

**Figure 2.1**

*Rates for Violent Crimes and Homicides with Years*



Note: Source: UCR Program

The graph above illustrates violent crime and homicide rates from 1990 to 2020. In 1991, the violent crime rate peaked at 758.2 per 100,000 people (about the seating capacity of the Los Angeles Memorial Coliseum) and fluctuated until 2004. Though there was a slight increase between 2005 and 2006, rates declined thereafter. The most recent surge occurred from 2019 to 2020, with rates moving from 380.8 to 398.5 per 100,000 people (about the seating capacity of the Los Angeles Memorial Coliseum). Notably, assault rates escalated from 250.4 to 279.4, while rates of rape and robbery dropped (CRS Reports (2022, December 12)).

**Homicide Trends**

The homicide rate also surged in 1991, recording 9.8 per 100,000 people (about the seating capacity of the Los Angeles Memorial Coliseum). Subsequent years saw a decline, except for minor upticks. From 2014 onwards, there was a slight increase in homicide rates, reaching 6.5 per 100,000 in 2020 (CRS Reports (2022, December 12)).

**Statistical Analysis for 2020-2021**

The FBI's analysis for 2020 and 2021 indicated a 1% decrease in violent crimes and a 4.3% increase in homicides, both statistically insignificant. Thus, crime rates remained relatively stable during these years (CRS Reports (2022, December 12)).

**Policy Considerations**

The UCR aims to modernize data reporting to align with the National Incident-Based Reporting System (NIBRS) through long-term strategies that address data gaps. The Justice Assistance Grant (JAG) program uses UCR data to allocate funds to governmental entities based on crime rates. Noncompliance with NIBRS adoption results in funding disqualification. Congress is exploring the use of incentives or penalties to encourage NIBRS adoption by law enforcement agencies (CRS Reports (2022, December 12)).

German Lopez (19 July 2016) addressed an unanswered question at a Republican convention focused on "making America safe again": how could nonpartisan policies reduce crime without resorting to gun control? Lopez (19 July 2016) began with an assumption against gun control due to political tensions, particularly from Republicans, even though his evidence suggested that high levels of gun ownership correlate with increased violence in the U.S.

Lopez identified six strategies for reducing crime:

* **Stricter Alcohol Policies-**Studies suggest a strong link between alcohol consumption and violent crimes. Lopez proposed solutions like higher alcohol taxes and reduced numbers of alcohol outlets.
* **Hot-Spot Policing-**Criminologist (David Kennedy) highlighted that crimes are often concentrated in specific areas. Increased police presence in these "hot spots" could lower crime rates.
* **Focused Deterrence Policing -**Kennedy also advocated for focused deterrence, where police concentrate on individuals with criminal records or suspicious activities, leading to low crime rates.
* **Raising the Dropout Age -**Extending the time children spend in school could potentially reduce crime, though the effectiveness of this policy is still debated.
* **Behavioral Intervention Programs-**Programs like the University of Chicago's Youth Guidance’s "Becoming a Man" aim to educate youth at risk of violent behavior.
* **Eliminating Blighted Housing-**Repairing or cleaning dilapidated buildings can also contribute to reducing crime.

**Economic Theories:**

Regarding economics and its relationship with crime rates, a nuanced correlation has been observed across various studies and analyses. Economic conditions do not operate in a vacuum but often interplay with other social, cultural, and political factors. This makes it complex to definitively conclude that economic downturns will unequivocally lead to a rise in crime. Nonetheless, several theories present compelling cases for the role of economics in influencing crime rates (Lehrer, E. (n.d.)).

One such theory comes from a professor at UCLA, who has postulated that declining wages catalyze criminal activities, particularly among the youth. In his theory, reduced income opportunities generate a sense of frustration and disillusionment. Young individuals may find that the expected returns on illegal activities, such as drug trading, outweigh the benefits of engaging in legitimate, low-wage work. This perspective aligns with a broader economic theory known as the "rational choice theory," which suggests that individuals weigh the costs and benefits of illegal activities against those of lawful employment (Lehrer, E. (n.d.)).

Jeff Grogger (n.d) theory becomes particularly noteworthy when considering the socio-economic environment that disadvantaged youths face. A backdrop of declining wages can synergize with other adverse conditions such as inadequate education, limited access to quality jobs, and social marginalization, further nudging these youths toward illicit activities as a perceived viable alternative. This raises critical questions about the policy interventions necessary to tackle the root causes of crime, which may be deeply embedded in economic structures (Lehrer, E. (n.d.)).

In summary, while a direct, causal relationship between economic conditions and crime rates is difficult to establish, theories like those proposed by Jeff Grogger (n.d)) provide valuable frameworks for understanding how economic factors, such as wage stagnation, can contribute to increased criminal activities. These insights warrant consideration in the formulation of both economic and criminal justice policies aimed at mitigating crime rates (Lehrer, E. (n.d.)).

**Patterns in Crime**

The relationship between environmental conditions and human actions is a subject of interest across multiple fields. In criminology, how crime rates change with the seasons is a significant focus. This change in crime rates, influenced by weather and holidays, offers vital information on the variables that impact criminal conduct. This literature review explores three major themes: the influence of weather-related factors, the effect of holiday seasons on crime rates, and the day of the week and crime rates.

**Weather-Related Influences on Crime Rates**

Several studies have elucidated the correlation between weather patterns, particularly temperature and precipitation. Higher temperatures and lower precipitation tend to escalate the rates of both violent and non-violent crimes as people are more active and interact more under such conditions (Kamide, 2021). The transient relationship between crime and weather reveals a substantial escalation in crime with rising temperatures. Moreover, the study found that crime rates tend to increase with temperature up to a certain point, after which they plateau or decline (Kamide, 2021). An investigation into the daily shootings in Chicago showed a significant association, where a 10-degree higher temperature was related to a 34% and 42% increase in shootings on weekdays and weekends or holidays respectively. Furthermore, understanding the impact of weather on crime necessitates a good comprehension of exogenous variables like weather and holidays which influence temporal trends in crime (Kamide, 2021).

**Holiday Seasons and Crime Rates**

Empirical evidence suggests a surge in particular crimes during holiday seasons, driven by the stresses and financial pressures associated with holidays. Contrary to common perception, property crimes peak during summer rather than holiday seasons (Reeping, 2020). However, there is a noticeable spike in theft and property crime rates during the holiday season, particularly between Thanksgiving and the New Year, attributed in part to the financial pressure to provide holiday gifts. The Criminal Victimization Survey of 2019 also noted a rise in personal larceny and robbery by approximately 20% during the holiday season (Reeping, 2020). The increase in crime rates by about 20% during November and December every year emphasizes the impact of holiday seasons on criminal activities (Reeping, 2020).

**The Day of the Week and Crime Rates**

The temporal distribution of crime extends to weekly cycles, where certain days exhibit higher crime rates. Research posits that weekends, particularly Fridays, often witness a higher incidence of certain crimes due to increased social interaction and alcohol consumption. Moreover, an analysis of 911 call records from a recent one-year period revealed variations in burglary calls on different days of the week, with Monday having the highest percentage of calls at 16.19% (Van Sleeuwen, 2021). A study on street robbery location choices in Chicago also explored how spatial crime patterns depended on the time of the day and the day of the week. Additionally, a temporal analysis of crime pairs for the hour of the day and hour of the week demonstrated varying crime rates across various times (Van Sleeuwen, 2021). Paydays, often falling at the end or beginning of a week, may also correlate with spikes in specific criminal activities. This topic intertwines with the previously discussed themes, as weekends and holidays may share common factors influencing crime rates.

**Budget Constraints and the Erosion of Law Enforcement Efficacy**

A perceived absence of criminal activity is often equated with societal stability, growth, and security. However, recent budgetary allocations have presented challenges for law enforcement agencies in fulfilling their mandated duties effectively. Research indicates that financial constraints significantly affect the ability of Police Departments to manage increasing crime rates effectively (Borjas, 2019).

The primary focus of this research is to scrutinize whether budgetary and manpower constraints of major Police Departments have a direct impact on the crime rates within their jurisdictions. Budget trends at federal, state, and local government levels will be compared with historical crime rates and Department of Justice statistics.

**Societal Impact of Crime Rates**

Changes in crime rates have a multifaceted impact on the well-being of a society. Crime is considered one of the most detrimental societal issues, albeit varying estimates regarding its social cost. Budgetary limitations not only restrict law enforcement from reaching ideal staffing levels but also impede the acquisition of essential equipment and ongoing officer training (Borjas, 2019).

**Resource Scarcity and Its Consequences**

The scarcity of resources poses risks not only to law enforcement personnel but also to local communities. Inadequate funding leads to an inability to recruit qualified staff and maintain consistent training programs. For example, a 2018 survey indicated that nearly 50% of Americans viewed gun violence as a significant concern.

**Budgetary Impact on Specialized Units**

Due to budget cuts, some police agencies have been forced to discontinue key units. One such example is the Vice unit, instrumental in operations that maintain public order and safety. A reduction in adequate funding also results in a high attrition rate among officers, as exemplified by the early retirement of more than 1,000 police officers in New York City since 2012 (Borjas, 2019).

**Local Examples and Police-to-Civilian Ratio**

The current budgetary plan of Los Angeles serves as a case in point, where the spending hardly suffices for routine law enforcement operations (Lin & Ming-Jen, 2009). The police-to-civilian ratio in LA, one officer for every 433 inhabitants, underscores the magnitude of understaffing.

**Budget Governance and Public Discourse**

Typically, local community police boards collaborate with other supporting agencies to design a police department's budget, as governance is generally local (Vaughan, A. D., & Andresen, 2018). In the 21st century, topics such as financial responsibility, policing costs, and budget stagnation have gained increasing public attention (Vaughan et al., 2018).

**The Complexity of Police Funding and Its Impact on Crime Rates in the United States**

**Is the Police Force Underfunded?**

The United States grapples with an enduring problem of criminal violence, indicated by a homicide rate far exceeding those in comparable wealthy democracies. Paradoxically, this comes amid ongoing calls to "defund the police" and allegations that the nation overspends on law enforcement. However, data shows that less than 1% of the GDP is allocated for policing, a figure that has notably declined post-Great Recession. The downward trend in police employment rates has been a concern since as early as 2019.

**The Role of Private Funding in Law Enforcement Research**

Amidst the shortfall in federal funding, private philanthropic organizations have played a significant role, particularly in research on gun violence (Morral).

**Focus on Prevention and Intervention**

Effective U.S. attorneys serve as problem solvers, employing a multifaceted approach that includes prevention initiatives. The Department's grantmaking bodies aim to bolster these efforts through improved communication and information-sharing regarding evidence-based strategies.

**Evaluation of Grant-Making Support**

Key offices within the Department of Justice, including the Office of Community-Oriented Policing Services and the Office of Justice Programs, continue to enhance efforts to combat gun violence and other violent crimes. These offices focus on research funding, preventative measures, and providing technical assistance to state and local entities. Legislation has further aided this, as exemplified by the 1998 revision of the Crime Enforcement and Accountability Challenge Grant Program, which allocated an additional $60 million for demonstration awards.

**Contested Budget Claims and the 'Defund the Police' Narrative**

Despite claims of budget reductions affecting public safety, an examination of 109 law enforcement budgets showed only eight were reduced by more than 2%, while 91 saw an increase of at least 2%. For instance, the Los Angeles Police Department's budget increased by 9.4% since 2019, yet the county sheriff contends that 'defunding has consequences.' A subset within law enforcement argues that anti-police rhetoric has had a more detrimental impact than actual budget cuts.

**The Correlation Between Police Budgets and Crime Rates**

The relationship between police department budgets and crime rates involves intricate variables. This study primarily aimed to examine the direct correlation between the number of officers and the available budget in relation to local crime rates. Existing literature broadly corroborates that fiscal constraints adversely impact the efficacy of policing strategies (Morral). Financial inputs come from a diverse range of sources including federal, state, and local governments, as well as private philanthropy, suggesting the need for a multifaceted financial strategy.

**Figure 2.2**

*Los Angeles County Sheriff total budget*

A graph of blue bars

Description automatically generated

The relationship between law enforcement funding and crime rates is complex and influenced by various factors, including public opinion and managerial efficacy. Despite the constraints imposed by inadequate resources, the literature generally supports the idea that budgetary limitations adversely impact the effectiveness of police departments (Cooper, 2003; Vaughan et al., 2018). Thus, a multi-pronged approach involving not just adequate funding but also its judicious utilization for preventive and intervention measures is imperative for effective policing.

The financial structure of a police department influences its capacity to effectively combat crime, but it's not the sole variable. Factors like public opinion and departmental efficiency also play pivotal roles. Although limitations in resources can constrain a department's functionality, the relationship between budget and crime rates is modulated by multiple factors. A nuanced approach, encompassing not only sufficient funding but also its judicious deployment for preventive and interventional strategies, is crucial for maximizing police effectiveness.

