Analysing New York City Police Department (NYPD) Arrest Data For Crime Insights And Policing Strategies

Sravani Pulusu
Data Analytics Engineering
George Mason University
Fairfax, VA
spulusu@gmu.edu

Sai Manish Reddy Pannala
Data Analytics Engineering
George Mason University
Fairfax, VA
spannala@gmu.edu

Priyadarsini Devarapalli
Data Analytics Engineering
George Mason University
Fairfax, VA
pdevara@gmu.edu

***Abstract*— In this project, our team aims to analyze the New York City Police Department (NYPD) Arrest Data to derive insights into crime patterns, law enforcement practices, and community impact. Focused on evidence-based strategies, we employ data science and analytics to scrutinize the extensive dataset, exploring crime trends, hotspots, and demographic disparities. By leveraging machine learning techniques, we intend to predict crime patterns, aligning with the growing trend in data-driven decision-making within law enforcement. Our motivation encompasses enhancing public safety, evaluating community policing initiatives, promoting equity, and optimizing resource allocation. The project involves data cleaning, exploratory data analysis, clustering, and geospatial visualization, ultimately providing valuable recommendations for improving policing strategies, fostering community relations, and ensuring public safety in the dynamic context of New York City.**

Keywords— NYPD Arrest Data, Crime Patterns, Predictive Policing, Community Impact

I. Introduction

New York City, one of the world's busiest cities, is home to a diverse range of neighbourhoods, cultures, and dynamics. As such, policing this city requires evidence-based strategies that are both effective and equitable. The NYPD Arrest Data (Year to Date) dataset provides a wealth of information, including the type of crime, location, time of enforcement, and suspect demographics. The New York City Police Department (NYPD) Arrest Data (Year to Date) dataset, available on data.gov, is a comprehensive repository of information regarding arrests made within the city. This dataset offers a unique opportunity to investigate the particulars of law enforcement, crime, and community dynamics in New York City's diverse and dynamic geography. The NYPD's goal is to safeguard public safety, uphold the law, and build positive relationships with the community, thus it is vital to scrutinize and interpret the massive amounts of data generated by these operations. This project aims to analyze the NYPD Arrest Data using data science and analytics to answer crucial questions regarding crime patterns, law enforcement practices, and community impact. We recognize the significance of this endeavor considering its potential to not only inform policing policies, but also to improve transparency, accountability, and public trust in law enforcement institutions. [1] Through an in-depth analysis of the NYPD arrest data, this project seeks to shed light on the current state of crime and the effectiveness of law enforcement strategies. By leveraging this data, we aim to uncover patterns and trends that can inform policing strategies and contribute to public safety.[2]

In the era of big data, law enforcement agencies around the world are increasingly turning to data-driven approaches to enhance their effectiveness. The NYPD is no exception. By analyzing the NYPD Arrest Data, we can gain insights into crime patterns and trends, and evaluate the effectiveness of current policing strategies. This data-driven approach allows for more informed decision-making, helping the NYPD to allocate resources more efficiently, respond more quickly to emerging crime trends, and ultimately, better serve and protect the communities of New York City [18][26].

One of the most promising applications of data science in law enforcement is predictive analytics. NYPD is already using machine learning to spot crime patterns.[20] By analyzing historical crime data, predictive models can help identify potential crime hotspots, anticipate future crime trends, and even predict the likelihood of an individual reoffending. In this project, we will explore the potential of predictive analytics by applying machine learning algorithms to the NYPD Arrest Data. Our goal is to develop a model that can accurately predict crime occurrences, providing a valuable tool for proactive policing.[21]

The findings from this project have important implications for policy and practice. By identifying factors associated with crime and arrest trends, we can inform the development of evidence-based policies and interventions. Moreover, by highlighting disparities in arrest rates across different demographic groups, we can draw attention to potential issues of equity and fairness in law enforcement. Ultimately, the goal of this project is not just to analyze the data, but to use the insights gained to make a positive impact on policing practices and community safety in New York City.

II. Literature Review

Several literature searches were conducted for the purpose of this study. In this literature review, we can explore relevant research, including studies on predictive policing, law enforcement strategies, and geospatial analysis, to provide valuable context and insights for our analysis.

The research "Minority Report" a Reality? The NYPD's Big Data Approach to Predicting Crime by Clemens discusses how the NYPD has developed a data-driven approach to fight, and even predict, crime. Clemens' study investigates the utilization of big data and predictive policing by the New York City Police Department (NYPD) as a means to address and potentially predict criminal activity. The study highlights the incorporation of vast amounts of data, including crime locations, times, and types, into sophisticated computer models and algorithms. These models enable the prediction of areas with a higher likelihood of criminal activity. The study also emphasizes the use of "Big Data"-driven predictive policing by analyzing regularly recorded crime data (location, time, and crime), using sophisticated computer models and algorithms, to predict places of expected criminal activity. This approach is consistent with the core principles of our project since it prioritizes the implementation of data-driven strategies. This study offers valuable insights on the utilization of crime data for the purpose of predicting areas with high criminal activity.[3]

The research article "The Effects of Local Police Surges on Crime and Arrests in New York City" published in PLOS ONE tested the effects of Operation Impact on reported crimes and arrests from 2004 to 2012 using a difference-in-differences approach. According to the findings of the study, Operation Impact was significantly associated with reductions in total reported crimes, assaults, burglaries, drug violations, misdemeanour offences, felony property crimes, robberies, and felony violent crimes. This study holds significant relevance with our project as it investigates the consequences of particular police techniques, providing valuable insights into the potential outcomes of data-driven policing initiatives. [4]

The study by Chainey et al. highlighted hotspot mapping as a useful method for predicting spatial crime patterns. These insights are critical when we analyse the NYPD Arrest Data, emphasizing data-driven ways to improving New York City law enforcement strategies. Chainey stresses the importance of geographical analysis, particularly the value of hotspot mapping, as a complement to the temporal dimension. Their study demonstrates how spatial analysis methods can reveal spatial trends in crime, an important component of crime analysis. Geospatial analysis reveals crime hotspots—specific geographic areas with higher concentrations of criminal activity—when applied to datasets like the NYPD Arrest Data. This knowledge enables law enforcement organizations to concentrate their efforts on specific regions, using manpower, resources, and preventive measures to effectively target and deter criminal activity. It also helps to comprehend crime displacement by demonstrating whether criminal activity moves from one site to another because of law enforcement efforts, allowing for the development of adaptive measures. These studies provide unique and significant insights into the study of NYPD arrest data, which may be relevant to our project. They offer insights on predictive policing, the influence of various law enforcement techniques, and the significance of geospatial data. We can obtain a better knowledge of the background for our analysis, identify areas of interest, and evaluate the findings of our investigation by reviewing these papers. They collectively emphasize the role of data-driven methods in improving law enforcement strategies and enhancing public safety. [5]

The "Utilization of Power 2020" NYPD report analyzes police use of force in 2020, categorizing incidents by factors like suspect resistance. It emphasizes transparency, presenting detailed statistics and insights, and discusses training protocols for responsible force use.[6]

