**Predictive Crime Analytics: Geospatial Insights into Toronto’s Homicide and Crime Patterns**

Name: Syed Muzaffar Hasan Kazmi

Student ID: 501271881

Supervisor: Ceni Babaoglu, PhD

Date of Submission: October 28, 2024

Table of Contents

[**Abstract** 3](#_Toc181049105)

[**Introduction** 5](#_Toc181049106)

[**Background on Toronto’s Crime Landscape** 5](#_Toc181049107)

[**Overview of Predictive Crime Analytics and Geospatial/Temporal Crime Analysis** 5](#_Toc181049108)

[**Gaps in Existing Research** 5](#_Toc181049109)

[**Research Questions** 7](#_Toc181049110)

[**Literature Review** 8](#_Toc181049111)

[**Data Description** 14](#_Toc181049112)

[**Proposed Methodology** 17](#_Toc181049113)

[**Data Source** 20](#_Toc181049114)

[**GitHub Repository** 20](#_Toc181049115)

[**References** 21](#_Toc181049116)

# **Abstract**

In urban environments, public safety is a critical issue, with crime patterns and trends directly influencing policymaking, law enforcement strategies, and community welfare. In this project, the goal is to analyze Toronto Police Service crime data to provide predictive insights and identify patterns that could aid in proactive crime prevention. This project focuses on two distinct datasets from Toronto Police Services: one pertaining to homicide crimes and the other encompassing non-homicide crimes such as assault, break and enter, auto theft, robbery, theft, etc. The theme of this analysis is predictive analytics, using machine learning techniques to uncover trends and predict future crime occurrences.

The primary research objectives are as follows: First, to identify and visualize crime hotspots across Toronto by neighborhood and geographical coordinates, exploring which areas exhibit specific types of crimes. Second, to analyze trends in crime types over time, examining which locations and premises are increasingly susceptible to particular crimes. Third, to assess temporal crime trends, investigating how factors such as specific days of the year or week, and seasonal influences, correlate with crime activity. Building on these insights, the study aims to leverage machine learning to develop predictive models, identifying potential criminal activity patterns across Toronto’s neighborhoods and assessing how attributes like premises type, location, and timing might inform these predictions.

The publicly available datasets from Toronto Police Services offer detailed attributes such as offense date, neighborhood, location type, and specific crime type, enabling a comprehensive analysis of Toronto’s crime landscape. By employing classification models (e.g., decision trees, logistic regression) to forecast crime types based on temporal and spatial variables, and clustering methods (e.g., K-means, DBSCAN) to detect crime hotspots, this study aims to identify high-risk zones and provide spatial visualizations. These visualizations, such as heatmaps and geospatial plots, will illustrate the frequency and distribution of crimes across the city.

This project utilizes Python, with libraries such as pandas for data manipulation, scikit-learn for model development, Folium for interactive mapping, and Matplotlib and Seaborn for data visualization. By applying these methods to real-world data, the study demonstrates how predictive analytics can inform data-driven policing strategies, thereby supporting public safety efforts in Toronto and contributing to safer urban environments.

In conclusion, this project offers an end-to-end approach to analyzing and predicting crime patterns in Toronto, providing essential insights for urban safety initiatives and policy development. Through the integration of machine learning and geospatial analysis, this research highlights the potential of predictive tools in enhancing crime prevention and community safety.

# **Introduction**

## **Background on Toronto’s Crime Landscape**

Toronto, Canada’s largest city, has seen significant fluctuations in crime rates over the years, reflecting broader social dynamics, economic conditions, and policing strategies. As of 2023, Toronto reported an overall increase in crime rates compared to previous years, particularly in violent crime categories such as shootings and homicides (Toronto Police Service, 2023). Factors contributing to crime in urban areas like Toronto include socioeconomic disparities, neighborhood environments, and changing community dynamics (Crime Severity Index, 2023). Understanding the underlying patterns of crime is essential for informing public safety policies and resource allocation within the city.