**The Economic Theory of Crime and Increased Police Force**

Mello's 2019 study, titled "More COPS, less crime," seeks to evaluate the effectiveness of increasing the number of police officers in reducing crime rates. This work falls under the broader framework of economic theories of crime, most notably Gary Becker's 1968 model, which posits that crime rates surge when economic conditions deteriorate. Mello's inquiry is particularly relevant given the significant rise in funding for the Community Oriented Policing Services (COPS) hiring grant program—from less than $20 million between 2005-2008 to $1 billion in 2009.

**Methodology and Allocation of Funds**

Mello employs a difference-in-differences model using an application score cutoff approach, effectively creating a quasi-experimental design to allocate grants. Cities scoring above the threshold received more funding and subsequently expanded their police forces. Importantly, these high-scoring cities were generally larger in population and exhibited higher crime rates. According to the COPS Universal Hiring Program (CHP) regulations, at least 1.5% of total funds must be allocated to each state and a minimum of 50% to areas with populations exceeding 150,000, thereby directing more funding to high-score areas.

**Spillover Effects**

Blattman (2017) suggests that bolstering police numbers in one jurisdiction may result in the displacement of criminal activity to neighboring areas, an important consideration when evaluating the overall effectiveness of increasing police presence.

**Longitudinal Trends and Economic Benefits**

Mello’s research observes a significant divergence in trends post-2009 between high- and low-scoring cities. For high-scoring cities that received funding, each additional police officer was associated with a reduction of four violent crimes and fifteen property crimes. Mello's cost-benefit analysis further revealed that the cost to a city of being a crime target would decrease by approximately $35 per additional officer. Strikingly, the benefits in terms of reduced crime amounted to nearly $350,000, far exceeding the governmental expenditure of $95,000 per officer hired.

**Implications and Policy Considerations**

Mello concludes that the COPS program represents a wise fiscal investment, especially during economically challenging times such as the Great Recession. Each added police officer not only directly contributes to reduced crime rates but also offers a compelling economic advantage, thereby aligning with Becker’s economic theory of crime.

**The Complex Relationship Between Police Employment and Crime**

Emily K Weisburst's study, titled "Safety in Police Numbers: Evidence of Police Effectiveness from Federal COPS Grant Applications," contributes to an ever-growing literature aimed at dissecting the relationship between police presence and crime. This is crucial in the United States, where incarceration rates outstrip the world average by 300%, yet police employment rates are 35% lower (Walmsley, 2016). Weisburst underscores the difficulty of isolating the causal effect of police presence on crime rates, as variations in crime rates are often both an antecedent and a consequence of police recruitment efforts.

**Methodological Rigor in Analyzing COPS Funding**

Weisburst employs a robust methodological approach using COPS (Community Oriented Police Services) hiring grants allocated between 2000 and 2014 as a natural experiment. The study incorporates the use of instrumental variables, controls for grant application evaluations, employs panel data analysis across a broad sample of municipalities, and scrutinizes diverse aspects of police hiring grants. This methodological rigor affords a granular understanding of the causal links between police staffing and crime.

**Statistical Significance and Policy Implications**

The findings are compelling, revealing significant elasticities of -1.28 for violent crime and -0.73 for property crime when correlated with an increase in police presence. However, the study stops short of delving into other significant outcomes of increased police hiring, such as the potential impact on racial bias or the use of force. Despite this limitation, Weisburst argues that these results can guide policy to effectively utilize police funding, not just to enhance community safety, but also to foster public confidence in law enforcement.

**Balancing Community Safety and Public Confidence**

The research concludes that an increase in police numbers significantly reduces severe crime types. However, it opens up a conversation for further research on the nuanced impacts of increased policing, such as community relations and the equitable use of force. Policymaking in this context, therefore, must be multifaceted, addressing both the public’s demand for safety and the imperative for equitable and just policing practices.

**Conclusion**

In conclusion, crime rates spiked in 2020 after decades of decline, likely exacerbated by the COVID-19 pandemic and economic downturn. However, non-partisan policies like those proposed by Lopez and supported by UCR data offer actionable solutions to mitigate this uptick in crime.  The seasonal patterns in crime rates underscore a complex interplay between environmental factors and human behavior. Understanding these patterns is pivotal for law enforcement agencies to allocate resources efficiently and for policymakers to formulate strategies aimed at crime prevention and community safety. The existing body of literature highlights the significance of weather conditions and holiday seasons in influencing crime rates, providing a pathway for further research to delve deeper into these correlations and their implications in the broader context of criminology and public policy. In summary, Mello’s 2019 study affirms the hypothesis that an increase in police presence leads to reduced crime rates, particularly in high-scoring cities. The economic benefits of this reduction present a compelling argument for the sustained or increased funding of community-oriented policing services. However, future research should account for potential spillover effects to provide a more comprehensive understanding of the program's overall impact. Weisburst's study offers valuable empirical evidence on the effectiveness of police presence in reducing crime rates. By leveraging COPS grants as a natural experiment and employing rigorous methodologies, the research establishes a significant causal link between increased police numbers and decreased crime. Nonetheless, the study also raises critical questions about the broader societal impacts of increased police staffing, thereby emphasizing the need for a balanced approach to police funding that prioritizes both community safety and public trust.

**CHAPTER 3 – METHODOLOGY: DATA PREPARATION**

**Software**

All data manipulation was done in Excel and analysis was performed using Python (Jupyter Notebook).

**Data Collection for Crime Rates**

The “Louisville Metro Crime Rates from 2017 - 2021 were extracted from the source data.louisvilleky.gov. The dataset was available from different sources from the same site. We collected the data from various sources and wrangled it into a single Excel sheet. Here is the process of how we wrangled the data:

**Figure 3.1**

*2017 Dataset*

A screenshot of a computer

Description automatically generated

[Louisville Metro KY - Crime Data 2017 | Louisville Kentucky Open Data (louisvilleky.gov)](https://data.louisvilleky.gov/datasets/LOJIC::louisville-metro-ky-crime-data-2017/about)

**Figure 3.2**

*2018 Dataset*

A screenshot of a computer

Description automatically generated

[Louisville Metro KY - Crime Data 2018 | Louisville Kentucky Open Data (louisvilleky.gov)](https://data.louisvilleky.gov/datasets/LOJIC::louisville-metro-ky-crime-data-2018/about)

**Figure 3.3**

*2019 Dataset*

A screenshot of a computer

Description automatically generated

[Louisville Metro KY - Crime Data 2019 | Louisville Metro KY - Crime Data 2019 | Louisville Kentucky Open Data (louisvilleky.gov)](https://data.louisvilleky.gov/datasets/LOJIC::louisville-metro-ky-crime-data-2019-1/explore)

**Figure 3.4**

*2020 Dataset*

A screenshot of a computer

Description automatically generated

[Louisville Metro KY - Crime Data 2020 | Louisville Metro KY - Crime Data 2020 | Louisville Kentucky Open Data (louisvilleky.gov)](https://data.louisvilleky.gov/datasets/LOJIC::louisville-metro-ky-crime-data-2020/explore)

**Figure 3.5**

*2021 Dataset*

A screenshot of a computer

Description automatically generated

[Louisville Metro KY - Crime Data 2021 | Louisville Metro KY - Crime Data 2021 | Louisville Kentucky Open Data (louisvilleky.gov)](https://data.louisvilleky.gov/datasets/LOJIC::louisville-metro-ky-crime-data-2021/explore)

The above screenshots where the different excel sheets and the links below them are the links from where we extracted the datasets.

**Data Wrangling Cleaning and Description for Crime Rates**

Once we had the complete data, we added it all to the on sheet and removed the variables that weren’t helpful for our research questions and could bring limitations, before deleting we had 15 variables, and 377894 the dataset looked like as shown in the screenshot below.

**Figure 3.6**

*2017-2021 Dataset Before omitting variables*

A screenshot of a computer

Description automatically generated

As in any analytical endeavor the significance doesn’t lie only in data but also lie in the ones that are omitted. Following are the omitted columns for the Louisville Crime Rate dataset:[1]

* NIBRS\_CODE: National Incident Based Reporting System.[1]
* UCR\_HIERACH: Uniform Crime Reporting.[1]
* ATT\_COMP: Attempted crimes status.[1]
* BLOCK\_ADDRESS: The exact address of the crime.[1]
* INCIDENT\_NUMBER: Unique identified code for the crime report.[1]
* BADGE-ID: It is the Badge Id number of the police officer. [1]

Here is the explanation behind omitting the columns, The columns such as BADGE\_ID and LMPD\_BEAT were removed as there was a susceptibility that those columns are highly multicollinear which can lead to raising concerns about violation of statistical assumptions inherited in many modeling techniques. BLOCK\_ADDRESS was omitted as it would lead to an ethical concern for the victim. NIBRS\_CODE and UCR\_HIERARCHY are omitted because these columns were domain-specific, they hold no interpretative power in the context of fiscal analysis and could make it complex for us to interpret analysis. [1]

***SOURCE-1: From our concept paper.***

This is how our data looked like after wrangling.

**Figure 3.7**

*2017-2021 Dataset After omitting variables*

A screenshot of a computer

Description automatically generated

**Variables:**

We had had 9 variables and 307661 rows; from 9 variables we had 5 categorical variables and 4 numerical variables. Below is the explanation of variables we will be using for the analysis:[1]

Object Id: Unique Identifier for the crime report. [1]

* DATE\_OCCURED: Actual date of the incident occurrence. [1]
* DATE\_REPORTED: Date when the incident was reported to LMPD. [1]
* CRIME\_TYPE: The category of crime type. [1]
* UOR\_DESC: Uniform Offense Reporting Code for the criminal act committed. [1]
* LMPD\_DIVISION: The division of LMPD in which the incident occurred. [1]
* PREMISE\_TYPE: Type of location in which the incident occurred. [1]
* City: The city associated with the incident block location. [1]
* ZIP\_CODE: The zip code associated with the incident block location.[1]

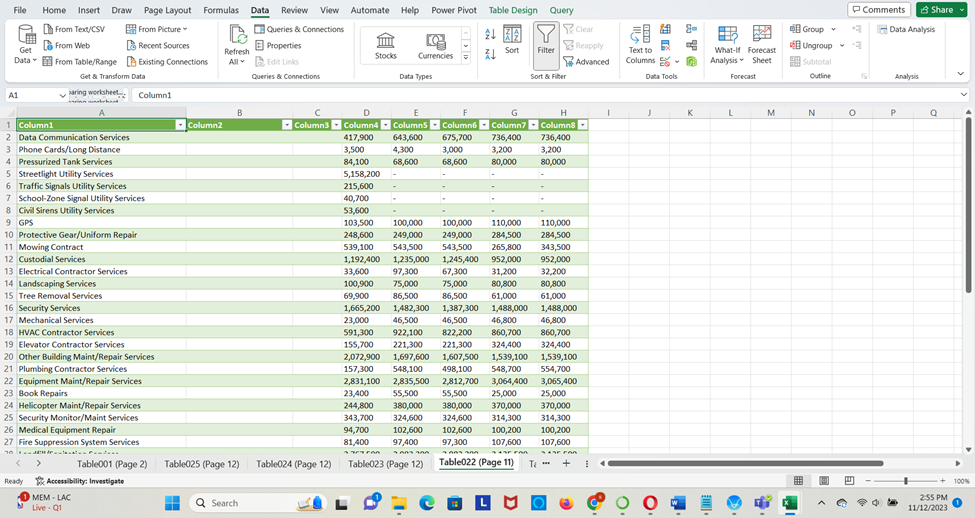
**Data Collection for Funding:**

Extraction of Funding Data:[**https://louisvilleky.gov/government/management-budget/fy21-budget**](https://louisvilleky.gov/government/management-budget/fy21-budget)

We have tried to clean the data using Excel. It is challenging to clean the data using Excel because we need to extract the data from multiple documents.

**Figure 3.8**

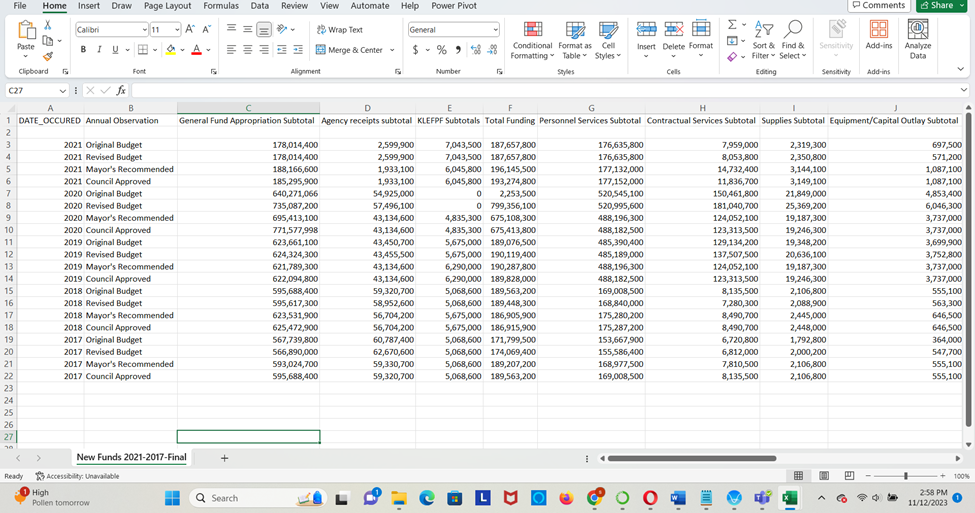
*2017-2021 Dataset before Omitting Variables*



Similar to all the above screenshots we have multiple sheets. We don’t need all those details. Hence, we have extracted the data we need manually from the funding data documents.