The study "Policing the pandemic: estimating spatial and racialized inequities in New York City police enforcement of COVID-19 mandates" conducted a retrospective spatial analysis of demographic factors and public health policing in New York City from 12 March–24 May 2020. The findings demonstrated pronounced spatial and racialized inequities in pandemic policing of public health that mimic historical policing practices deemed racially discriminatory. The study found that ZIP codes with higher percentages of lower income and Black residents experienced disproportionately high rates of policing during the COVID-19 pandemic [7].The "Investigation of the CompStat Cycle" report on the Department of Justice Assistance (BJA) website evaluates the CompStat model's impact on U.S. police departments. Produced by PERF, it examines crime reduction, resource allocation, and accountability. Valuable for policing organizations, it contributes to discussions on modern policing. In a 2021 Taylor and Francis paper, "Utilization of Power in Metropolitan Police Work," the study compares American and Dutch police in urban settings. Through case studies, it explores contextual factors, tactics, and training methods influencing force use. Valuable for policymakers, it bridges theory and practice, fostering discussions on cross-cultural perspectives and enhancing policing strategies. [8]

The article "Analysis of NYC Reported Crime Data Using Pandas" on Towards Data Science is a practical guide to crime data exploration in New York City. Using Python's Pandas library, the article simplifies complex concepts, making it useful for both data science enthusiasts and those interested in data-driven insights into crime dynamics. The article adds to the literature on leveraging data science for public safety.[9]

The New York Times' intelligent component, "The amount Change Do the Police Need?" distributed in Walk 2021, is a provocative and outwardly convincing assessment of policing rehearses in New York City. Through a blend of insightful reporting and intuitive information perceptions, the article fundamentally examines the endeavors and effects of the New York City Police Division (NYPD) in embracing changes. The vivid show permits peruse to investigate key measurements and accounts connected with policy change, for example, the utilization of power, official discipline, and local area relations. The intuitive idea of the element works with a nuanced comprehension of the intricacies encompassing police change drives. The incorporation of individual stories and genuine models carries a human aspect to the examination, making it open to a vast crowd. This piece from The New York Times fills in as an essential asset for people looking for an extensive and outwardly captivating investigation of the continuous difficulties and headways in police change inside the setting of one of the country's most significant police divisions.[10]

The New York Magazine's article named "The Wrongdoing Battling Project That Changed New York Everlastingly," distributed in Walk 2018, offers an intelligent review of the advancement and effect of the CompStat program in forming the scene of policing New York City. Wrote by specialists in the field, the article gives a nitty gritty assessment of how CompStat, at first presented in 1994, reformed the way to deal with wrongdoing decrease through information-driven methodologies. The piece digs into the program's parts, remembering its concentration on ideal and precise data, quick organization of assets, compel policing strategies, and persevering development stressing how these components added to a massive decrease in crime percentages throughout the long term. By winding together, a verifiable setting, interviews, and measurable proof, the article portrays the program's victories, recognizing its extraordinary job in making New York City more secure. This review examination is fundamental for anyone with any interest in understanding the significant crossroads throughout the entire existence of present-day policing and the enduring effect of creative methodologies like CompStat on metropolitan security.[11]

The proposal for an NYPD Inspector General by the Brennan Center for Justice emphasizes the need for increased oversight of the NYPD’s intelligence operations, which have expanded significantly in their efforts to keep New York safe from terrorism. The Brennan Center suggests that oversight mechanisms have not kept pace with the police’s new and expanded roles and recommends that an independent inspector general be established for the NYPD. This proposal is particularly relevant to our project on NYPD arrest data, as it underscores the importance of transparency, accountability, and independent oversight in law enforcement operations.[15]

The report named "New York City Police Division's (NYPD) Reaction to Exhibitions Following the Passing of George Floyd," delivered by the Workplace of Head legal officer (OAG) for the Province of New York in 2020, offers a thorough assessment of the NYPD's activities during the fights set off by the unfortunate demise of George Floyd. This definite report carefully investigates the different techniques and strategies utilized by policing the elevated time of common distress. By giving a granular record of explicit episodes, the report takes into consideration an exhaustive assessment of the NYPD's adherence to legitimate guidelines and the possible effect on people's privileges to serene gatherings. Focused on legitimate experts, policymakers, and the overall population, the report fills in as a significant asset, improving comprehension of how we might interpret the complex elements of policing protestors in the midst of cultural strife. Past simply recognizing areas of concern, the report presents helpful proposals for strategy changes and improved preparation, with the overall objective of resolving fundamental issues and directing positive changes in how public exhibitions are policed [12].

The document titled "Discipline in the NYPD 2016-2017" offers a thorough exploration into the intricate landscape of disciplinary procedures within the New York Police Department (NYPD) during the specified period. This comprehensive report functions as a pivotal resource in cultivating transparency and insight into the internal mechanisms employed by the NYPD to address instances of misconduct and ensure adherence to professional standards among its officers. Through a meticulous analysis of disciplinary actions, case studies, and overall trends, the report provides a multifaceted view of the disciplinary landscape, unravelling the diverse factors that contribute to corrective measures. By offering a detailed breakdown of the disciplinary process, from investigations to hearings, the report not only serves as an informational asset for internal stakeholders but also acts as a crucial tool for external oversight entities and the general public. In essence, this document contributes significantly to the ongoing dialogue surrounding accountability, ethics, and the continual enhancement of law enforcement practices within the NYPD [13] .

Each of these papers offers a distinctive viewpoint on the analysis of NYPD arrest data, and they may be able to provide us useful information for this project. These help us in understanding the larger context of our analysis, identifying prospective areas of interest, and evaluating the results of our investigation.

III. Description of the Problem

Policing a city as diverse and dynamic as New York needs constant adaptation. Crime is a significant problem in New York City, and understanding crime patterns and trends is crucial for developing successful crime prevention and reduction tactics[16]. The New York Police Department (NYPD) keeps a detailed record of all arrests made throughout the year [14].While this dataset is rich in information, it is also complex and large, making it difficult to extract useful insights manually. Although the NYPD Arrest Data (Year to Date) dataset contains a large amount of arrest information, it is not used to guide crime prevention measures, analyze community policing activities, or analyze the dynamics of crime in New York City.**Error! Reference source not found.**. The aim of this project is to analyze and predict crime patterns using machine learning techniques based on this dataset. In the broader context, this project aligns with the growing trend in law enforcement towards data-driven decision-making. It aims to bridge the gap between the massive amounts of arrest data generated by the NYPD and the practical use of this data to enhance public safety and the efficiency of law enforcement operations. By identifying crime trends, hotspots, and crime types, governments and police departments can allocate resources more effectively and deploy their resources more strategically. By addressing these challenges and objectives, this project aims to provide valuable insights and recommendations for improving policing strategies, enhancing public safety, and building stronger community relations in the dynamic and diverse environment of New York City.