## **Overview of Predictive Crime Analytics and Geospatial/Temporal Crime Analysis**

Predictive crime analytics involves utilizing statistical methods and machine learning techniques to forecast potential criminal activity, enabling law enforcement agencies to allocate resources more effectively (Lau, 2020). Geospatial and temporal analysis further enhances this approach by incorporating location and time into predictive models, allowing for a nuanced understanding of crime trends and patterns (Ioannidis, 2024, p. 14). Through the use of geospatial data, analysts can identify crime hotspots, revealing areas that are more susceptible to criminal activities (He et al., 2022).

## **Gaps in Existing Research**

While a substantial body of research examines the influence of socioeconomic factors—such as income levels, education, and age demographics—on crime activity, there is limited open-access research focusing on a more granular view of Toronto’s crime landscape. Specifically, I was unable to find publicly accessible studies that provide a detailed analysis of how temporal and geographical factors impact the likelihood of different crime types across Toronto neighborhoods. Understanding these localized patterns could offer valuable insights, particularly into the timing and frequency of specific crimes within certain neighborhoods.

Additionally, there appears to be a scarcity of research addressing how this granular crime data could inform police readiness in Toronto. By predicting the types of crimes more likely to occur in particular areas and at specific times, law enforcement could strategically allocate resources and prepare appropriate response measures tailored to the nature and severity of anticipated incidents. This project aims to bridge this gap by leveraging machine learning and geospatial analytics to develop insights that could enhance the effectiveness of crime prevention and response strategies in Toronto.

# **Research Questions**

1. Where are the primary hotspots of crime in Toronto, and how do crime types vary across different neighborhoods and geographical coordinates?

Goal: Identify and map crime hotspots across Toronto, examining prevalent types of crimes by neighborhood to support targeted safety initiatives.

1. What types of locations and premises are most susceptible to specific types of crime, and how have these susceptibilities changed over time?

Goal: Determine trends in location-based crime patterns, analyzing whether certain types of locations (e.g., residential, commercial) and premises (e.g., schools, parks) exhibit increasing or decreasing rates for particular crimes.

1. How do temporal factors—such as the time of day, day of the week, and season—correlate with fluctuations in crime rates across Toronto?

Goal: Analyze time-based patterns in crime occurrences to identify whether specific times or days are consistently associated with higher crime activity and understand how this varies by crime type and location.

1. Can machine learning models effectively predict the type of crime based on attributes such as location, time, and premises type, and what are the primary indicators for each crime type?

Goal: Use classification models to predict crime types based on key factors and uncover the most significant predictors, providing actionable insights for preventive measures.

# **Literature Review**

Understanding crime patterns through spatial and temporal analysis is crucial for effective law enforcement and urban planning. Numerous studies have explored the intricate relationships between geographical locations, time variables, and crime incidents, yielding insights that significantly influence policy decisions and strategic resource allocation. This literature review examines the existing body of research, critiques the methodologies employed, and highlights gaps that the current study aims to address.

In the doctoral thesis by Ioannidis (2024), titled *Understanding Crime Patterns Using Spatial Data Analysis: Case Studies in Stockholm, Sweden*, the author explores spatial data analysis as a tool for understanding urban crime patterns, using Stockholm as a case study. The study integrates spatial crime data with remote sensing and traditional socio-demographic indicators to enhance crime prediction models. Ioannidis' work particularly focuses on cannabis-related crimes, street theft, residential burglaries, and sexual crimes, applying a combination of classical regression and machine learning techniques, notably the Random Forest classifier, to assess crime hotspots and criminogenic environments.

The methodology employs high-resolution imagery and spatio-temporal data, which provide granular insights into environmental and urban factors influencing crime, such as vegetation, built density, and illumination levels. A notable strength in Ioannidis' approach is the integration of remote sensing data with spatial crime analysis, enabling detailed predictions and insights into specific urban features contributing to crime occurrences.