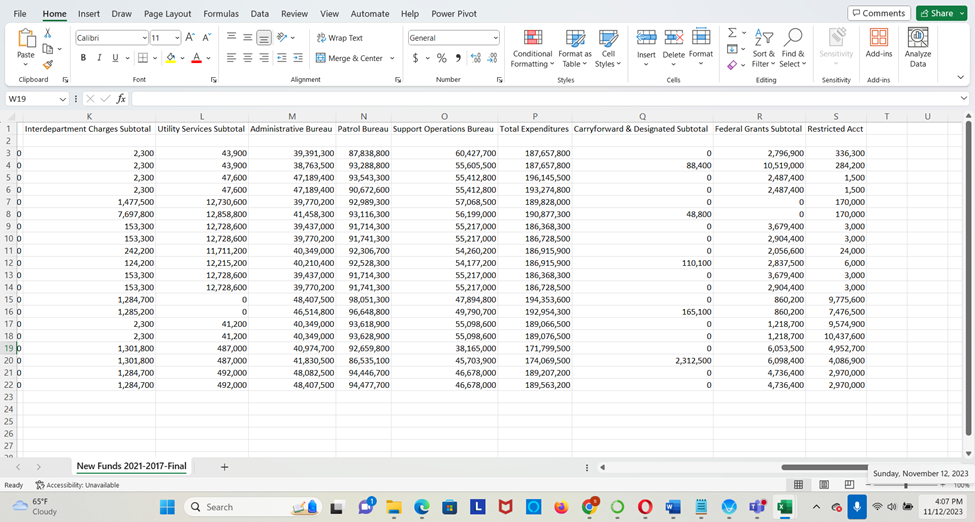
**Figure 3.9**

*2017-2021 Dataset after Omitting Variables*



**Figure 3.10**

*2017-2021 Dataset after Omitting Variables*



**Data Wrangling Cleaning and Description for Funding:**

Louisville's general fund derives from various components for different fiscal periods. The data encompasses a range of variables that shed light on annual observations, appropriations, receipts, subtotals, and expenditures, among other elements. [1]

* Annual Observation: Prior Year Actual.[1]
* General Fund Appropriation Subtotal: Total amount allocated to the general fund.[1]
* Agent's receipts subtotal: Receipts collected by agents. Taxes.[1]
* KLEFPF Subtotals: Mandatory Training.[1]
* Total Funding: Grand total of funding.[1]
* Personnel Services Subtotal: Personnel services, such as salaries and benefits.[1]
* Contractual Services Subtotal: Indicates the total amount spent on contractual services.[1]
* Supplies Subtotal: Represents the total expenditure on supplies.[1]
* Equipment/Capital Outlay Subtotal: Total amount spent on equipment/capital outlay. [1]
* Interdepartmental Charges Subtotal: Charges that occur between departments. [1]
* Utility Services Subtotal: Represents the total expenditure on utility services. [1]
* Administrative Bureau: Expenditure allocated to the administrative bureau. [1]
* Patrol Bureau: Expenditure allocated to the patrol bureau. [1]
* Support Operations Bureau: Expenditure allocated to the support operations bureau. [1]
* Total Expenditures: Grand total of all expenditures. [1]
* General Fund Appropriation Subtotal: General fund subtotal. [1]
* Carryforward & Designated Subtotal: Donations giving for the year. [1]
* Federal Grants Subtotal: Funds received from federal grants. [1]
* Restricted Acct: Indicates specific uses or emergencies. [1]

In summary, the dataset provides a comprehensive view of the fund allocation and expenditure across different bureaus and categories in Louisville for various fiscal years. [1]

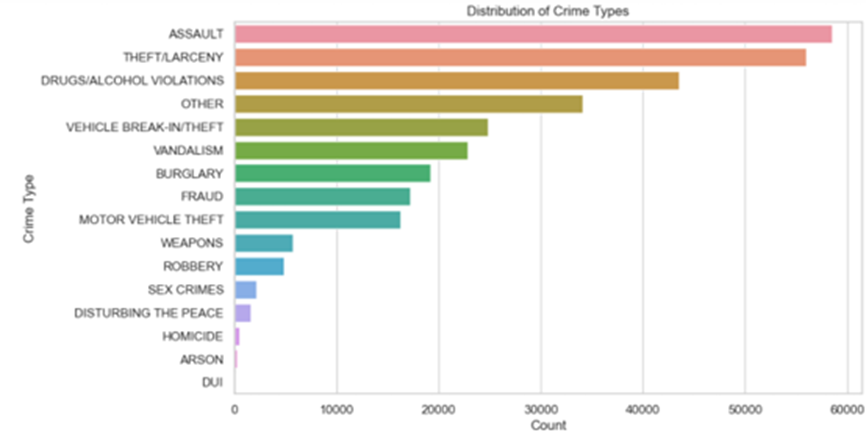
**CHAPTER 4 – METHODOLOGY: EXPLORATORY DATA ANALYSIS**

**EDA and Methods for Crime Rates**

We are checking for the Distribution of Crime types and from the graph we can say that ASSAULT had the highest number of any other crimes and DUI (Driving Under the Influence) had the least number.

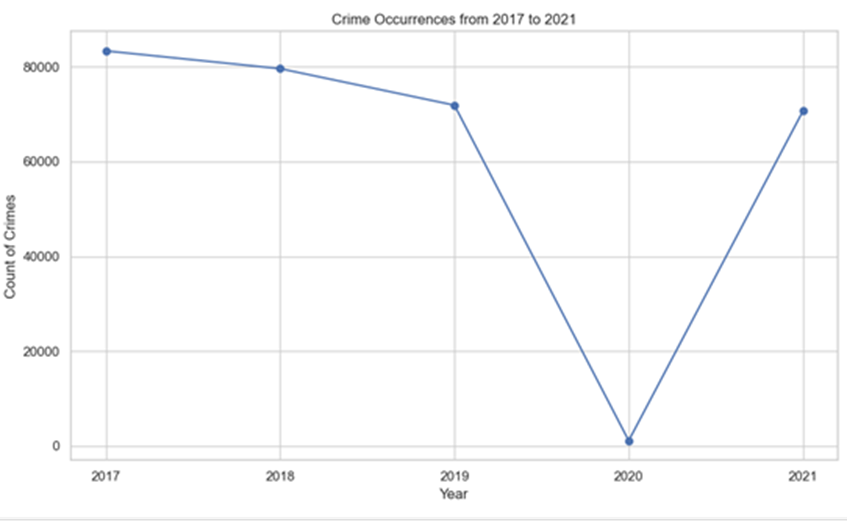
**Figure 4.1**

*Distribution of Crime Types*



**Figure 4.2**

*Crime Rates from 2017- 2021*



Here we looked for the crime rates over the years and found that there was a decrease in crime rates from 2019-2020 as the lockdown was imposed but from 2020-2021 the crime rates elevated but still the uptick was not as much as it was during 2017.

**Missing Values**

We looked for the missing values and found 1478 missing values for the following variables.

**Figure 4.3**

*Checking for missing values*

A white rectangular object with a black border

Description automatically generated with medium confidence

So, we replaced it with zero because, replacing it with zero would show us that for the crime type the DATA\_OCCURED, DATA\_REPORTED, PREMISE\_TYPE, City, ZIP\_CODE were not recorded.

**Figure 4.4**

*Replacing missing values with zero*

A close-up of a computer screen

Description automatically generated

**Outliers**

As most of the columns weren’t numerical, we didn’t consider checking for their outliers, but we did check the outliers for ZIP\_CODE the results for it.

**Figure 4.5**

*Checking for outliers*

A white background with text

Description automatically generated

A graph with dots and numbers

Description automatically generated

**Funding Methods and EDA**

In this comprehensive analysis, we delved into a dataset covering financial records from 2017 to 2021. The dataset primarily focuses on various financial metrics related to a specific agency or department, offering a detailed view of budget allocations, expenditures, and other monetary aspects. The dataset comprises 21 entries and 20 columns, encompassing a range of financial data. Each entry represents a unique record, with columns detailing various financial metrics such as fund appropriations, agency receipts, expenditures across different bureaus, and more.

**Missing Data**

The dataset predominantly consists of object data types, with 'Year' being the only float64 type. This indicates that most data are categorical or string-based, including financial figures that were initially formatted as strings. Upon initial inspection, we noted missing values in several columns. These were addressed by filling numeric columns with zeroes and categorical columns, like 'Annual Observation', with 'Unknown'. This approach ensured completeness and consistency in the dataset. Outlier detection, particularly in 'Total Expenditures', was conducted to identify any unusual or aberrant data points. These outliers are essential for highlighting anomalies that could indicate significant financial events or discrepancies needing further investigation.

**Figure 4.6**

*Checking for Missing Values*

A graph with different colored squares

Description automatically generated

**Variables**

The dataset doesn't explicitly specify a response variable; however, 'Total Funding' and 'Total Expenditures' could serve as primary metrics for analyses, such as trend analysis or budget forecasting. The dataset includes categorical variables like 'Annual Observation', providing descriptive labels for each financial year. These variables are crucial for classifying and segmenting the data for specific analyses. Most of the variables in the dataset are numeric, representing various financial figures. These include 'Year', financial subtotals, and totals across different categories. The numeric nature of these variables allows for statistical analysis and trend observation.

**Figure 4.7**

*Variables*

A screenshot of a computer program

Description automatically generated

A statistical summary of the dataset reveals insights into the numeric variable 'Year', with a count of 20, a mean year of 2019, and a standard deviation of approximately 1.45. This summary underscores the time frame of the financial data. This analysis provided a thorough understanding of the dataset, highlighting key aspects like missing data, outliers, and the types of variables present. The comprehensive approach to data cleaning and formatting laid a robust foundation for further analysis, ensuring accuracy and reliability in the findings. The insights gathered here will be instrumental for subsequent analyses, including trend observation, budget comparisons, and anomaly detection, contributing to informed financial planning and decision-making.

The 'Budget Comparisons' aspect of the Exploratory Data Analysis delves into contrasting the budgeted amounts against the actual expenditures recorded in the dataset spanning from 2017 to 2021. This comparison is crucial in understanding how well the budgeting aligns with the actual financial execution. The dataset, meticulously cleaned and formatted for accuracy, includes key columns such as 'General Fund Appropriation Subtotal' (representing the budgeted amount) and 'Total Expenditures'. The methodology adopted for this analysis primarily involved line plots, which succinctly illustrate the relationship between planned and actual spending over the years. These visualizations provided a clear and concise representation of the financial management efficacy. The comparison revealed how closely the actual spending adhered to the budgeted amounts each year. Discrepancies, whether as over-expenditures or under-expenditures, were readily apparent, offering insights into the financial discipline and adaptability of the agency or department.

Key findings from this analysis highlighted years where significant variances occurred between the budgeted and actual figures. Such deviations could indicate a variety of factors including unexpected financial demands, changes in policy, or shifts in departmental priorities. Understanding these differences is critical for refining future budgeting processes and improving financial accuracy and reliability. The insights gained from the budget comparisons underscore the importance of robust and flexible financial planning. The ability to adapt to changing circumstances while maintaining fiscal responsibility is a key takeaway from this analysis. For future planning, it is recommended to investigate the causes behind significant budget variances to enhance the accuracy of future budget forecasts. In conclusion, the Budget Comparisons analysis offers an in-depth look at financial planning efficacy, revealing areas of strength and opportunities for improvement. It plays a crucial role in financial oversight and forms an essential part of comprehensive fiscal management and strategic planning.

**Figure 4.8**

*Budget vs Expenditure Over Years*

A graph with a line and a line

Description automatically generated

The 'Bureau-wise Analysis' forms a significant part of our Exploratory Data Analysis, focusing on the financial dataset that chronicles records from 2017 to 2021. This analysis is aimed at understanding how funds are allocated and utilized across different bureaus or departments within the organization. The dataset, after undergoing rigorous cleaning and formatting, includes detailed financial information across various bureaus, such as the 'Administrative Bureau', 'Patrol Bureau', and 'Support Operations Bureau'. Our approach in this analysis involved leveraging line plots to visualize the expenditure trends of these bureaus over the years. This method effectively illustrates the comparative financial trajectory of each bureau, providing a clear visual representation of their respective budget utilization. By examining these trends, we were able to discern how financial priorities and allocations have shifted over the years, which is crucial in understanding the operational focus and changes within the organization. One of the primary observations from this analysis was the identification of distinct financial patterns and trends for each bureau. These patterns offer insights into the operational dynamics and resource allocation strategies within the organization. Notably, the analysis helped in highlighting bureaus with increasing or decreasing financial trends, which could be indicative of strategic shifts, evolving operational demands, or changes in resource prioritization.