IV. Importance of the Problem

Crime is a major problem in New York City, and it is important to develop effective crime prevention and reduction strategies. Effective crime analysis can significantly enhance public safety and resource allocation in law enforcement. Understanding and analyzing the NYPD Arrest Data has a direct impact on public safety because it serves as the foundation for building successful law enforcement techniques that aid in the maintenance of security in a metropolis as dynamic and diverse as New York. Law enforcement organizations can strategically deploy resources for maximum impact by identifying potential crime hotspots and periods of increased criminal activity. Furthermore, understanding the factors influencing crime can inform policy decisions and community engagement efforts. By examining the impact of community policing activities and resolving demographic disparities in arrests, this study can also strengthen community bonds and promote fairness and equity within the criminal justice system. Furthermore, using this dataset to feed predictive policing models, proactive law enforcement can be enabled, optimizing resource allocation for more efficient crime prevention and response. [17]

V. Motivation

The motivation behind selecting the NYPD Arrest Data (Year to Date) dataset for this project is driven by several key factors:

- **Enhancing Public Safety:** In cities like New York City, crime is a major concern. By analyzing this dataset, law enforcement might better understand crime trends and take proactive measures to deter and prevent criminal activity, which will eventually improve public safety.
- **Community Policing Enhancement:** The dataset enables an evaluation of community policing initiatives. Understanding the influence of community engagement on crime rates is crucial for establishing confidence between law enforcement and the community.
- **Equity and Fairness:** The dataset allows us to look into demographic differences in arrests. It is essential to identify and eliminate potential biases in law enforcement practices in order to ensure equity and fairness in the criminal justice system.
- **Resource Allocation Optimization:** This data can be used to create predictive policing models. These models can help with the optimal allocation of law enforcement resources, resulting in more effective crime prevention and response techniques.
- **Historical Crime Trends:** Comparing current arrest trends to historical data offers insight into how crime dynamics have changed over time. This historical framework is essential for understanding and adapting to changing patterns of crime.
- **Data-Driven Decision-Making:** The idea is in line with a broader trend in law enforcement to use data-driven techniques. The insights gained from this analysis can contribute to more informed decision-making and evidence-based policing strategies.

In summary, the NYPD Arrest Data (Year to Date) dataset was chosen to improve public safety, improve community relations, promote fairness within law enforcement, optimize resource allocation, and provide a data-driven approach to addressing crime dynamics in New York City.

VI. Research Questions

1) Are certain crimes more common during specific seasons or times of the day?

2) Which area has more sex related crimes?

3) Are there difference in arrest rates for different racialor ethnic groups?

4) Which age group has committed more crimes?

5) Where are the areas with the most arrests for different crimes in NYC?

VII. Data Source

The primary data source for this project is the “NYPD Arrest Data (Year to Date)” dataset, which is publicly available on the NYC Open Data portal and on the DATA.gov website. This dataset provides comprehensive information about every arrest made by the NYPD during the current year, including details about the type of crime, the location and time of enforcement, and suspect demographics. The data is manually extracted every quarter and reviewed by the Office of Management Analysis and Planning. This dataset will serve as the foundation for our analysis and insight generation.[1]:

VIII. Description of the Dataset

The given dataset is made up of arrest records, each of which is uniquely recognized by an ARREST_KEY. It also includes a variety of attributes pertaining to each arrest. The date of the

arrest is indicated by the ARREST_DATE, while the offense description and penal code are specified by the PD_DESC and PD_CD, respectively. KY_CD and OFNS_DESC provide additional classification and explanation of the transgression. Legal codes and the associated legal category of the offense (such as felony) are provided by LAW_CODE and LAW_CAT_CD. ARREST_PRECINCT (precinct) and ARREST_BORO (borough) are two specifics regarding the arrest location. The jurisdiction that is involved in the arrest is indicated by JURISDICTION_CODE. The offender's PERP_SEX, PERP_RACE, and AGE_GROUP demographic data are included. The terms Latitude, Longitude, X_COORD_CD, and Y_COORD_CD are used to denote geographic coordinates.
Furthermore, for every arrest record, a New Georeferenced Column displays the georeferenced point that combines latitude and longitude. This dataset is crucial for helping with crime analysis and law enforcement activities by illuminating trends and demographics related to arrests.

1. ARREST_KEY: A unique identifier for each arrest record.
2. ARREST_DATE: The date when the arrest occurred.
3. PD_CD: The penal code associated with the offense.
4. PD_DESC: Description of the offense corresponding to the penal code.
5. KY_CD: The internal classification code for the offense.
6. OFNS_DESC: Description of the offense category.
7. LAW_CODE: The legal code associated with the offense.
8. LAW_CAT_CD: The legal category of the offense (e.g., felony).
9. ARREST_BORO: The borough where the arrest occurred.
10. ARREST_PRECINCT: The precinct where the arrest occurred.
11. JURISDICTION_CODE: The jurisdiction code related to the arrest.
12. AGE_GROUP: The age group of the perpetrator.
13. PERP_SEX: The gender of the perpetrator.
14. PERP_RACE: The race of the perpetrator.
15. X_COORD_CD: The X-coordinate (geospatial coordinate) associated with the arrest location.
16. Y_COORD_CD: The Y-coordinate (geospatial coordinate) associated with the arrest location.
17. Latitude: Latitude of the arrest location
18. Longitude: Longitude of the arrest location.
19. New Georeferenced Column: A georeferenced point that shows where the arrest happened by combining longitude and latitude.

IX. Proposed approach

1. Data Preparation, Cleaning and Preprocessing: Data preparation and cleaning are essential to ensure the quality and integrity of the dataset. We will address missing values, outliers, and inconsistencies to create a reliable dataset for analysis.

2. Exploratory Data Analysis (EDA) and Statistical Analysis: To find patterns, trends, and linkages in the data, EDA and statistical analysis will be used. To acquire preliminary insights, we will use descriptive statistics and visualize the data.

3. Clustering and Identifying Influencing Factors: To discover groups or trends in the data, we will use clustering algorithms. This step seeks to identify connections or influencing elements that lead to crime trends.

4. Data Visualization: Extensive data visualization will be used to illustrate and make accessible the findings. Visualization will be critical in identifying high- and low-crime regions, as well as the SDOH characteristics connected with them. Visualizations will be used to analyze and pick the most and least observed locations for further investigation.

5. Communicating results: Finally, the data analysis insights and results are clearly and concisely presented.

We propose to use a combination of exploratory data analysis and statistical analysis techniques. Explorative data analysis helps in understanding the data and identifying potential trends and patterns. Following that, we will use statistical analysis to validate these patterns and gain deeper insights. We plan to experiment with various methods and then choose the one that provides the most useful insights.

X. Proposed Method for evaluation

The insights derived from the analysis will be evaluated based on their relevance to the problem statement and their potential impact on policing strategies. We will also use a portion of the dataset as a hold-out test set to evaluate the robustness of our insights on unseen data. The visualizations and outputs will be achieved using various techniques, analytics methods, and tools.

XI. Data Analysis and Results

The NYPD Arrest data analysis can make us understand the crime trends and take proactive measures to determine and prevent criminal activity, which will eventually improve public safety. The following methodology has been employed for the analysis of the NYPD Arrest Data:

A. Data Collection And Cleaning

The NYPD Arrest Data (Year-to-Date) dataset was successfully collected and subsequently put through a thorough examination process, that involved a detailed analysis of the entire dataset. During the course of this analysis, several instances of missing and null values were identified within the dataset.

Python was used to carry out the data cleaning process to keep the data's dependability and integrity. The impacted rows were eliminated where there were only a small number of missing values, and they did not significantly affect the analysis.

In this data cleaning and preprocessing effort, we transformed an initial dataset comprising 112,507 rows into a refined dataset containing 3,104 entries. The data cleaning process involved addressing critical issues related to missing values and outliers to ensure that the dataset was in a suitable state for analysis. Missing data, when present, can significantly impact the reliability of any analysis. To tackle this issue, we employed various techniques such as imputation and removal of rows with missing values, where appropriate. Furthermore, outliers, which can skew statistical analyses and visualizations, were identified and either removed or adjusted as necessary to enhance the dataset's overall quality. By diligently executing these data cleaning and preprocessing steps, we have successfully transformed the initial dataset into a more manageable and reliable dataset, ready for in-depth analysis and meaningful insights. This rigorous data preparation is essential for ensuring the accuracy and credibility of our subsequent research and findings.