Compared to Ioannidis’ research, my study on Toronto crime patterns has a similar focus on predictive analytics but diverges in the data sources and crime types. While Ioannidis employs remote sensing data for physical environmental factors, my project leverages Toronto Police Service data to explore crime across various neighborhoods, focusing on patterns in non-homicide and homicide incidents. Ioannidis’ limitations include potential biases from data aggregation at different spatial scales, which might affect prediction accuracy—a limitation my study partially mitigates by employing a neighborhood-specific approach without relying on high-resolution remote sensing data.

In summary, Ioannidis' research contributes significantly to understanding how urban characteristics impact crime, particularly through the novel use of remote sensing in criminology. However, my project builds on this by focusing more specifically on the spatiotemporal attributes of crime within Toronto, aiming to create insights for local law enforcement that address neighborhood-specific crime types and trends.

The study by He et al. (2022) explores four prominent spatial analysis techniques—spatial scan statistics, local Moran’s I, Getis-Ord G\*, and the AMOEBA method—to detect crime hotspots with a focus on balancing concentration and shape characteristics. Their findings emphasize that, while each method has merits, AMOEBA uniquely excels in identifying irregular, complex hotspot shapes, a factor often overlooked in traditional methods focused solely on intensity. This ability to capture non-standard shapes makes AMOEBA a robust tool for detailed hotspot analysis, though its irregularity poses challenges in practical deployment, as the delineation of amorphous hotspot boundaries complicates consistent tracking and resource allocation.

This research aligns closely with my study’s goal of developing predictive models for crime patterns, especially in refining hotspot detection and understanding the spatial dimensions of crime occurrences. While He et al. prioritize shape complexity and hotspot definition, my research could build on these insights by examining how temporal factors influence these spatial hotspots over time, potentially applying AMOEBA or similar models to Toronto’s distinct crime landscape.

However, He et al. also reveal limitations in the methods’ ability to integrate temporal dynamics and socio-environmental factors alongside spatial clustering. This gap indicates an opportunity for my research to contribute by incorporating these dynamic variables, thereby enhancing the predictive power and operational relevance of hotspot detection in urban crime analysis.

The article, *The Spatial and Social Patterning of Property and Violent Crime in Toronto Neighbourhoods: A Spatial-Quantitative Approach* by Wang, Lee, and Williams (2019), presents a spatial analysis of property and violent crime across Toronto's neighborhoods, aiming to identify crime patterns and associated socioeconomic factors. The authors leverage spatial techniques like Local Moran's I for clustering analysis and apply both OLS and GWR models to assess relationships between crime rates and socioeconomic variables, including demographic characteristics of offenders. This study reveals significant clustering of both property and violent crimes, with high-crime areas predominantly located in the city core and areas of socioeconomic disadvantage. The findings highlight the complex socio-spatial patterns in crime, suggesting that factors like age and residential instability have varying impacts on crime rates across different areas.

While this research significantly contributes to understanding spatial crime patterns in Toronto, it primarily focuses on neighborhood-level influences and uses aggregated data, which could mask finer details. The reliance on neighborhood-level socioeconomic indicators, though informative, may limit insights into individual-level influences and more granular neighborhood characteristics, which our study attempts to address by focusing on detailed temporal and spatial variables within neighborhoods.

In contrast to Wang et al.'s emphasis on broad spatial clustering, our project further explores temporal patterns (e.g., days of the week or year) and integrates predictive modeling for crime types based on location and premises details. By employing machine learning techniques, our research seeks to predict crime occurrences in more precise locations and times, aiming to provide actionable insights for proactive crime prevention strategies.

The study "Homicide Rates Are Spatially Associated with Built Environment and Socio-Economic Factors" by Mohammadi et al. (2022) explores spatial and temporal patterns of homicides in Toronto between 2012 and 2021. The study uses Kernel Density Estimation, Moran’s I, Kulldorff’s SaTScan spatio-temporal analysis, Geographically Weighted Regression (GWR), and Multi-Scale GWR (MGWR) to analyze homicide data. It finds that homicide rates are clustered in downtown and northwest Toronto and significantly influenced by factors such as population density, commercial establishment density, and material deprivation. MGWR explains these variations well, demonstrating higher model accuracy than OLS.