The insights from this bureau-wise analysis are instrumental in several ways. Firstly, they help in identifying the financial health and stability of each bureau. Secondly, they allow organizational leaders to pinpoint areas requiring financial adjustments or additional resources. For future planning, a deeper dive into the causes behind significant trends in each bureau could be beneficial. It would enable a more nuanced understanding of the operational effectiveness and strategic planning within each bureau. In conclusion, the Bureau-wise Analysis provides a comprehensive view of the financial distribution and trends across different departments within the organization. It is an essential tool for evaluating operational efficiency, financial management, and strategic alignment across the various bureaus, thereby forming a crucial component of organizational analysis and planning.

**Figure 4.9**

*Expenditure Distribution Among Bureaus Over Years*

A graph of a number of years

Description automatically generated with medium confidence

The 'Correlation Analysis' is a vital part of our Exploratory Data Analysis, focusing on uncovering relationships between different financial variables in the dataset, which spans from 2017 to 2021. This segment of analysis specifically aims to discern any significant correlations between key financial metrics such as 'Total Funding' and 'Total Expenditures'. Properly cleaned and formatted, the dataset allows for an accurate assessment of these relationships. Our methodology for this analysis involved the use of scatter plots, a powerful tool for visualizing and assessing the strength and direction of relationships between two variables. In this context, scatter plots were utilized to plot 'Total Funding' against 'Total Expenditures', providing a visual representation of how these two variables interact with each other. This visual exploration is essential for identifying patterns, trends, or anomalies in the data that might not be apparent through a cursory glance or simple statistical summaries.

One of the key findings from this correlation analysis was the understanding of how closely related the funding and expenditures are. A strong correlation would suggest that as funding increases, expenditures also increase proportionally, and vice versa. This relationship is crucial for financial planning and management, as it helps in forecasting and budgeting. The scatter plot can also highlight any outliers or deviations from the general pattern, which could warrant further investigation. The insights garnered from this analysis are invaluable for the financial oversight of the organization. Understanding the relationship between funding and expenditures assists in making informed decisions regarding budget allocations and financial planning. It also aids in identifying areas where financial efficiency can be improved.

For future planning, it might be beneficial to conduct a more in-depth statistical analysis, such as calculating the correlation coefficient, to quantify the strength of the relationship between these variables. Additionally, exploring correlations between other financial variables could provide a more holistic view of the organization's financial dynamics. In conclusion, the Correlation Analysis is a critical component of the financial EDA, providing key insights into the interplay between different financial metrics. This analysis not only aids in current financial assessment but also serves as a foundational tool for future financial strategy and planning within the organization.

**Figure 4.10**

*Correlation Between Total Funding and Total Expenditure*

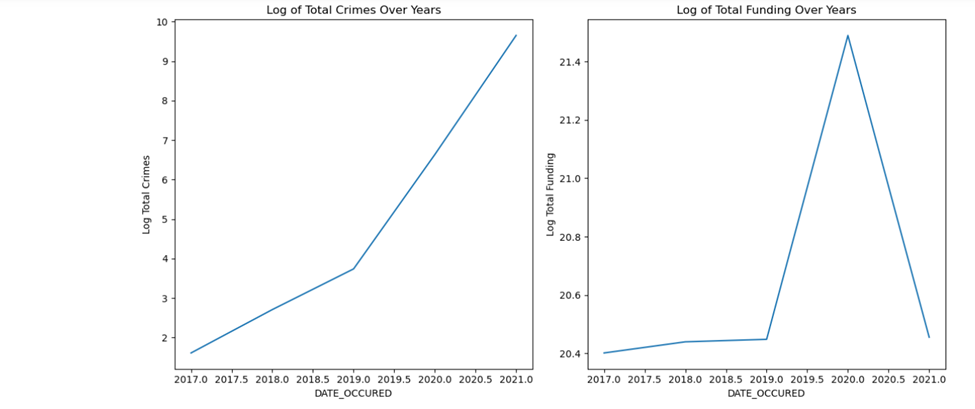
A graph with blue dots

Description automatically generated

**Time series analysis**

**Figure 4.11**

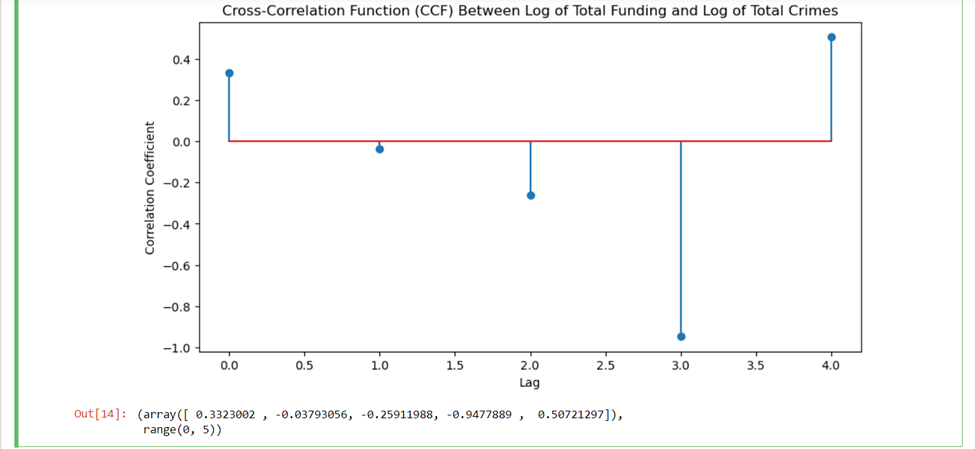
*Log of Total Crimes over years with Date Occurred*



The above-left graph shows a steady increase in total crimes from the year 2017 to 2021. The right graph shows relatively stable funding from 2017 to just before 2020 and there is a spike around 2020. The spike represents a one-time increase in funding.

**Figure 4.12**

*Cross-Correlation Function between Log of Total Funding and Log of Total Crimes*



There is a significant positive correlation at lag 0. It indicates that the funding and crimes are directly related to the same time point. There are smaller and negative correlations at lag 1 and lag 2, which shows that funding is shifted by one- or two-time units. There is a very strong negative correlation for lag 3. This shows the perfect inverse relationship between funding and crimes.

The negative peak suggests that there appears to be a temporal lag before the distribution of funding appears to have the greatest impact on crime rates. However, it's challenging to determine the precise duration of this lag without further context about the data. It shows a temporal relationship where the effects of changes in funding are strongly correlated with changes in crime rates after a certain period.

**CHAPTER 5 – METHODOLOGY: MODELING**

To answer our first research question: How does the allocation of various types of public safety funding relate to distinct categories of crime rates?

First, we started by seeing the correlation between Public Funding and Crime Categories.

**Figure 5.1**

*Correlation Matrix Between Public Funding and Crime Categories*

A screenshot of a computer

Description automatically generated

Based on the graph we can conclude that all the crime types are positively correlated but Total Funding is negatively correlated to all the crime types.

**Linear Regression**

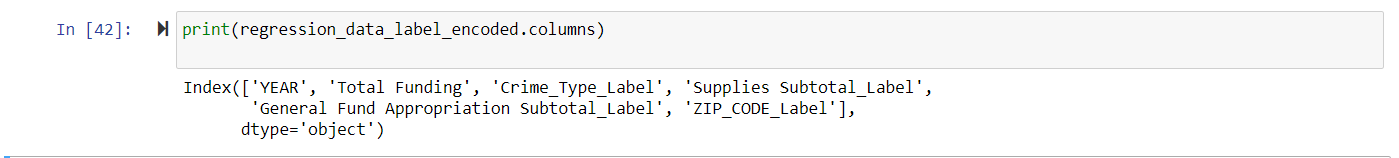
To find the correlation between public funding and various crime types, we started with Linear Regression, as we had two datasets, we merged them together and performed Label Encoding on the CRIME\_TYPE variable to make it numerical to get good results.

**Figure 5.2**

*Merging Datasets and Label Encoding*

A screenshot of a computer

Description automatically generated



**Figure 5.3**

*Linear Regression Output*

A screenshot of a computer program

Description automatically generated

We performed Linear Regression for the variables General Fund Appropriation, Subtotal Label and Crime Type Label. Because the Total Funding had negative correlation with crime types, we selected these variables. For Linear Regression, we divided the dataset into test and train, then moved forward with calculation and the values of MSE and R- R-squares.

**Random Forest Classifier Algorithm**

As our previous model wasn’t a good fit, now we proceeded with the Random Forest Algorithm.

**Figure 5.4**

*Random Forest Classifier Algorithm*

A group of numbers on a white background

Description automatically generated

For the Random Forest Algorithm, we used the variables Total Funding, Year, and CRIME\_TYPE as the other variables that we used for linear regression didn’t give us good results, so this type we tried different variables that aligned with the research question. We did Label Encoding and merged both datasets then divided it into test and train, performed a random forest algorithm on it, and printed the results.

**K-means Clustering Analysis**

The above two results didn’t provide us the satisfying results, so we wanted to try another modeling before drawing a conclusion about the research question, so we opted for clustering analysis. For clustering analysis, we did One Hot Encoding for the CRIME\_TYPE and merged the datasets, then standardized the variables selected the cluster number as 3, Applied K-Mean Clustering, and printed its results.

**Figure 5.5**

*K-Means Clustering Analysis*

A white background with black text

Description automatically generated

A white background with black and white text

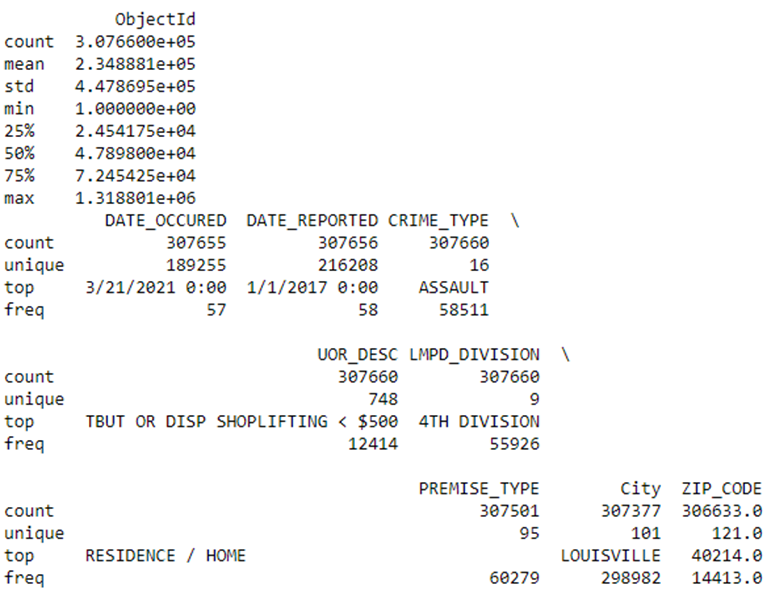
Description automatically generated

To answer the second question: How do temporal patterns influence the prevalence of crime, with a focus on the variation across seasons, the incidence during holidays, and the fluctuation throughout the days of the week?

The dataset under examination will be reviewing crime rate correlations in Lewisville, which covers multiple aspects of criminal activities. It includes complete timestamp data, with 307,655 occurrences and 307,656 report dates, encompassing a wide range of dates. Crime types are well-represented with 16 unique categories, notably dominated by 'ASSAULT'. Every record in the dataset contains a crime description, totaling 748 unique descriptions, with 'TBUT OR DISP SHOPLIFTING < $500' being the most frequent. The dataset includes information on 9 unique police divisions, with the '4TH DIVISION' being the most common. Additionally, it has extensive data on-premises types, predominantly 'RESIDENCE / HOME', and covers 101 cities and 121 zip codes, with 'LOUISVILLE' and '40214' being the most common city and ZIP code, respectively. This comprehensive dataset offers valuable insights for time-series analysis, crime pattern analysis, geographical analysis, and strategic crime prevention efforts, particularly in residential areas.