B. Exploratory Data Analysis And Visualizations:

1. Temporal Trends of Arrests in a Year

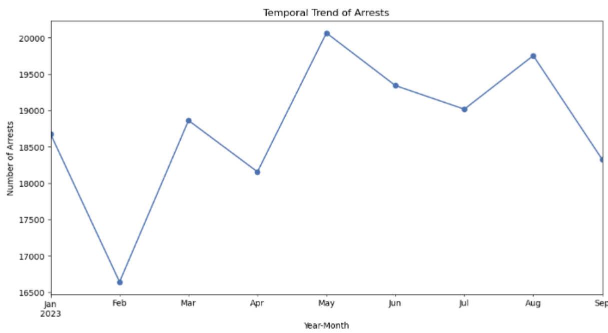

Figure-1: Temporal Trends of Arrests in a Year

The analysis of temporal trends in arrests is in response to the first research question showing arrests from January to September reveals a fluctuating pattern, ranging between 17,000 and 20,000 incidents. May witnessed the highest number of arrests, approaching the upper limit, while February recorded the lowest, slightly exceeding 16,500. This observed variability underscores the dynamic nature of law enforcement outcomes over time. These findings suggest a significant variation in law enforcement activities over time. The peak in arrests during May suggests a notable surge in law enforcement activities, possibly influenced by various factors such as changes in policing strategies, increased law enforcement presence, or heightened criminal activities during that period. Conversely, the decline observed in February may indicate a relative lull in law enforcement actions during that month. Possible explanations for these fluctuations include shifts in law enforcement strategies, seasonal changes in crime rates, or broader societal factors. This temporal analysis not only provides a snapshot of arrests but also prompts further inquiry into the dynamic interplay of law enforcement and contextual influences, offering valuable insights for policymakers, law enforcement agencies, and researchers.

2. Spatial Distribution of Sex-Related Crimes by Borough in New York City

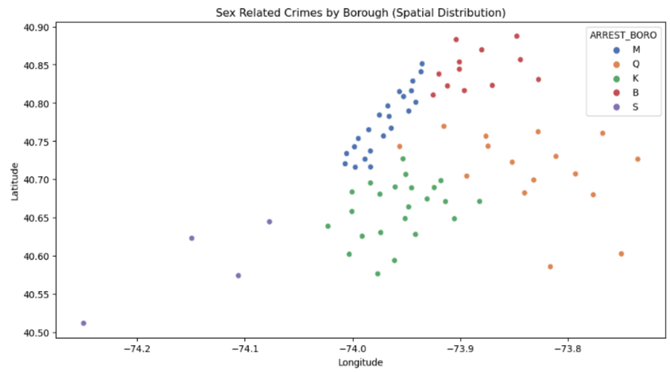

Figure-2: Sex Related Crimes By Borough

The spatial distribution of sex-related crimes in New York City, corresponding to the second research question as depicted in the scatter plot, illuminates distinct patterns across the boroughs. Notably, Queens and Brooklyn exhibit the highest concentrations of such crimes, represented in orange and green respectively, with Manhattan following in blue, and the Bronx in red. Staten Island, portrayed in purple, also registers instances of sex crimes, albeit at a lower concentration compared to the other boroughs. The observed disparities suggest that the Bronx, Queens and Brooklyn stand out with a higher prevalence of sex-related crimes. These findings hold practical significance, providing valuable insights for law enforcement agencies to strategically allocate resources and intensify efforts in these identified areas. By acknowledging and addressing the spatial nuances of sex crimes, law enforcement can enhance targeted interventions and contribute to the overall safety and well-being of communities within New York City.

3. Comparative Analysis of Arrest Rates Among Racial and Ethnic Groups in New York City

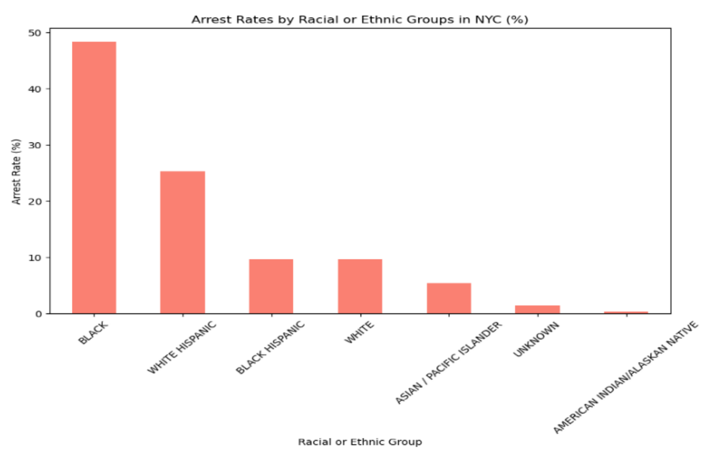

Figure-3: Arrest Rates by Racial or Ethnic Groups in NYC

The comparative analysis of arrest rates among racial and ethnic groups in New York City which is in response to our third research question underscores stark disparities, revealing that

Black and White Hispanic individuals face significantly higher arrest rates than other groups, with rates at approximately 50% and 25% respectively. In contrast, White and Black Hispanic individuals exhibit considerably lower arrest rates, hovering around 10%. The arrest rates for Asian/Pacific Islander, Unknown, and American Indian/Alaska Native individuals are notably the least, standing at less than 10%. These findings illuminate a pressing concern regarding unequal treatment within the criminal justice system. The observed disparities prompt a critical examination of the underlying factors contributing to these arrest rate variations. This research serves as a foundational step, signaling the need for in-depth investigations into the root causes of these disparities and the development of targeted strategies to address the pronounced imbalances in arrest rates among different racial and ethnic groups in New York City.

4. Analysis of Crime Frequency Across Different Age Groups

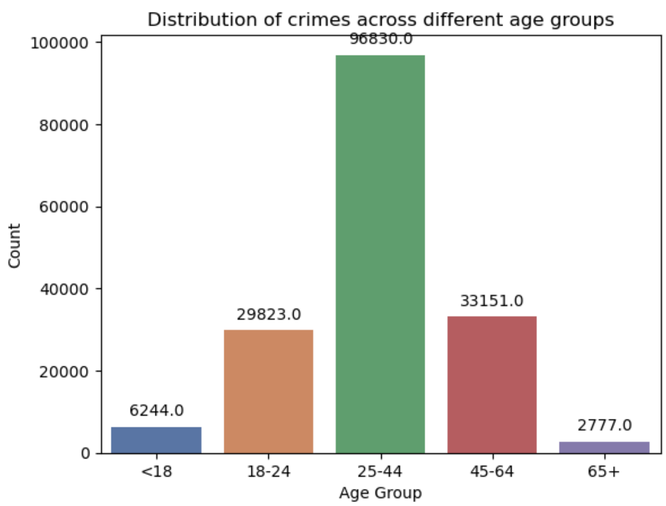

Figure-4: Distribution of Crimes across different age groups.