In contrast, my research focuses on predictive analytics using machine learning to determine temporal and geographic crime patterns across Toronto, focusing specifically on the timing, location, and nature of crimes without prioritizing socio-economic factors. Where Mohammadi et al. highlight socio-economic and environmental determinants, my study aims to model and predict high-risk areas and crime types, utilizing attributes like neighborhood, time of day, and crime type rather than broader socio-economic variables. This allows for proactive crime prediction and response strategies relevant to police readiness and resource allocation, bridging a gap in predictive geospatial crime modeling.

The paper by Xue Luo (2012), titled *Spatial Patterns of Neighbourhood Crime in Canadian Cities: The Influence of Neighbourhood and City Contexts,* investigates crime distribution in six Canadian cities, including Toronto, through a multilevel analysis of neighbourhood and city factors. Using data from the Uniform Crime Reporting Survey and Census data, Luo explores spatial dependencies and finds that crime rates are often clustered in areas near city centers, influenced by both local characteristics and broader city contexts. Luo employs spatial autocorrelation and multilevel regression methods to account for these patterns, highlighting the complexity of crime factors that extend beyond individual neighbourhoods.

Luo's approach aligns with existing ecological and multilevel theories, such as Social Disorganization Theory, to explain neighborhood variance but expands upon them by emphasizing spatial dependence and broader city factors. This study's findings are valuable for my research as it reinforces the importance of spatial factors in crime distribution, a key focus of my own project. While Luo's work contributes significantly to understanding multilevel crime dynamics, my research will delve deeper into predictive modeling, particularly focusing on time and location as crime determinants, filling a gap in the prediction of crime patterns.

**Conclusion**

The review of existing literature reveals a multifaceted landscape of crime analysis, with a strong emphasis on spatial and socio-economic factors. While significant strides have been made in understanding crime patterns through various methodologies, notable gaps remain, particularly regarding the integration of temporal factors and community perspectives. The current study aims to address these gaps by providing a comprehensive analysis of crime patterns in Toronto, focusing on both homicide and non-homicide data while incorporating qualitative insights. By leveraging advanced data analysis techniques and community engagement, the research aspires to contribute valuable knowledge to the field of predictive crime analytics, ultimately supporting more effective law enforcement strategies and community safety initiatives.

# **Data Description**

This project utilizes two primary datasets provided by the Toronto Police Service: the Homicide dataset and the Major Crime Indicators (MCI) dataset, referred to as "Non-Homicide" data in this report. These datasets offer a comprehensive view of reported crimes in Toronto, covering both violent and property crimes. Each dataset includes essential temporal, geographical, and categorical details, allowing for a multidimensional analysis of crime patterns across the city.

The **Homicide dataset** focuses exclusively on cases of homicide, providing detailed information on each incident, such as the unique offense identifier (EVENT\_UNIQUE\_ID) and the date and time (OCC\_DATE) when the crime occurred. Temporal variables like the year (OCC\_YEAR), month (OCC\_MONTH), and day (OCC\_DAY) are included, alongside day of the week (OCC\_DOW) and day of the year (OCC\_DOY), enabling an in-depth examination of temporal patterns in homicides. The dataset also includes geographical attributes such as the police division (DIVISION) where the incident occurred, as well as neighborhood identifiers based on both the old 140-neighborhood structure (HOOD\_140, NEIGHBOURHOOD\_140) and the current 158-neighborhood structure (HOOD\_158, NEIGHBOURHOOD\_158). These fields provide critical location-specific data that can help pinpoint and analyze spatial concentrations of homicides. Furthermore, coordinates are provided in latitude and longitude (LAT\_WGS84, LONG\_WGS84), as well as projected coordinate system fields (x and y), allowing for precise geospatial mapping and analysis. The dataset also categorizes homicides into types (HOMICIDE\_TYPE), such as shootings or stabbings, aiding in categorizing patterns across different crime methods.