**Figure 5.6**

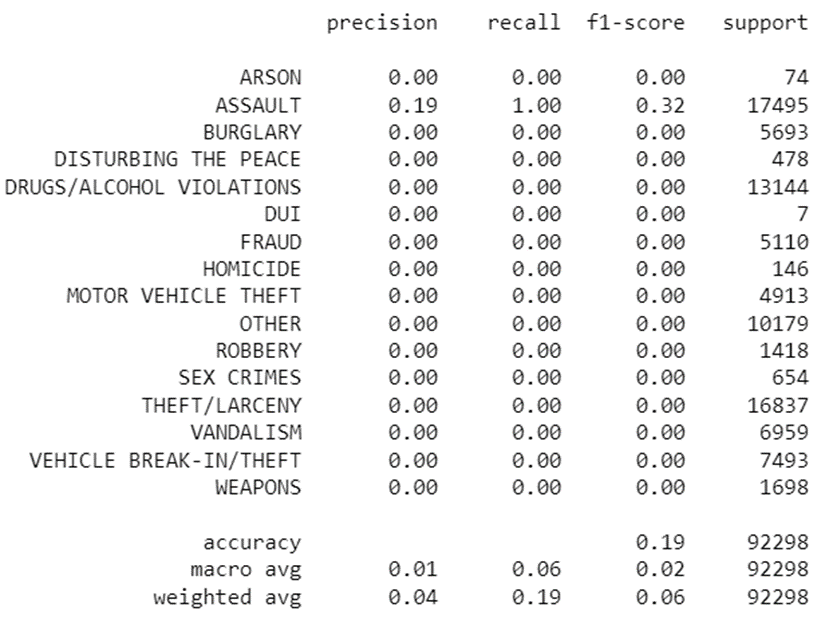
*Crime Dataset Description*



The classification report reveals critical insights about the model's performance in crime type prediction. Notably, the model excels in identifying 'ASSAULT' cases, reflected in its perfect recall of 1.00, but suffers from a low precision of 0.19, suggesting a high number of false positives. For all other crime types, the model fails entirely, with both precision and recall at 0.00. Overall, the model achieves an accuracy of 0.19, heavily skewed by its success in predicting 'ASSAULT'. The macro average scores (precision: 0.01, recall: 0.06, F1-score: 0.02) highlight the model's poor performance across all classes, while the weighted averages (precision: 0.04, recall: 0.19, F1-score: 0.06) offer a slightly better but still inadequate view, considering class imbalances. These results suggest a pronounced bias in the model towards 'ASSAULT', potentially due to class imbalances or unique features of this crime type. Additionally, the model's inability to distinguish other crime types, points to potential issues in learning or feature similarity. The impact of class imbalance is evident, leading to high accuracy in the dominant class but poor overall model performance.

**Figure 5.7**

*Classification Crime Prediction*



**Chi-Squared**

Three separate Chi-Squared tests were conducted to explore associations between the types of crimes committed and three different variables: the season of the year, the day of the week, and whether the day was a holiday or not. The results of these tests provide insights into whether the distributions of crime types vary significantly across these different time-related categories.

**Crime Type Vs Season**

The Chi-Squared test for crime type versus season yielded a Chi-Squared value of approximately 647.753 with a highly significant p-value of about 9×10−108. This suggests a statistically significant association between the type of crime committed and the season of the year. In other words, the likelihood that the observed distribution of crime types across different seasons is due to random chance is extremely low. This results in certain types of crimes that may be more prevalent in specific seasons compared to others.

**Crime Type Vs Day of the Week**

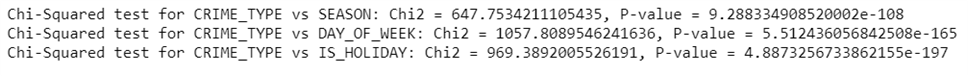
The analysis of crime type versus the day of the week produced a Chi-Squared value of approximately 1057.809 with a p-value of 5.51×10−165. Which shows a statistically significant relationship between the type of crime and the day of the week. It implies that certain crimes may occur more frequently on specific days of the week. The very low p-value rejects the null hypothesis that the distribution of crime types is uniform across all days of the week.

**Crime Type Vs Holiday Status**

The test comparing crime type with holiday status resulted in a Chi-Squared value of approximately 969.389 and an exceedingly small p-value of 4.89×10−197. This strongly suggests that different types of crimes are significantly associated with whether the day is a holiday. This could mean that certain crimes are more likely to happen on holidays, or conversely, some types of crimes might decrease on these days.

**Figure 5.8**

*Chi-squared for all observations*



The results of these statistical tests provide strong evidence that the type of crime committed is not uniformly distributed across different seasons, days of the week, or holiday statuses. The significant p-values in all tests indicate that these variations are unlikely to be due to chance. Such findings can be invaluable for law enforcement and public safety agencies in planning and resource allocation, tailoring crime prevention strategies to specific times when certain crimes are more likely to occur.

**Day of the Week**

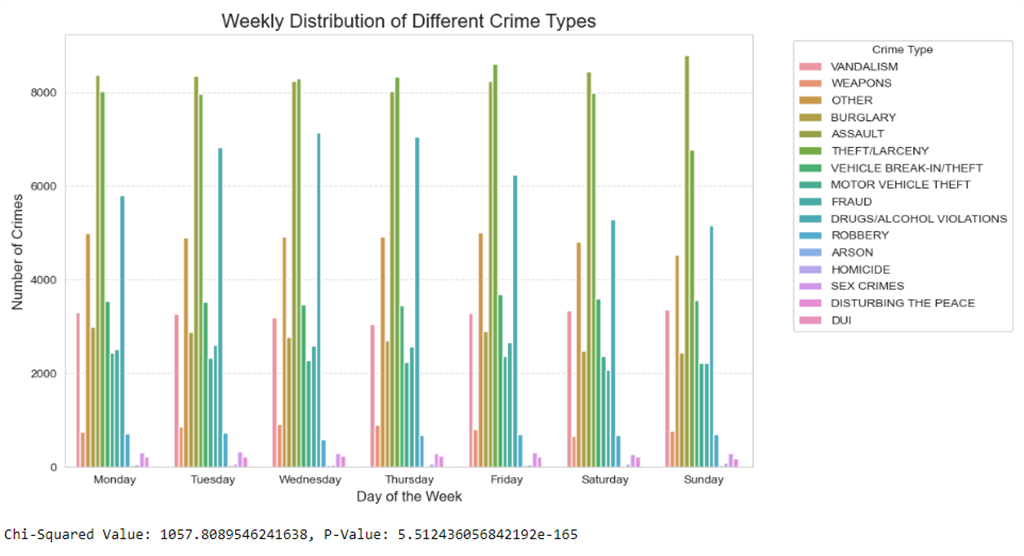
The count plot presented visualizes the frequency of various crime types across different days of the week, offering an intuitive grasp of daily crime trends. The data indicates that 'THEFT/LARCENY' incidents are the most frequently occurring crime, consistently high from Monday to Sunday. However, there are discernible patterns throughout the week, such as the peak in 'DRUGS/ALCOHOL VIOLATIONS' on weekends, hinting at a correlation between leisure activities and substance-related offenses. Additionally, 'ASSAULT' cases show substantial numbers, with slight variations, but remain a significant concern every day. In contrast, 'HOMICIDE' remains the least frequent crime but shows a slight uptick during the weekend.

The midweek days - Tuesday and Wednesday - exhibit a noteworthy surge in 'FRAUD' cases, suggesting a pattern that could be related to the typical scheduling of financial transactions and activities. Meanwhile, 'DUI' offenses rise markedly towards the end of the week, culminating on Saturday, possibly reflecting increased social gatherings and alcohol consumption. 'SEX CRIMES' and 'ROBBERY' maintain a relatively steady occurrence throughout the week, with no substantial spikes on any particular day, implying a lack of strong temporal influence on these crimes.

This visual representation serves as a foundation for hypotheses regarding behavioral patterns associated with different crimes. For example, the increase in 'VANDALISM' and 'WEAPONS' violations as the week progresses could be linked to escalating social interactions and conflicts.

**Figure 5.9**

*Weekly Distribution of Different Crime Types*



**Seasonal Crime Distribution**

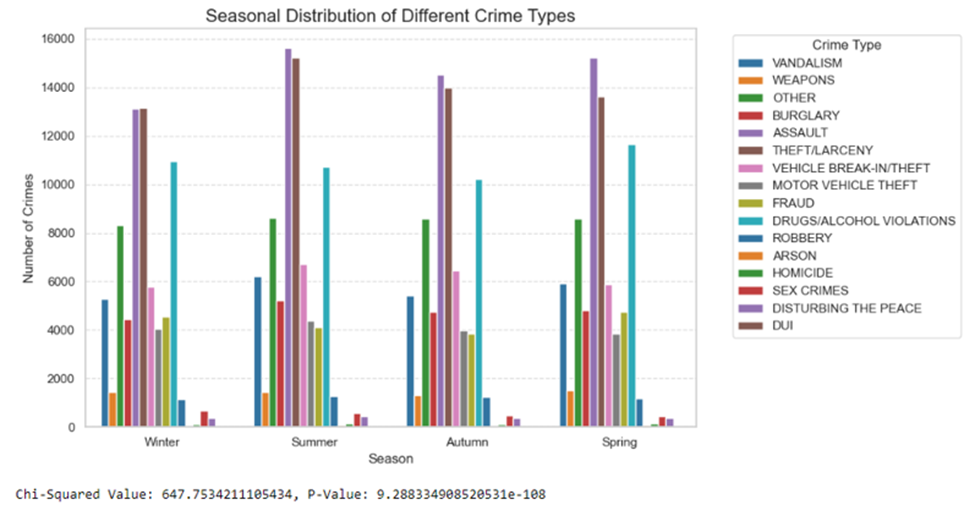
An additional count plot helps provide a visual representation to explore the correlation between seasonal changes and the frequency of assaults. According to the data presented, Assault is one of the most common crime types across all seasons. However, it is observable that the frequency of assaults rises significantly during the summer season.

This pattern could be attributed to various factors associated with warmer weather, such as increased social interaction, more outdoor activities, and the influence of elevated temperatures on human behavior. The increased frequency during the summer months may suggest that there is a higher propensity for conflicts that can escalate to assaults when people are more likely to gather in public spaces, and the discomfort from the heat can exacerbate tensions. The statistical significance of the observed seasonal trend in assaults can be further substantiated through analytical methods such as the Chi-Squared test for independence. A low p-value from such a test would indicate that the observed seasonal distribution of assaults is unlikely to be due to random chance, thereby supporting the hypothesis that assaults are more frequent during warmer seasons.

Overall, the count plot suggests a correlation between warmer seasons and the occurrence of assaults. The marked increase in assaults during summer can inform law enforcement and public safety agencies in strategic planning and preventive action. Further statistical analysis would be necessary to confirm the strength and significance of this correlation.

**Figure 5.10**

*Seasonal Distribution of Different Crime Types*



**Holidays Vs Non-Holidays**

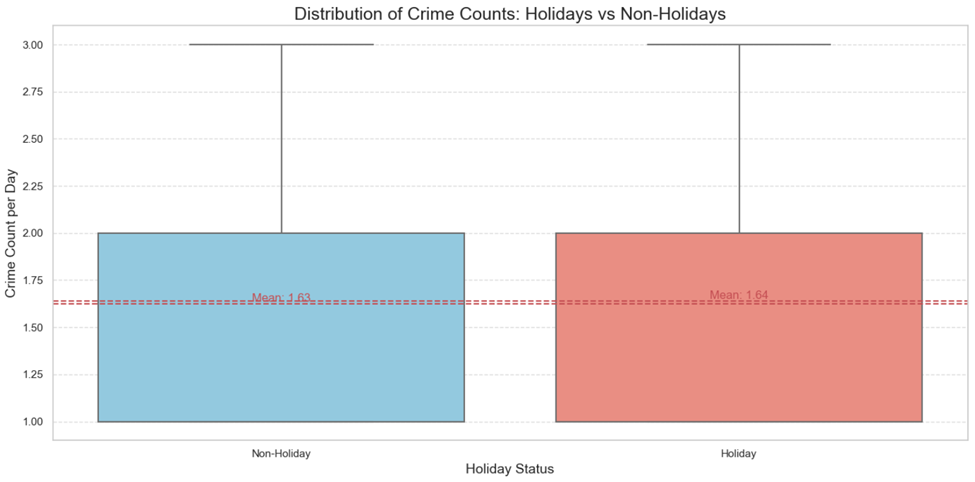
The boxplot analysis focuses on the distribution of crime counts comparing holidays to non-holidays. It features two bars—one blue for non-holidays and one red for holidays—each representing the average daily crime count. The mean crime count is annotated with horizontal dashed lines, showing a mean of approximately 1.63 crimes per day for non-holidays and a slightly higher mean of 1.64 for holidays. Despite this marginal difference, the similarity in height between the two bars suggests an almost equivalent average number of crimes per day during holiday and non-holiday periods. This small variance implies a potential, albeit minor, increase in crime on holidays compared to non-holidays.

The chart does not provide evidence of statistical significance, and the proximity of the means suggests that any difference in crime rates may not be substantial in practical terms. A statistical hypothesis test, such as a t-test for independent samples, would be necessary to determine if the observed difference is statistically significant. From a practical perspective, the data indicates that crime rates do not undergo a dramatic shift during holidays. This may suggest that specialized law enforcement measures on holidays might not need to deviate significantly from regular days in terms of resource allocation for crime prevention.

In terms of policy and resource allocation, these initial findings could imply that policing resources may not require substantial adjustments during holiday periods. However, more detailed data and additional statistical testing would be prudent to determine the significance and consistency of the observed difference in relation to different crime categories and specific holidays. In summary, the boxplot suggests only a negligible distinction in the average daily crime count between holidays and non-holidays. For policymaking and law enforcement resource distribution, it is crucial to dive deeper into granular data and appropriate statistical methods to affirm if the slight variation observed holds true across diverse crime types and specific holidays.