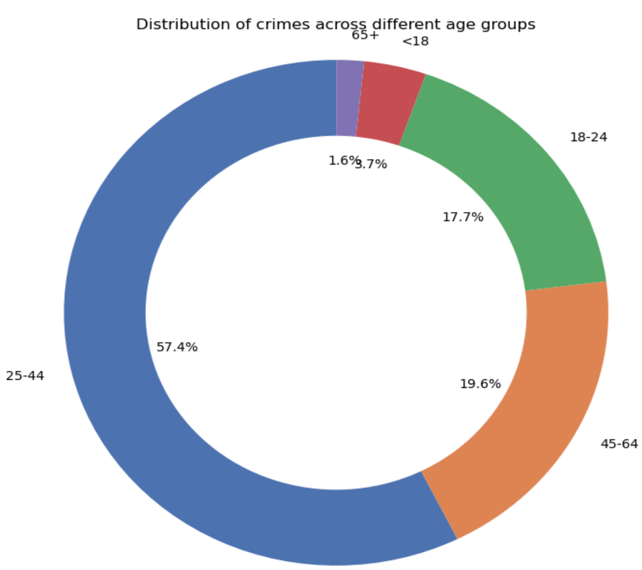

Figure-5: Distribution of Crimes across different age groups.

The analysis of crime frequency across different age groups corresponding to our fourth research question, as depicted in the pie chart and bar graph, reveals compelling patterns in criminal involvement. Notably, the 25-44 age group emerges as the primary contributor, with an alarming count exceeding 96,000 crimes, constituting 57.4% of the total. Following closely, the 45-64 age group registers over 33,000 crimes, representing 19.6% of the total, and the 18-24 age group exhibits more than 29,000 crimes, accounting for 17.7%. Conversely, individuals under 18 and those over 65 are implicated in the least amount of crimes. These findings prompt consideration of various factors influencing crime rates within specific age cohorts, such as the demographic distribution, engagement in activities leading to arrests, or broader societal influences. The prominence of the 25-44 age group in criminal activities suggests the need for a closer look into the various factors at play, helping shape more effective strategies for law enforcement and broader societal interventions to prevent crime across different age groups.

5. Distribution of Top Crimes Across Different Boroughs in New York City

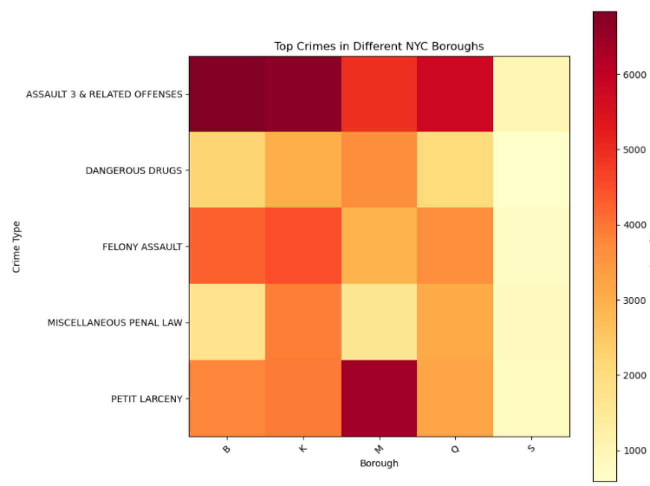

Figure-6: Heat Map of Top Crimes in different NYC Boroughs.

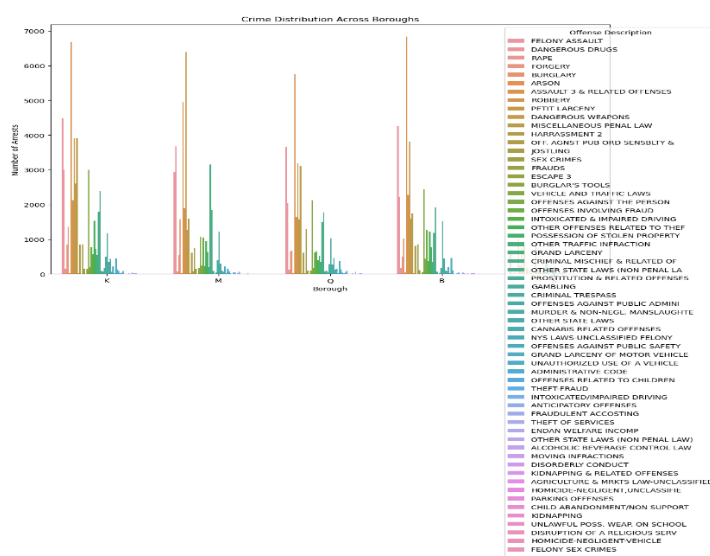

Figure-7: Stacked Bar Chart of Top Crimes in different NYC Boroughs.

The examination of crime distribution across different boroughs in New York City referring to our fifth research question, as portrayed in the stacked bar graph and heat map, unveils distinct patterns in the prevalence and variety of crimes. Notably, Assault 3 & Related Offenses, Dangerous Drugs, Felony Assault, Miscellaneous Penal Law, and Petit Larceny emerge as the top crimes in various NYC boroughs. Brooklyn, Manhattan, and the Bronx notably exhibit significantly higher numbers of crimes, with Queens following closely behind, while Staten Island registers the least number of crimes. Moreover, the types of crimes committed in Brooklyn, Manhattan, and the Bronx display a greater diversity compared to other boroughs. These findings not only shed light on the distribution of top crimes across different boroughs but also underscore the varying crime landscapes, emphasizing the need for nuanced and targeted law enforcement strategies. This information serves as a valuable snapshot of the crime situation in NYC boroughs, aiding law enforcement agencies in directing resources and efforts effectively to address specific crime challenges in each area.

6. Gender Disparities in Arrest Frequencies

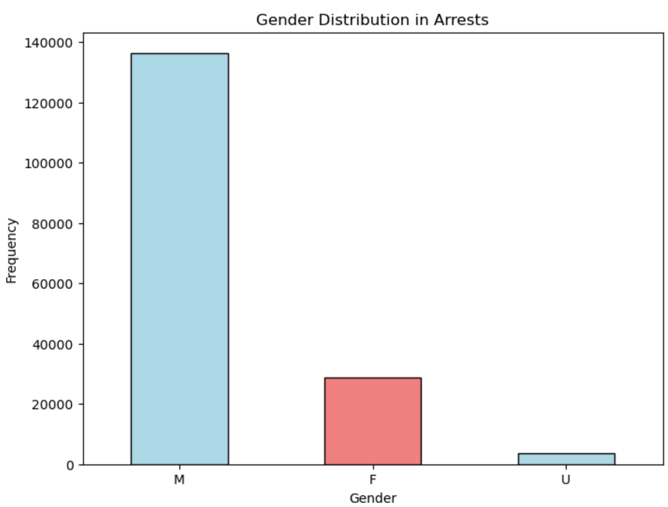

Figure-8: Gender Distribution in Arrests

The analysis of arrest distribution by gender, as depicted in the bar graph, unveils a notable gender disparity in arrest frequencies. Males significantly outnumber females in arrests, with nearly 140,000 instances compared to approximately 35,000 arrests for females. Individuals with unknown gender face the lowest arrest count, barely reaching 15,000. These findings highlight a pronounced discrepancy in arrest rates between males, females, and those with an unknown gender. The reasons for this contrast could stem from a myriad of factors, including divergent behavioral patterns, societal expectations, or potential biases within the criminal justice system. Understanding and addressing the root causes of these gender-based arrest differentials is pivotal for fostering a more equitable and just legal landscape. These findings contribute valuable insights into the dynamics of gender-related arrest frequencies, prompting further exploration into the multifaceted factors influencing law enforcement interactions and outcomes.

