AIT 664-009: Represent, Process & Visualize Applied Information Technology

PROJECT PROPOSAL

Prof. Ebrima N Ceesay, Ph.D., CISSP

Analyzing New York City Police Department (NYPD) Arrest Data for Crime Insights and Policing Strategies

Introduction:

New York City, one of the world's busiest cities, is home to a diverse range of neighborhoods, 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 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]

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. The New York Police Department (NYPD) keeps a detailed record of all arrests made throughout the year. 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. The aim of this project is to analyze and predict crime patterns using machine learning techniques based on this dataset. By identifying crime trends, hotspots, and crime types, governments and police departments can allocate resources more effectively and deploy their resources more strategically.

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.

Literature Review:

Several literature searches were conducted for this study**.**

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. The study highlights 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. [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, misdemeanor offences, felony property crimes, robberies, and felony violent crimes. [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 analyze 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. [5]

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.

Research Questions:

The research questions that can be answered by exploring the dataset include:

1. Are certain crimes more common during specific seasons or times of the day?
2. Where are the areas with the most arrests for different crimes in NYC?
3. Are there differences in arrest rates for different racial or ethnic groups?
4. Which age group has committed more crimes?
5. Which area has more sex crimes?

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]

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.

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.

Project Time Plan:

The project is expected to be completed in 10 to 12 weeks, with the following timeline:

PROJECT TIMELINE	*TASK DESCRIPTION*
WEEK 1-2	Project Proposal and Data Collection
WEEK 3-4	Data Cleaning, Pre-processing and Exploratory data analysis
WEEK 5-6	Descriptive analysis, Insight Generation and Data visualization
WEEK 7-8	Interpretation of Statistical Analysis and Findings
WEEK 9-10	Evaluation, Refinement of Insights and Finalization
WEEK 11-12	Report Writing and Project presentation

References

[1] ., "NYPD Arrest Data (Year to Date)," 14 July 2023. [Online]. Available: https://catalog.data.gov/dataset/nypd-arrest-data-year-to-date. [Accessed 20 September 2023].

[2] L. E. P. V. a. N. R. V. Mandalapu, "Crime Prediction Using Machine Learning and Deep Learning: A Systematic Review and Future Directions," [Online]. Available: https://ieeexplore.ieee.org/abstract/document/10151873. [Accessed 20 September 2023].

[3] Clemens, ""“Minority Report” a Reality? The NYPD’s Big Data Approach to Predicting Crime," [Online]. Available: https://d3.harvard.edu/platform-rctom/submission/minority-report-a-reality-the-nypds-big-data-approach-to-predicting-crime/. [Accessed 20 September 2023].

[4] J. F. a. A. G. J. MacDonald, "The Effects of Local Police Surges on Crime and Arrests in New York City," [Online]. Available: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0157223. [Accessed 20 September 2023].

[5] L. T. a. S. U. S. Chainey, "‘The Utility of Hotspot Mapping for Predicting Spatial Patterns of Crime’, Secur J, vol. 21, no. 1, pp. 4–28,," Feb 2008. [Online]. Available: https://doi.org/10.1057/palgrave.sj.8350066. [Accessed 20 September 2023].