**Figure 5.11**

*Distribution of Crime Counts: Holidays vs Non-Holidays*



**PCA: Assaults Vs Season**

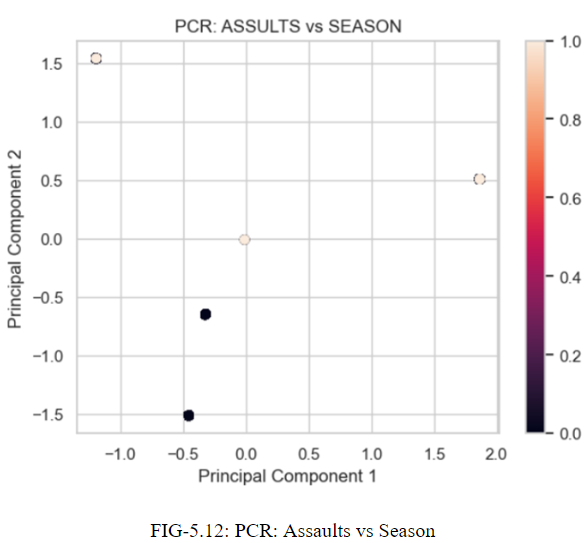
The Principal Component Analysis (PCA) plot for assaults versus season condenses data into two principal components, simplifying the exploration of the dataset's variance. The first principal component (PC1) along the x-axis captures the maximum variance and may correspond to the most influential variables in the data related to assaults and seasons. The second principal component (PC2) along the y-axis, orthogonal to PC1, reveals additional, albeit less, variance.

The plot shows data points transformed from the original dataset, distributed mainly towards the lower left quadrant, suggesting a potential clustering. The color gradient of the points, from light to dark, may indicate a variable of interest, such as the frequency or intensity of assaults, with darker shades potentially representing higher values. A notable observation is the concentration of data points in a specific area of the plot, which could imply that a significant amount of the variability in the assault data is explained by these two components. Additionally, a few points are distant from the main cluster, indicating potential outliers or unique seasonal patterns.

The PCA plot implies that it is possible to reduce the complexity of the dataset by focusing on these two dimensions, which may retain the essential information about seasonal variations in assaults. Although the plot hints at seasonal patterns, it does not provide a definitive mapping of these patterns to the principal components. To further understand the dataset, it is recommended to analyze the correlation between the individual seasons and the principal components, possibly through additional plots or statistical tests. Outliers should be examined to ascertain whether they represent anomalies or data entry errors. Moreover, this PCA plot offers a visual abstraction of the relationship between assaults and seasonal trends, indicating an underlying structure that may be pertinent for more in-depth analysis or modeling. Further analytical steps are required to draw concrete conclusions about the specific impacts of seasonal changes on the incidence of assaults.

**Figure 5.12**

*Assaults Vs Season*

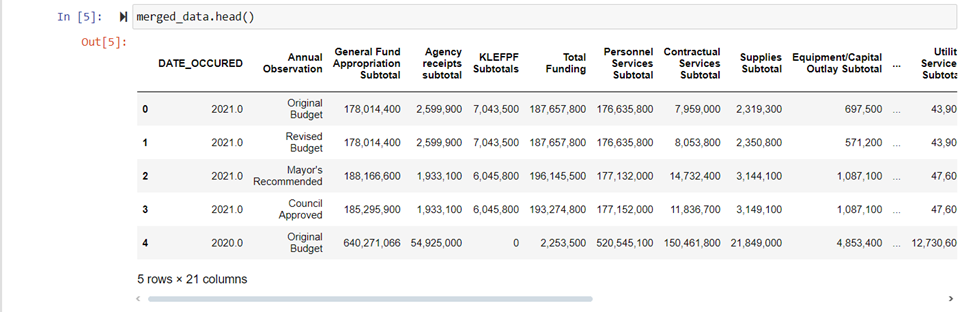


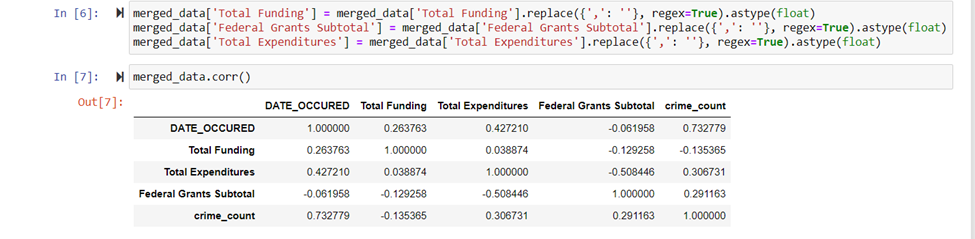
To answer the third research question: Is there a time lag between the allocation of funds and its effect on crime rates?

We have converted the crime data df into a datetime object. It helped us with manipulation and analysis. Once the data is converted the year is retained using .dt.year. This analysis is based on the year. In this, the data is grouped by the DATE\_OCCURED variable which has only a year. In the funds data the missing values in the variable DATE\_OCCURED are filled using the fill method. Here, the funding data and crime data are merged based on the DATE\_OCCURED column. We have preprocessed and combined the two datasets for analysis of the relationship between funds and crime types on a yearly basis.

**Figure 5.13**

*Merged the Crime Data and Funding Data*





In the Total Funding, Federal Grants Subtotal, and Total Expenditures columns we have cleaned the dataset by removing the commas and then converted the cleaned values to floats. Initially, the data was in a string format with commas as thousands of separators. Hence, we converted to floats for calculations. From the above code, we can see that DATE\_OCCURED has no strong correlation with any of the variables. The Total Funding has a weak positive correlation with the DATE\_OCCURED and there is no strong correlation with crime\_count. The variable Total Expenditures shows a moderate positive correlation with the variable DATE\_OCCURED and a weak positive correlation with crime\_count. Federal grants subtotal doesn't show any strong correlation with DATE\_OCCURED and a weak negative correlation with crime\_count. crime count has a strong positive correlation with DATE\_OCCURED. Looking at the correlation matrix we can say that crime counts increase with the years.

**Figure 5.14**

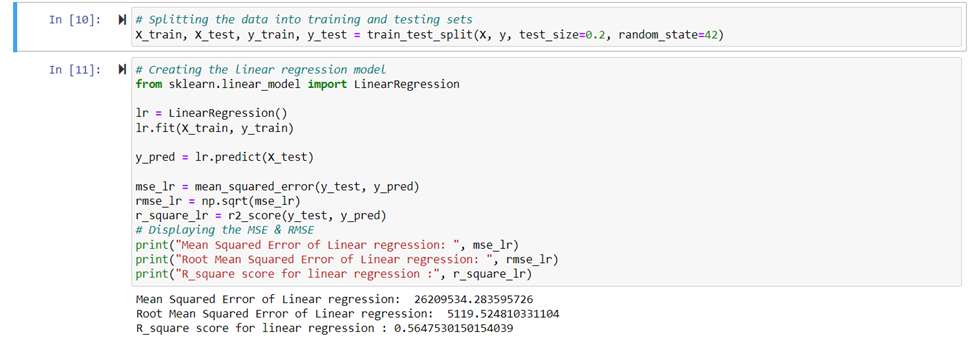
*Correlation between Crime Data and Funding Data*



**Linear Regression**

**Figure 5.15**

*Output for Linear Regression*

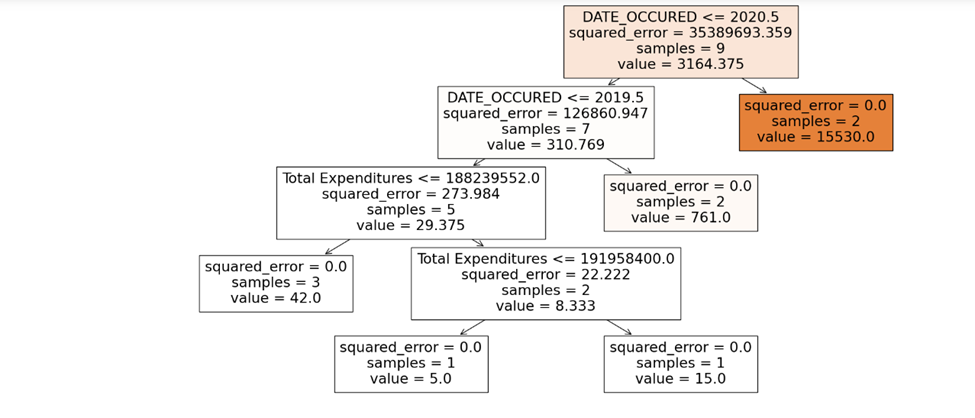


This model's MSE is quite high, indicating that the model's predictions differ significantly from the actual values. An RMSE of 5119 indicates that, on average, the model's predictions differ by 5119 points from the actual crime rate. This appears to be a significant inaccuracy. An RMSE of 5119 would be appropriate if crime\_count usually falls into the tens or hundreds of thousands range. This error may be very large if crime\_count is normally smaller. Hence, it is a less accurate model. At around 0.565, the R2 value indicates that the model explains 56.48% of the variation in the crime data. The R2 score of 0.565 indicates that the model has an appropriate predictive capability, but the high MSE and RMSE values show that there is a significant degree of inaccuracy in the predictions. Here, the model's accuracy is not good, so we have evaluated further using another model (Random Forest Model).

**Random Forest Classifier Algorithm**

**Figure 5.16**

*Comparing Total Expenditures with Date Occurred*



The total expenditures are the first decision node of the tree. The dataset is split based on whether the total expenditures are less than or equal to 188,239,552. The average value for these samples is 29.375.

The above tree has a good number of nodes, it suggests a model that is not too complex. Here, the squared error of the few leaf nodes is 0, this indicates that it perfectly fits in the training set. The higher squared error at the root node indicates the predictions in the branch may be less accurate.

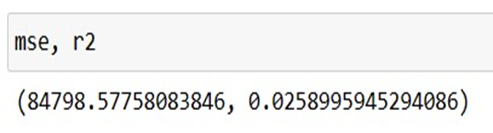
The research question "Are there lower crime rates in areas receiving more funding?" makes use of two different datasets: one that provides funding details, while the other one focuses on crime data. The intention is to determine if places with higher funding levels also exhibit lower crime rates. First, a linear regression was employed in order to evaluate the relationship between funding and crime rate.

**Linear Regression**

Both datasets are merged on the same column ‘YEAR’ and 'Crime\_Count' is a new column that was created by adding up all the crimes in each category inside the crime data dataset by year and ZIP code. Linear regression was performed on Total Funding (Independent variable) and Crime\_Count (dependent variable).

**Figure 5.17**

*Results of Linear Regression*

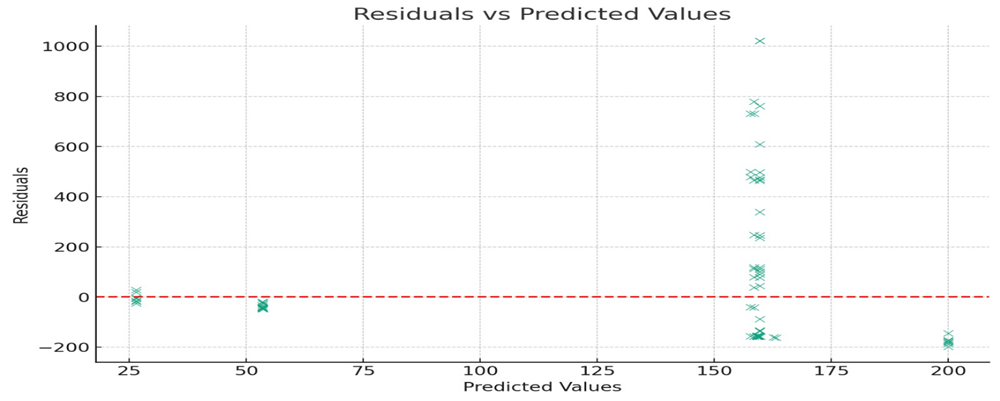


Mean squared Error (MSE) is 84798.577 and R-square value is 0.0258. The MSE value is very high, which suggests that the model is not a better fit for the data. 0.025 value of R- Square explains that only 2.5% of the variability in the crime counts, which is very low.

Plotted a Residuals Plot to show the predicted values on the x-axis and residuals on the y-axis.

**Figure 5.18**

*Residual Plot*



Around the zero line, the residuals are not randomly distributed. Rather, they exhibit a pattern in which the residuals are higher for both low and high predicted values.

The pattern of this plot and the results of MSE and R-square suggests that the linear regression model may not be the best fit for the data.

**Logistic Regression**

The funding data was grouped by year, calculating the meaning of 'Total Funding' for each year.

**Data Merging**

The aggregated crime and funding data were merged on the 'YEAR' column, forming a combined dataset with crime counts and average funding per year. 'Crime\_Count' is a new column that was created by adding up all the crimes in each category inside the crime data dataset by year and ZIP code.