7. Comparative Analysis of Offenses Among Different Age Groups in the United States

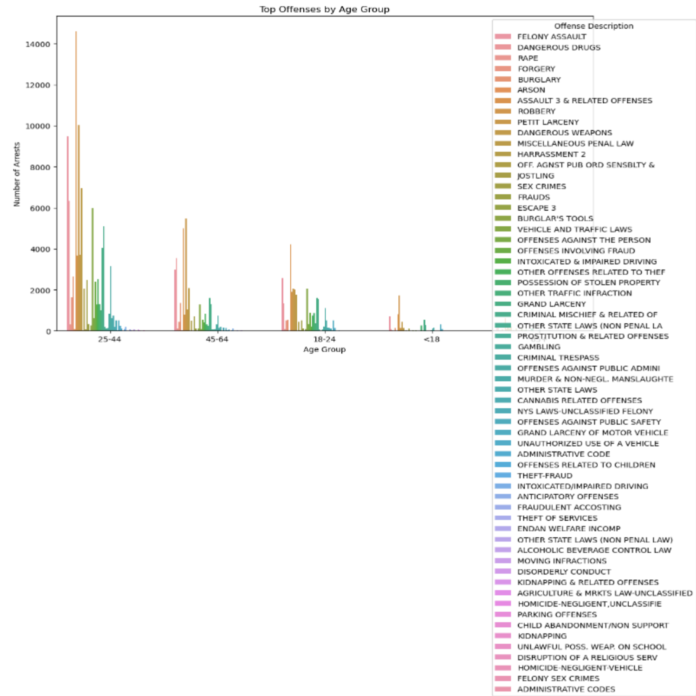

Figure-9: Stacked Bar Chart of Top Offenses among Different Age Groups.

The bar graph illuminates the predominant offenses committed by various age groups, offering insightful distinctions in criminal patterns. Among individuals aged 25-44, "Dangerous Drugs" emerges as the top offense, while the 18-24 age group is characterized by a higher prevalence of "Felony Assault." In the 45-64 age category, "Miscellaneous Penal Law" claims the top spot, while individuals aged 65 and above predominantly engage in "Offenses Against the Person." These findings underscore the dynamic nature of criminal behavior across age groups, suggesting that the most common types of offenses vary distinctly. The observed variations could be attributed to a myriad of factors encompassing differences in behavior, available opportunities, and inherent risk factors associated with each age cohort. This comparative analysis of offenses among different age groups not only contributes to our understanding of criminological trends but also provides a nuanced lens through which law enforcement and policymakers can tailor strategies to address specific age-related crime challenges effectively.

C. Criminal Hotspots Unveiled: Geospatial Visualization of High-Crime Areas

As a noteworthy aspect of our research, we used the longitude and latitude coordinates found in the dataset to visualize the geographic information. We developed visualizations to identify regions on a map with higher crime rates using geospatial frameworks in Python. To create a map display of the dataset's longitude and latitude information, we used the Python module Folium. With the help of this visualization tool, we were able to pinpoint geographical areas with higher crime rates, giving our project's goals a valuable perspective.

Below is an example of the map view to get a detailed understanding about the map and the details of the crimes. Please investigate the pdf document attached to the project and the website for navigating through the map and better view.

Figure-10: Geospatial Distribution of Crime Counts in New York City

The map identifies regions with larger concentrations of criminal activity, which can help law enforcement better use their resources to combat crime in those areas. The visualizations not only pinpoint areas with higher crime concentrations but also afford valuable insights for law enforcement, aiding in the strategic allocation of resources to combat crime effectively. Local governments and policymakers can utilize this information to tailor crime prevention initiatives, addressing the specific challenges faced in each area. Notably, the identification of crime clusters around transportation hubs underscores the potential need for heightened security measures in these locations. Overall, our geospatial analysis contributes to a more informed and targeted approach to crime prevention, aligning with the overarching goal of enhancing public safety in New York City.

D. Predictive Modeling: Random Forest and Gradient Boosting Classifiers

We employed machine learning models such as Random Forest Classifier and Gradient Boosting Classifier, to predict the gender of individuals involved in criminal activities based on selected features.

Accuracy: 0.81

	precision	recall	f1-score	support
F	0.41	0.01	0.03	5775
M	0.81	0.99	0.89	27538
U	0.21	0.01	0.02	706
accuracy			0.81	34019
macro avg	0.47	0.34	0.31	34019
weighted avg	0.73	0.81	0.73	34019

Figure-11: Random Forest Classifier Model Results.

Accuracy (Gradient Boosting Classifier): 0.81

```
C:\Users\munna\anaconda3\Lib\site-packages\sklearn\metr
re are ill-defined and being set to 0.0 in labels with
vior.
  _warn_prf(average, modifier, msg_start, len(result))
```

	precision	recall	f1-score	support
F	0.00	0.00	0.00	5775
M	0.81	1.00	0.89	27538
U	0.00	0.00	0.00	706
accuracy			0.81	34019
macro avg	0.27	0.33	0.30	34019
weighted avg	0.66	0.81	0.72	34019

Figure-12: Gradient Boosting Classifier Model Results.

Both the Random Forest and Gradient Boosting Classifier models have an overall accuracy of 0.81. However, they differ significantly in their performance across different classes:

- The Random Forest model has a high precision, recall, and f1-score for class ‘M’ but struggles with classes ‘F’ and ‘U’. It has a particularly low recall for these classes, indicating a high number of false negatives. These findings suggest that the model performs well in correctly identifying positive cases but struggles with accurately identifying negative cases.
- The Gradient Boosting Classifier model performs well for class ‘M’ but is not able to correctly predict classes ‘F’ and ‘U’ at all, as indicated by the precision, recall, and f1-score of 0.00 for these classes. These findings suggest that while the model performs well overall and particularly for class ‘M’, it struggles with accurately predicting classes ‘F’ and ‘U’.

In summary, while both models have the same overall accuracy, the Random Forest model provides a more balanced performance across different classes compared to the Gradient Boosting Classifier model. These findings suggest that the model performs well in correctly identifying positive cases but struggles with accurately identifying negative cases. However, both models could benefit from further tuning or additional features to improve their performance for classes ‘F’ and ‘U’.

XII. Related Work

In delving into the complex dynamics of crime patterns, law enforcement strategies, and community impact in the vibrant backdrop of New York City, our research stands at the intersection of several seminal studies and analyses. The multifaceted nature of our inquiry necessitates a comprehensive understanding of existing research to contextualize our findings. Here, we provide an overview of relevant studies that have laid the groundwork for our investigation, illuminating key themes such as temporal trends, spatial distribution, racial disparities, age-related crime dynamics, and gender-based variations.

- Temporal Trends in Arrests: According to the article, by Carolyn Rebecca Block Statistical Analysis Centre, some crimes are more common during specific seasons or times of the day. For example, burglary and larceny are more common in the summer, while homicide has no clear seasonal pattern.[22] The reasons for these patterns are not fully understood. Our research into the temporal trends in arrests aligns with the broader exploration of seasonal crime patterns, as noted in Block's study1. While Block's work touches on variations during different times of the year, our analysis takes a step further, uncovering nuanced temporal fluctuations from January to September. The peak in May suggests a surge in law enforcement activities, possibly influenced by factors like changes in policing strategies, increased law enforcement presence, or heightened criminal activities during that period.