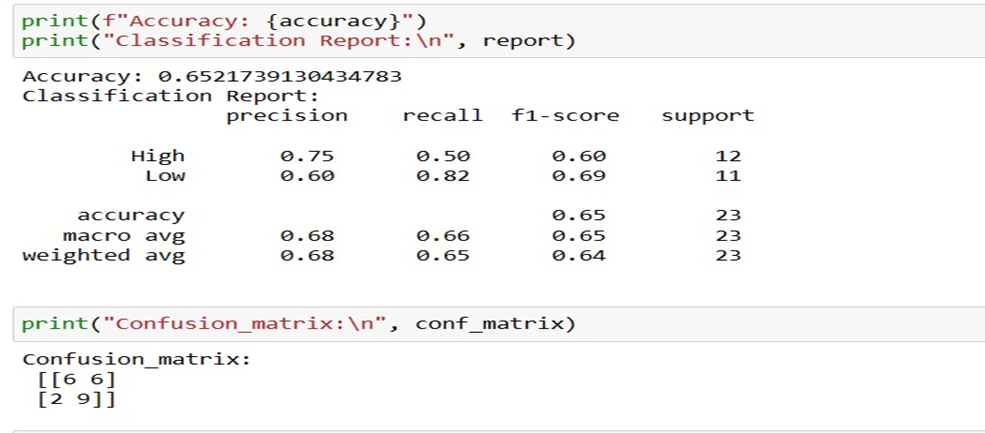
The 'Total Funding' column was cleaned by eliminating commas and converting their values to numeric. Missing values are filled using the median for the Total Funding column.

After categorizing the funding data by year, the meaning of "Total Funding" for each year was determined. A combined dataset containing average funding per year and crime counts was created by merging the aggregated crime and funding data on the 'YEAR' column. Based on the median crime count, a binary classification of crime rates ('High' or 'Low') has been established.

'Total Funding' (the independent variable) and 'CRIME\_TYPE' (the dependent variable) were compared using logistic regression.

**Figure 5.19**

*Logistic Regression Output*



**Logistic Regression Model Output**

The logistic regression achieved an accuracy of 65.2%. This implies that there is an acceptable level of dependability in the model's ability to predict the crime rate category from funding levels.

The F1-score, recall, and precision for 'High' crime rate prediction were 0.75, 0.50, and 0.60, respectively. The precision, recall, and F1-score scores for the model were 0.60, 0.82, and 0.69 for "Low" crime rates, respectively. 'Low' crime areas are more accurately identified by the model than 'High' crime areas, according to these measures.

**Confusion Matrix**

* True positive (accurate identification of high crime): 6
* A correctly diagnosed low crime rate is a true negative. 9
* False positive: 6 Low crime cases that were mistakenly classified as high.
* False negative: two high crimes that are mistakenly classified as low crimes.
* This emphasizes the model's ability to identify places with low crime rates even further.

The findings point to a modest correlation between funding levels and crime rates. It's a common belief that places receiving more funding will also have lower crime rates.

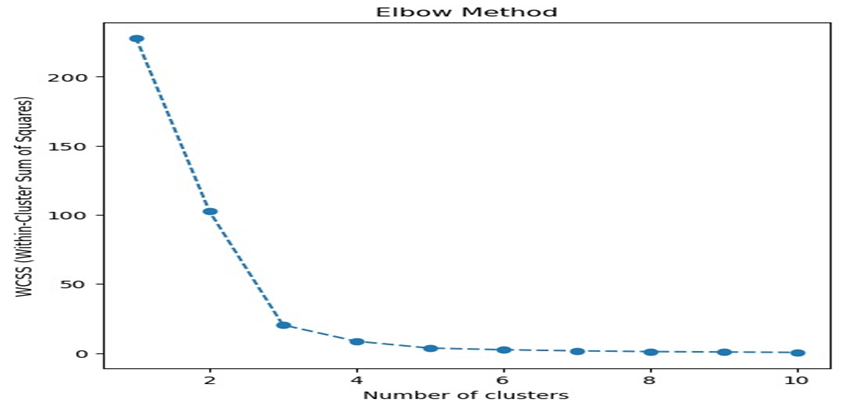
**Clustering Algorithm (K-Means)**

To examine the connection between funding and crime rates, we first combined the funding and crime data according to appropriate variables (such as year or ZIP code), and then united the two sets of data. The total number of crimes (Crime Count) for each year (YEAR) per ZIP code (ZIP\_CODE) and the related amount of funding (Total funding) for that year are now included in the combined dataset.

Next, to find groups of places with related funding and crime rates, we can apply clustering techniques utilizing algorithms such as K-means. This will make it easier to determine whether there are different areas with similar financing and crime patterns.

**Figure 5.20**

*Elbow Method Graph*



The best number of clusters for K-means clustering can be found using the Elbow Method visualization. The 'elbow' point, or the point at which the Within-Cluster Sum of Squares (WCSS) declines more slowly, is what we're looking for in the above figure. Because it does not result in much-improved data modeling when more clusters are added, this point is seen to be a good estimate for the number of clusters.

The 'elbow' point appears to occur around two or three clusters based on the figure. This shows that finding separate patterns in crime rates related to funding may be best achieved by clustering the data into two or three clusters. Clustered the data into 3 clusters. Performed K-means clustering with 3 clusters, which are distinguished by the average number of crimes and the average amount of Total funding.

**Output for the K-Means Clustering**

**Figure 5.21**

*K-means Clustering Output*



**Figure 5.22**

*Cluster of Areas*

A graph with numbers and colored dots

Description automatically generated with medium confidence

Cluster 0 (blue): The average overall financing is roughly 2.15 billion, which is very high, while the average crime rate is only about 21. This graph's top left position denotes a high level of financing but a low rate of crime. This shows that, maybe because of increased spending, the crime rates in this cluster may be reduced.

Cluster 1 (orange): on average, there are 58 crimes committed there, and the total funding is roughly 758.11 million. Indicating moderate funding levels and moderate crime rates, this cluster is located near the middle of the graph.

Cluster 2 (green): The average overall funding is almost the same as Cluster 1 (764.74 million), but the average crime rate is much higher (709). Despite having comparable funding levels to Cluster 1, this cluster has greater crime rates.

The central point of each cluster, represented by the red point, is the average of the clustered data points (total financing and the number of crimes).

The lowest crime rates are seen in Cluster 0, which has the most funding, indicating that a higher funding level may be linked to a lower crime rate. Nonetheless, Clusters 1 and 2 had notably different crime rates despite receiving comparable financing, suggesting that variables other than funding may have a big impact on crime rates.

**CHAPTER 6 – MODEL EVALUATION**

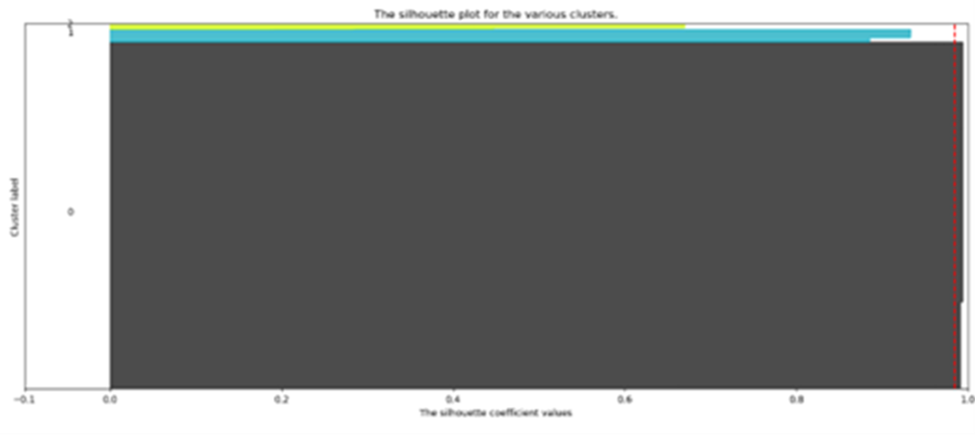
For research question 1 we performed three modelling techniques, Linear Regression, Random Forest, and K-Means Clustering. Below is the evaluation of each model:

Even though we have low MSE value which shows good model performance, the exceptionally low R-square value concludes that the model does not fit well. For Random Forest we can see that F1-scores indicate a superior performance but if we look at precision and then on F1-scores we can say that the model is not performing well. But when we compare both the models i.e., the Linear Regression and Random Forest Classifier we can see that the Regression performs a bit better with low MSE score but none of the models were giving us accurate results. Lastly, we did cluster we checked for Silhouette score.

**Figure 6.1**

*Silhouette Score*

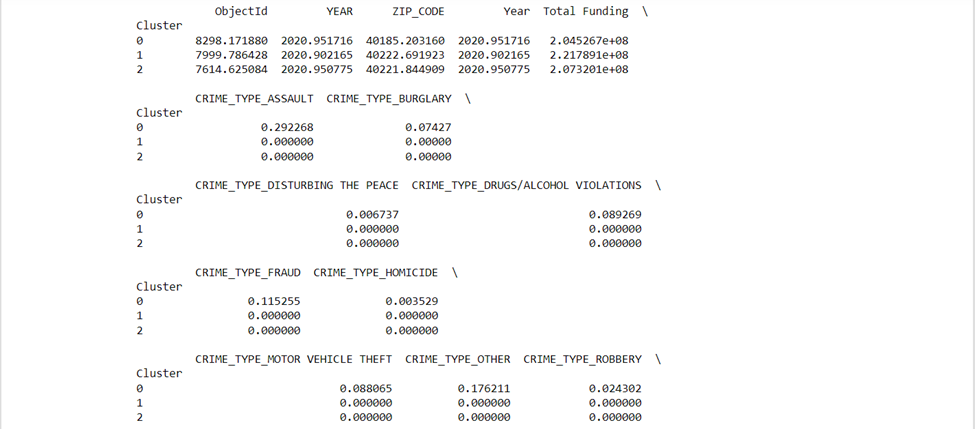


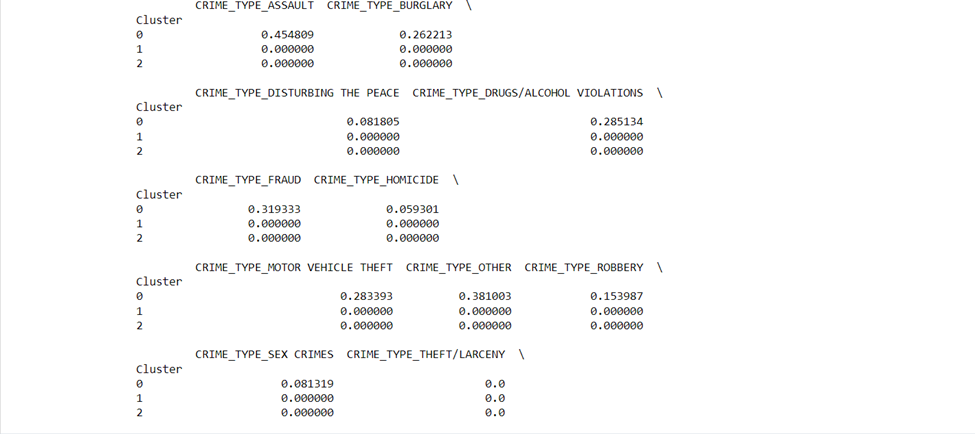
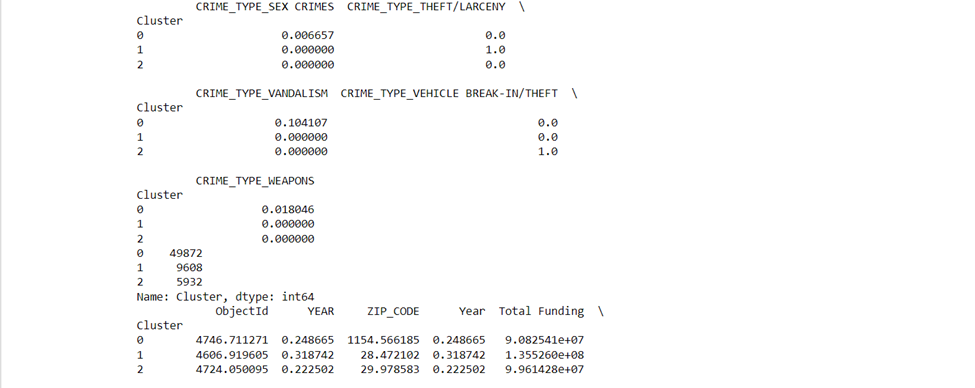


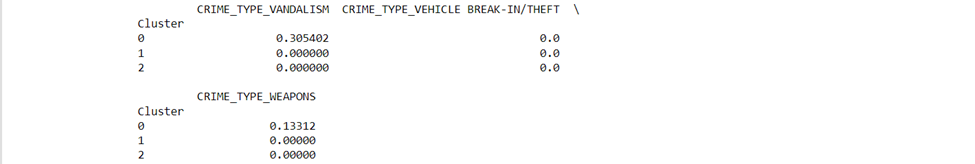
Even though we got a very good silhouette score of 0.98 but it doesn’t predict that there will be a link between public funding and Crime types, the good score just shows that the attributes are divided very well in clusters. So, to see whether the model fits with our research question we performed cluster categorization standard deviation to see if the Funding is having any effect on Crime Types or not. Below FIG shows the results of cluster categorization standard deviation.

**Figure 6.2**

*Cluster Categorization Standard Deviation*







We have the results for crime types and total funding for year 2020,as 2020 was our target year we considered seeing the difference in the crime rate for that year, the output we have had the average values of different crime types and Total Funding for each cluster, we can we observe that the crime types of rates differ significantly from before funding and after funding like if we look at ‘ASSAULTS’ ,’DISTURBING PEACE’ ,’HOMICIDES’, SEX CRIMES’, ‘OTHER 'crime types rates were lower before funding but after funding there was an uptick in them, but all the rest of the crime types had were having low rates after funding so we can conclude that there is a relationship between funds and crime rates.

A series of Chi-Squared tests have elucidated the associations between crime types and temporal factors such as season, day of the week, and holiday status. The statistical analyses revealed that the distribution of crime types is significantly influenced by seasonal changes, with a Chi-Squared value of approximately 647.753 and a p-value near 9×10^−108, indicating a notable variation in crime occurrences across seasons. Similarly, the day of the week was found to be a significant factor, evidenced by a Chi-Squared value of 1057.809 and a p-value of 5.51×10^−165, suggesting certain crimes are more likely on specific days. Additionally, holiday status also appeared to affect crime types, with a Chi-Squared value of about 969.389 and a p-value of 4.89×10^−197, highlighting a differential pattern of crime on holidays. Collectively, these findings strongly refute the null hypothesis of a uniform distribution of crime types across these temporal dimensions, suggesting that strategic law enforcement and preventative measures could be synchronized with these patterns to effectively address and mitigate crime rates.

**Figure 6.3**

*Linear Regression Output*



The MSE decreased from the MSE of the linear regression model, showing that the Random Forest model has, on average, lower squared differences between predicted and actual values. Compared to the linear regression model the RMSE is lower for this model. It indicates that this is a better fit for the data as the average error is smaller. The R2 value has increased to 0.77. It shows that there is an improvement from the linear regression model.

Upon comparison of both models, we can say that the Random Forest model has better accuracy than the linear regression model with the lower MSE, RMSE, and a higher R2 score. Random Forest model is a better fit for this dataset.

**Figure 6.4**

*Comparison of Linear Regression Model and Random Classifier Algorithm*



From the above results, we can see the average squared difference between the estimated and the actual values. The mean squared error is 13774851 for Random Forest model and 26209534 for Linear regression. The linear regression model has a higher MSE value than the Random Forest model. We can conclude that the Random Forest model's predictions are closer to the actual values.

The Root mean squared error is 3711.44 for Random Forest model and 5119.52 for Linear regression. Here, the lower the values are better. Random forest outperforms the linear regression model.

The R2 value for the Random Forest model is 0.77 and for the linear regression model is 0.56. The higher the R2 indicates the better fit. The Random Forest model has a higher R2 value. It explains the variance in the data is better than the linear regression model.

The linear regression model is not good enough to explain the variation in crime rates according to funding levels. A moderate degree of accuracy is offered by logistic regression, which is especially useful for locating places with low crime rates.

Although the clustering approach shows that funding is not the only factor influencing crime rates, it does provide insightful information about potential correlations between various funding levels and crime rates.

Although there is a correlation between funding levels and crime rates, the analysis demonstrates that it is not clear-cut and is probably impacted by a number of other socioeconomic factors. As demonstrated in Cluster 0, the models suggest that more funding may occasionally be linked to decreased crime rates; however, this is not a common pattern. The effective use of funds, socioeconomic circumstances, policing tactics, and community initiatives are among the other important elements that might influence crime rates.

**CHAPTER 7 – CONCLUSION**

**How does the allocation of various types of public safety funding relate to distinct categories of crime rates?**

To evaluate the effect of public safety funding, the analysis has looked at every type of crime, including assault, sex crimes, homicides, etc. Grants for police agencies, community policing projects, youth initiatives, and other interventions targeted at decreasing crime are examples of public safety funding. The methodology as involved comparing the rates of crime before and after the funding was granted. The conclusion is that increased financing has had a beneficial impact on lowering crime rates for all the crime type expects for few. Therefore, we can say that there is a relationship between crime rates and funds.

As we had funding data not helping to reduce the rates for ASSAULTS’,’DISTURBING PEACE’,’HOMICIDES’, SEX CRIMES’, ‘OTHER by providing funds to the victims. Instead of reducing these crime rates, we will need to implement the policies and for future work, we look forward to seeing how the policy implementation helps reduce them.

**How do temporal patterns influence the prevalence of crime, with a focus on the variation across seasons, the incidence during holidays, and the fluctuation throughout the days of the week?**

The research is focused to understand how temporal patterns influence the prevalence of crime, particularly focusing on variations across seasons, incidences during holidays, and fluctuations throughout the days of the week. The findings reveal that the distribution of crime is not uniform across these temporal variables. Statistical analysis, particularly Chi-squared tests, indicated significant p-values, confirming that the observed variations in crime occurrences are statistically significant and not due to random chance. This information is crucial for law enforcement and public safety agencies, as it allows for optimized planning and resource allocation. By aligning crime prevention strategies with periods of heightened criminal activity, these agencies can enhance their effectiveness.

Visual analyses, including count plots, have shown that certain crimes, like 'THEFT/LARCENY' and 'ASSAULT', are more frequent on specific days and during particular seasons, notably in the summer. The prevalence of 'DRUGS/ALCOHOL VIOLATIONS' on weekends and 'FRAUD' midweek indicates behavioral patterns that can be used for targeted interventions. The slight increase in average daily crime counts during holidays, while statistically insignificant, suggests a minimal need for deviation in law enforcement measures these days.

Principal Component Analysis (PCA) has been beneficial in simplifying the multidimensional crime data, revealing potential clusters and outliers indicative of unique seasonal patterns. While PCA provides a broad overview and suggests a structure in how assaults correlate with seasonal trends, more detailed analysis is needed to establish precise correlations and to understand the role of outliers.

The combination of statistical tests and visual data exploration techniques like PCA offers a comprehensive understanding of crime trends over time. This analysis highlights the complex nature of criminal activity and its interaction with temporal elements, emphasizing the need for adaptive strategies in law enforcement and public policy.

Outliers in the Principal Component Analysis (PCA) indicate unique crime patterns, warranting further investigation. Additionally, integrating socio-economic, demographic, and environmental data could offer a more comprehensive understanding of the factors influencing crime trends. Advanced analytical techniques, such as sophisticated statistical methods and machine learning models, should be employed to unravel the complex interplay between temporal factors and crime rates. It is crucial to translate these findings into actionable policy recommendations for law enforcement and public safety agencies. This involves operationalizing insights from temporal patterns to enhance crime prevention strategies and community safety measures, bridging the gap between academic research and practical application.

The analysis reveals a notable temporal lag between fund allocation and its impact on crime rates, suggesting the need for timely and strategic funding to effectively mitigate crime. Further research should explore the nuances of fund allocation, particularly its timing, magnitude, and targeted areas. Expanding the range of temporal variables studied, including long-term trends and specific daily time frames, could provide deeper insights.

**Is there a time lag between the allocation of funds and its effect on crime rates?**

We can clearly see there is a temporal lag between the allocation of funds and the effects of crime rates. The analysis shows us that there is a greatest impact on crime rates when there is a temporal lag in allocating funds.

The findings highlight how crucial it is to take temporal gaps into consideration when formulating policies aimed at reducing crime. Funding's immediate consequences might not be as significant as those seen over a longer time period. This analysis shows the usefulness of advanced predictive models like Random Forest in comprehending and predicting these intricate linkages, in addition to highlighting the delayed but substantial influence of funding on reducing crime rates.

**Are there lower crime rates in areas receiving more funding?**

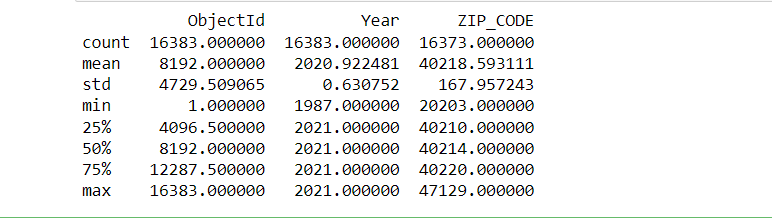
Finally, it should be noted that there is neither a clear-cut correlation nor a uniform link between Funding and crime rates. Funding is a significant but not the only element determining crime rates, as evidenced by the difference in model performances and cluster patterns. The uneven effects of funding on crime rates among various clusters emphasizes the significance of considering additional contextual elements, including but not limited to socioeconomic conditions, the effectiveness of law enforcement, and community involvement measures. These findings indicate that understanding and efficiently managing crime rates requires a more holistic strategy that considers a variety of factors in addition to funds.

Future studies could benefit from gathering more precise data that can be directly associated with crime figures, including a larger range of socioeconomic characteristics, and using sophisticated statistical methods to prove the relationship. Furthermore, a more complex model could more accurately reflect the complex link that exists between funding and crime rates, allowing for more efficient and strategic resource allocation.

**APPENDIX**

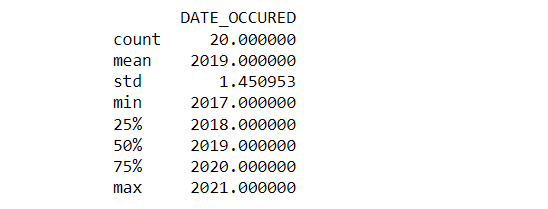
**Appendix A:**

Descriptive Statistics of Quantitative Variables for Crime Data



**Appendix B:**

Descriptive Statistics of Quantitative Variables for Funding Data



**REFERENCES**

Borjas, J. F. (May 2019). *Does a Police Department’s Budget Affect the Crime Index in Its Jurisdiction?* Retrieved from <https://www.google.com/books/edition/Does_a_Police_Department_s_Budget_Affect/6alDzgEACAAJ?hl=en>

CRS Reports. (2022, December 12). Violent Crime Trends, 1990-2021. Retrieved from <https://crsreports.congress.gov/product/pdf/IF/IF12281>

Jaffe, S. (2020). Decisions to be made on US gun violence research funds. *ScienceDirect*. Retrieved from [https://www.sciencedirect.com](https://www.sciencedirect.com/)

Hernandez, A. (2022, October 16). Despite 'defunding' claims, police funding has increased in many US cities. *ABC News*. Retrieved from <https://abcnews.go.com/US/defunding-claims-police-funding-increased-us-cities/story?id=91511971>

Kamide, S. (2021, December 27). It’s time for holiday cheer: Does it increase crime? *Pumphrey Law*. Retrieved from <https://www.linkedin.com/pulse/its-time-holiday-cheer-increase-crime-don-pumphrey/>

Lehrer, E. (n.d.). Crime and Economy: What Connection? *The Heritage Foundation*. Retrieved from <https://www.heritage.org/crime-and-justice/commentary/crime-and-economy-what-connection>

Lin, M.-J. (2009). More police, less crime: Evidence from US state data. *International Review of Law and Economics, 29*(2), 73-80. Retrieved from <https://www.sciencedirect.com/science/article/abs/pii/S0144818808000732>

Lopez, G. (2016, July 19). 6 proven policies for reducing crime and violence without gun control. *Vox*. Retrieved from <https://www.vox.com/2016/2/15/10981274/crime-violence-policies-guns>

Mello, S. (2019). More COPS, less crime. *Journal of Public Economics, 172*, 174-200. <https://doi.org/10.1016/j.jpubeco.2018.12.003>

Reeping, P. M., & Hemenway, D. (2020). The association between weather and the number of daily shootings in Chicago. *Injepi Journal, 7*(31). Retrieved from <https://injepijournal.biomedcentral.com/articles/10.1186/s40621-020-00260-3>

The Deputy Attorney General. (n.d.). *Comprehensive Strategy for Reducing Violent Crime*. Retrieved from <https://www.justice.gov/d9/pages/attachments/2021/05/26/comprehensive_strategy_for_reducing_violent_crime_memo.pdf>

Vaughan, A. D., & Andresen, M. A. (2018). The cost of mental health-related calls on police service: evidence from British Columbia. In *Evidence Based Policing: An Introduction* (pp. 173). Retrieved from <https://www.crimrxiv.com/pub/s0j8fqx1/release/1>

Van Sleeuwen, S. E. M., Steenbeek, W., & Ruiter, S. (2021). When Do Offenders Commit Crime? An Analysis of Temporal Consistency in Individual Offending Patterns. *Journal of Quantitative Criminology, 37*(4), 863–889. <https://doi.org/10.1007/s10940-020-09470-w>

Weisburst, K.E. (2018). Safety in Police Numbers: Evidence of Police Effectiveness from Federal COPS Grant Applications. *American Law and Economics Review, 21*(1), 81–109. <https://doi.org/10.1093/aler/ahy010>

Worrall, J. L. (2004). Funding Collaborative Juvenile Crime Prevention Programs: Does It Make a Difference? *Evaluation Review, 28*(4). Retrieved from <https://journals.sagepub.com/doi/abs/10.1177/0193841X04267908>