- Spatial Distribution of Sex-Related Crimes: A study by David Weisburd, Taryn Zastrow analyzing NYC crime data from 2010-2020 found that 1% of streets accounted for 25% of crime, with high-crime areas concentrated in Manhattan, Bronx, and Brooklyn. These areas likely had higher sex crime rates, and though overall crime declined, underreporting may mask the true prevalence of sex offenses.[27] Examining the spatial distribution of sex-related crimes, our findings resonate with Weisburd and Zastrow's study on NYC crime hotspots. Our research delves specifically into sex-related crimes, revealing distinct patterns across boroughs. Queens, Brooklyn, and the Bronx emerge as areas with a higher prevalence of sex-related crimes, providing actionable insights for law enforcement to strategically allocate resources based on borough-specific prevalence.
- Bill Hutchinson says, there are significant differences in arrest rates for different racial and ethnic groups in the United States. According to the article you linked, Black Americans accounted for 52% of all arrests in New York City in 2020, despite only making up 24% of the city's population.[23] This disparity is even more pronounced for certain offenses, such as low-level marijuana possession, where Black people were arrested at a rate 14 times higher than white people. The stark disparities in arrest rates among racial and ethnic groups, a key finding in our research, align with Hutchinson's exploration of racial disparities in arrest rates in NYC. While Hutchinson's work addresses the general racial disparities, our study contributes by highlighting specific racial and ethnic groups, prompting further investigations into the underlying causes of arrest rate imbalances and emphasizing the need for targeted strategies.
- Age-wise Distribution: The age-wise distribution of arrests highlighted the dominance of the 25-44 age group, mirroring the insights from Georgia Worrell and Tina Moore.[24] Their work emphasized the highest arrest rates in the 25-34 age range and the prevalence of violent crimes in the 18-24 age group. Our findings align, emphasizing the significance of understanding age-specific crime dynamics for effective law enforcement strategies.
- NYC's violent crime By Steve Cuozzo is concentrated in specific, mostly lower-income neighborhoods like Mott Haven and Brownsville, while areas like the Upper East Side and Park Slope remain safe. Our study on crime distribution across NYC boroughs aligns with Cuozzo's research on localized violent crime in NYC. While Cuozzo's work focuses on overall crime rates, our study goes beyond, providing insights into specific crimes in different boroughs. This nuanced understanding aids in targeted law enforcement interventions, acknowledging the varying crime landscapes in different parts of the city.

XIII. Limitations

Data Quality and Completeness: Incomplete or inconsistent reporting of crimes may produce biased results; the dataset may contain missing or erroneous information that could impair the effectiveness of machine learning models.

Representativeness: Due to reporting and enforcement biases, the dataset only contains recorded arrests; unreported crimes are excluded, which may introduce bias. As a result, demographic data may not accurately reflect the community.

Temporal Dynamics: Crime trends are subject to change over time, and the dataset might not take seasonality or current developments into account. It might also not consider outside influences like changes in policy.

Geospatial Resolution: Limited geographic information may overlook differences within smaller areas; precinct-level geographic coordinates may not be fine enough for accurate hotspot detection.

Data Imbalance: Unbalances in the distribution of crime types or locations may influence the evaluation and training of machine learning models.

Legal and Ethical Considerations: It is important to ensure that analytical results are used ethically and to prevent biases in law enforcement activities because sensitive information may give rise to privacy concerns.

Causation vs. Correlation: Additional context and domain expertise are necessary to comprehend the causal reasons underlying crime trends. Correlations in the data do not suggest causation.

Community Engagement: If the community isn't included, methods may be developed that don't specifically meet its wants and concerns.

Interpretability of the Model: It might be difficult to explain predictions in complex machine learning models if they are not interpretable.

Resource Allocation and Deployment: It is important to exercise caution when implementing predictive models for law enforcement resource allocation to prevent unintended outcomes or the reinforcement of biases.

XIV. Analytical Conclusion

The temporal analysis of arrests spanning January to September underscores a dynamic pattern, fluctuating between 17,000 and

20,000 incidents. The peak in May, nearing the upper limit, suggests a potential surge in law enforcement activities, influenced by factors like strategic shifts, increased police presence, or heightened criminal activities. Conversely, the dip in February indicates a relative lull, prompting exploration into the reasons behind these fluctuations. Shifts in law enforcement strategies, seasonal changes in crime rates, or broader societal factors may all contribute. This temporal lens not only provides a snapshot of arrests but prompts further inquiry into the dynamic interplay of law enforcement and contextual influences, offering valuable insights for policymakers, law enforcement agencies, and researchers.

The spatial distribution of sex-related crimes across NYC boroughs reveals distinct patterns. Queens and Brooklyn exhibit the highest concentrations, necessitating strategic resource allocation and intensified efforts in these areas. Understanding and addressing spatial nuances of sex crimes can enhance targeted interventions, contributing to overall community safety.

A comparative analysis of arrest rates among racial and ethnic groups exposes stark disparities, with Black and White Hispanic individuals facing significantly higher rates. White and Black Hispanic individuals exhibit lower rates, necessitating a critical examination of contributing factors and the development of targeted strategies. This research serves as a foundational step, signaling the need for in-depth investigations into the root causes of these disparities.

The analysis of crime frequency across age groups reveals compelling patterns. The 25-44 age group emerges as the primary contributor, prompting consideration of demographic distributions, engagement in activities leading to arrests, or broader societal influences. The prominence of this age group suggests the need for effective strategies for law enforcement and societal interventions to prevent crime.

Examining crime distribution across NYC boroughs uncovers distinct patterns in the prevalence and variety of crimes. Certain boroughs exhibit significantly higher numbers and a greater diversity of crimes, emphasizing the need for nuanced and targeted law enforcement strategies. This information aids law enforcement agencies in directing resources effectively to address specific crime challenges in each area.

The analysis of arrest distribution by gender unveils a notable gender disparity, with males significantly outnumbering females in arrests. Understanding and addressing the root causes of these gender-based arrest differentials is crucial for fostering a more equitable and just legal landscape.

The bar graph detailing predominant offenses by age groups provides insightful distinctions in criminal patterns. This comparative analysis contributes to understanding criminological trends and offers a nuanced lens for law enforcement and policymakers to tailor strategies to address specific age-related crime challenges effectively.

In conclusion, The analysis of NYPD Arrest Data reveals intricate patterns in temporal, spatial, demographic, and offense-related dimensions of arrests in New York City. The research transcends data examination, offering valuable insights into the dynamics of NYC crime. It informs policymakers, law enforcement, and researchers, striving for a safer, more equitable cityscape. The findings emphasize the need for targeted law enforcement strategies and policy interventions to address disparities and effectively allocate resources.

XV. Future Work

Further analysis of the data underscores the intricate interplay between temporal, spatial, demographic, and gender-related dimensions in New York City's crime landscape. Subsequent research on this subject might concentrate on improving prediction models with the addition of new datasets that include indicators for community involvement, urban growth, and socioeconomic issues. A more comprehensive picture of crime dynamics may be possible by combining data from social services, community organizations, and other law enforcement authorities. Moreover, implementing cutting-edge machine learning methods like deep learning may improve forecast accuracy. Examining how external events—like public gatherings or changes in policy—affect crime trends may shed light on how criminal activity is changing over time. Geospatial technologies and real-time data streams may also make it possible for crime prediction models to be more responsive and dynamic. Sustaining stakeholder interaction should be given top priority in future investigations to guarantee the applicability and moral application of predicted insights. Sustaining the success of predictive police initiatives in New York City will require regularly adjusting models to changing urban landscapes and assessing the long-term efficacy of methods that have been put into practice.

References

[1] 'NYPD Arrest Data (Year to Date)'. data.cityofnewyork.us, Nov. 08, 2023. Available: https://catalog.data.gov/dataset/nypd-arrest-data-year-to-date .

[2] V. Mandalapu, L. Elluri, P. Vyas, and N. Roy, 'Crime Prediction Using Machine Learning and Deep Learning: A Systematic Review and Future Directions', IEEE Access, vol. 11, pp. 60153–60170, 2023, doi: 10.1109/ACCESS.2023.3286344. Available: https://ieeexplore.ieee.org/abstract/document/10i1873 .

[3] '"Minority Report" a Reality? The NYPD's Big Data Approach to Predicting Crime', Technology and Operations Management. Available: https://d3.harvard.edu/platform-rctom/submission/minority-report-a-reality-the-nypds-big-data-approach-to-predicting-crime/ .

[4] J. MacDonald, J. Fagan, and A. Geller, 'The Effects of Local Police Surges on Crime and Arrests in New York City', PLOS ONE, vol. 11, no. 6, p. e0157223, Jun. 2016, doi: 10.1371/journal.pone.0157223. Available: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0157223 .

[5] S. Chainey, L. Tompson, and S. Uhlig, 'The Utility of Hotspot Mapping for Predicting Spatial Patterns of Crime', Secur J, vol. 21, no. 1–2, pp. 4–28, Feb. 2008, doi: 10.1057/palgrave.sj.8350066. Available: http://link.springer.com/10.1057/palgrave.sj.8350066 .

[6] '2020 Use of Force Report', Available: https://www.nyc.gov/assets/nypd/downloads/pdf/use-of-force/use-of-force-2020-issued-2021-12.pdf

[7] S. Kajeepeta, E. Bruzelius, J. Z. Ho, and S. J. Prins, 'Policing the pandemic: estimating spatial and racialized inequities in New York City police enforcement of COVID-19 mandates', Critical Public Health, vol. 32, no. 1, pp. 56–67, Jan. 2022, doi: 10.1080/09581596.2021.1987387. Available: https://www.tandfonline.com/doi/full/10.1080/09581596.2021.1987387 .

[8] 'Compstat: Its Origins, Evolution, and Future in Law Enforcement Agencies | Office of Justice Programs'. Available: https://www.ojp.gov/ncjrs/virtual-library/abstracts/compstat-its-origins-evolution-and-future-law-enforcement-agencies .

[9] B. Mendes, 'Analysis of NYC Reported Crime Data Using Pandas', Medium, Feb. 27, 2021. Available: https://towardsdatascience.com/analysis-of-nyc-reported-crime-data-using-pandas-821753cd7e22 . [Accessed: Nov. 13, 2023].

[10] 'Read the document', The New York Times, Mar. 20, 2021. Available: https://www.nytimes.com/interactive/2021/03/20/us/newyork-policing-DOI.html .

[11] C. Smith, 'The Crime-Fighting Program That Changed New York Forever', Intelligencer, Mar. 02, 2018. Available: https://nymag.com/intelligencer/2018/03/the-crime-fighting-program-that-changed-new-york-forever.html .

[12] 'New York City Police Department's Response to Demonstrations Following the Death of George Floyd', Available: https://ag.ny.gov/sites/default/files/2020-nypd-report.pdf

[13] 'Discipline Reports - NYPD'. Available: https://www.nyc.gov/site/nypd/stats/reports-analysis/discipline.page .

[14] 'NYPD Announces Citywide Crime Statistics for October 2022', The official website of the City of New York, Nov. 03, 2022. Available: http://www.nyc.gov/site/nypd/news/p00066/nypd-citywide-crime-statistics-october-2022 .

[15] 9211, 'A Proposal for an NYPD Inspector General | Brennan Center for Justice'. Available: https://www.brennancenter.org/our-work/policy-solutions/proposal-nypd-inspector-general .

[16] 'NYC Crime', New York Daily News, Dec. 09, 2023. Available: https://www.nydailynews.com/2023/12/09/rikers-island-investigators-demotion-raises-serious-concerns-to-nyc-jails-violence-monitor/ .

[17] D. Weisburd, C. W. Telep, and B. A. Lawton, 'Could innovations in policing have contributed to the New York city crime drop even in a period of declining police strength?: The case of stop, question and frisk as a hot spots policing strategy', Justice Quarterly, vol. 31, no. 1, pp. 129–153, 2014, doi: 10.1080/07418825.2012.754920. Available: http://www.scopus.com/inward/record.url?scp=84872198242&partnerID=8YFLogxK .

[18] W. Bratton, 'New Strategies for Combatting Crime in New York City', Fordham Urban Law Journal, vol. 23, no. 3, p. 781, Jan. 1996, Available: https://ir.lawnet.fordham.edu/ulj/vol23/iss3/8 .

[19] 'Crime and Enforcement Activity in New York City ', Available: https://www.nyc.gov/assets/nypd/downloads/pdf/analysis_and_planning/year-end-2022-enforcement-report.pdf

[20] C. Wood, 'How the NYPD is using machine learning to spot crime patterns', StateScoop, Mar. 19, 2019. Available: https://statescoop.com/how-the-nypd-is-using-machine-learning-to-spot-crime-patterns/

[21] 'Predictive policing algorithms are racist. They need to be dismantled.', MIT Technology Review. Available: https://www.technologyreview.com/2020/07/17/1005396/predictive-policing-algorithms-racist-dismantled-machine-learning-bias-criminal-justice/ .

[22] Carolyn Rebecca Block , 'IS CRIME SEASONAL?', Available: https://bjs.ojp.gov/content/pub/pdf/ics.pdf

[23] A. B. C. News, 'Blacks account for nearly half of all NYC arrests 6 years after end of stop-and-frisk: NYPD data', ABC News. Available: https://abcnews.go.com/US/blacks-account-half-nyc-arrests-years-end-stop/story?id=71412485

[24] V. A. Archive, G. author R. feed, V. A. Archive, E. the Author, F. on Twitter, and G. author R. feed, 'NYC violent crime arrests up for kids, NYPD data shows', May 27, 2023. Available:

https://nypost.com/2023/05/27/nyc-violent-crime-arrests-up-for-kids-nypd-data-shows/

[25] V. A. Archive, E. the Author, F. on Twitter, and G. author R. feed, 'Why NYC's violent crime is totally local', Jul. 29, 2023. Available: https://nypost.com/2023/07/29/why-nycs-violent-crime-is-totally-local/.

[26] 'NYPD launches new community policing plans', Police1, Jun. 26, 2015. Available: https://www.police1.com/patrol-issues/articles/nypd-launches-new-community-policing-plans-gMNNW6lVKcIUgvM6/

[27] David Weisburd and Taryn Zastrow, 'Crime Hot Spots: A Study of New York City Streets in 2010, 2015, and 2020'. Available: https://manhattan.institute/article/crime-hot-spots-a-study-of-new-york-city-streets-in-2010-2015-and-2020