The **Non-Homicide (MCI) dataset** encompasses a broader range of crime categories, including assault, break and enter, robbery, theft, and auto theft. This dataset contains a richer set of attributes, given the diversity of crime types it covers. Key temporal attributes include the report date (REPORT\_DATE) and the date the offense occurred (OCC\_DATE), as well as report-specific temporal information such as the year (REPORT\_YEAR), month (REPORT\_MONTH), day (REPORT\_DAY), day of the week (REPORT\_DOW), and report hour (REPORT\_HOUR). These time-related features allow for analysis of reporting patterns and the relationship between crime occurrences and reporting times. The dataset also includes additional fields for the occurrence year (OCC\_YEAR), month (OCC\_MONTH), day (OCC\_DAY), and occurrence hour (OCC\_HOUR), facilitating an investigation of crime timing in finer detail.

Spatial details are similarly comprehensive. The dataset specifies the police division (DIVISION) and location type (LOCATION\_TYPE), indicating the type of area where crimes occurred (e.g., residential or commercial locations). Additionally, premises type (PREMISES\_TYPE) offers further granularity, showing whether the crime took place in a private residence, public place, or business establishment. The inclusion of neighborhood fields (HOOD\_140, NEIGHBOURHOOD\_140, HOOD\_158, and NEIGHBOURHOOD\_158) mirrors the structure of the homicide data, supporting a consistent spatial analysis approach across both datasets. Geographical coordinates (LAT\_WGS84, LONG\_WGS84) allow for precise spatial mapping and enable the identification of crime hotspots across Toronto’s neighborhoods. Categorical fields, such as the Universal Crime Reporting (UCR) code (UCR\_CODE, UCR\_EXT), provide additional classification details, while a separate field for the offense name (OFFENCE) and major crime indicator category (MCI\_CATEGORY) helps differentiate among the various non-homicide crime types.

The **data sources** for both datasets stem from the Toronto Police Service’s open data initiative, ensuring that each dataset complies with privacy guidelines, including measures such as location offsetting to the nearest intersection rather than exact addresses. These datasets allow for a robust, multi-faceted analysis of crime in Toronto, exploring patterns across different crime categories, locations, and time periods, which will be foundational in examining crime trends, identifying hotspots, and developing predictive models to aid in urban crime prevention and resource allocation.

# **Proposed Methodology**

The methodology for this project focuses on analyzing Toronto’s crime patterns using machine learning and geospatial techniques to address critical research questions related to temporal patterns, spatial hotspots, and predictive modeling of criminal activity. Given the extensive structure of the datasets, several data preprocessing, visualization, and modeling steps are planned to extract meaningful insights.

**Data Preprocessing**:  
Data cleaning and transformation are essential steps to prepare the data for analysis. Steps include:

* + **Handling Missing Data**: Missing values, especially in location-specific fields (latitude, longitude, Occurrence, and neighborhood identifiers marked as NSA), were examined. Rows with unassigned coordinates or neighborhood data were removed.
  + **Date-Time Conversion**: Dates were converted to Python datetime format, allowing for the extraction of temporal features such as year, month, day of the week, and hour of the day.
  + **Data Type Conversion**: Numeric and categorical variables were converted to appropriate data types (e.g., integers for years) to optimize computational efficiency.

Combination of following techniques will be employed to answer questions laid out earlier in this report:

1. **Geospatial Analysis**:  
   Geospatial techniques will identify spatial patterns and hotspots across Toronto’s neighborhoods.
   * **Heatmaps and Density Plots**: Using libraries like Folium and Geopandas, interactive maps will illustrate areas with higher crime density. This visualization will help in understanding the concentration of specific crime types in different neighborhoods and over time.
   * **Hotspot Analysis**: Clustering techniques, such as DBSCAN and K-means, will be applied to locate crime hotspots. These methods will provide a quantitative assessment of clusters and highlight areas that may benefit from focused policing efforts.
2. **Temporal Pattern Analysis**:  
   Analyzing time-related variables helps uncover seasonal or weekly trends in crime.
   * **Trend Analysis**: Temporal features will be analyzed to find peaks in crime occurrences, helping to answer questions related to seasonal variations or high-risk periods (e.g., certain days of the week or times of day).
   * **Time-Series Decomposition**: Time-series techniques will break down trends, seasonality, and noise in the data, identifying temporal dependencies within crime occurrences.
3. **Predictive Modeling**:  
   Machine learning algorithms will be implemented to classify crime types and predict future occurrences.
   * **Classification Models**: Algorithms such as Decision Trees, Logistic Regression, or Random Forests will be trained to classify crime types based on attributes like location type, premises type, time, and day of the week.
   * **Evaluation Metrics**: Accuracy, precision, recall, and F1-score will evaluate model performance, ensuring that the prediction accuracy aligns with project goals.
   * **Predictive Hotspot Mapping**: Using predictive models, crime-prone areas will be visualized to identify which neighborhoods and premises types are likely to face increased crime activity.
4. **Tools and Libraries**:
   * **Python Libraries**: Key libraries include Pandas for data manipulation, Matplotlib and Seaborn for visualization, Scikit-learn for machine learning, and Folium for interactive mapping.
   * **GitHub**: The project’s code and results will be documented in a GitHub repository for transparency and reproducibility. <https://github.com/RiskSpecialist/CIND820_TorontoCrime/blob/main/CIND820_TorontoCrime.ipynb>

Through these methodological steps, the project aims to answer critical questions about Toronto’s crime patterns, enabling data-driven strategies for proactive crime prevention and resource allocation in urban policing.

# **Data Source**

Toronto Public Safety Data Portal: <https://data.torontopolice.on.ca/datasets/0a239a5563a344a3bbf8452504ed8d68_0/about>

Toronto Police Services - Major Crime Indicators (MCI) Open Data – (excluding Homicide and Sexual Violations)

<https://data.torontopolice.on.ca/datasets/0a239a5563a344a3bbf8452504ed8d68_0/explore?location=43.733276%2C-79.369718%2C9.78&showTable=true>

Toronto Police Services - Major Crime Indicators (MCI) Open Data – (Homicide)

<https://data.torontopolice.on.ca/datasets/TorontoPS::homicides-open-data-asr-rc-tbl-002/explore?location=43.722485%2C-79.393300%2C9.87&showTable=true>

# **GitHub Repository**

<https://github.com/RiskSpecialist/CIND820_TorontoCrime>

Python Code in GitHub: <https://github.com/RiskSpecialist/CIND820_TorontoCrime/blob/main/CIND820_TorontoCrime.ipynb>

# **References**

Crime Severity Index (CSI) in Canada. (2021). *Statistics Canada*. Retrieved from <https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=3510002601>

He, Z., Lai, R., Wang, Z., Liu, H., & Deng, M. (2022). Comparative study of approaches for detecting crime hotspots with considering concentration and shape characteristics. *International Journal of Environmental Research and Public Health, 19*(21), 14350. Retrieved from <https://www.mdpi.com/1660-4601/19/21/14350>

Ioannidis, I. (2024). *Understanding crime patterns using spatial data analysis: Case studies in Stockholm, Sweden* (Doctoral dissertation, KTH Royal Institute of Technology). KTH Royal Institute of Technology. Retrieved from

<https://kth.diva-portal.org/smash/get/diva2:1902051/FULLTEXT02.pdf>

Lau, T. (2020, April 1). *Predictive policing explained*. Brennan Center for Justice. Retrieved from <https://www.brennancenter.org/our-work/research-reports/predictive-policing-explained>

Luo, X. (2012). *Spatial patterns of neighbourhood crime in Canadian cities: The influence of neighbourhood and city contexts* (Master’s thesis, University of Waterloo). Retrieved from <https://dspacemainprd01.lib.uwaterloo.ca/server/api/core/bitstreams/46e84cb5-bcbf-4645-9482-9c06bd6e2cf3/content>

Mohammadi, A., Bergquist, R., Fathi, G., Pishgar, E., Nogueira de Melo, S., Sharifi, A., & Kiani, B. (2022). Homicide rates are spatially associated with built environment and socio-economic factors: A study in the neighborhoods of Toronto, Canada. *BMC Public Health, 22*, 1482. Retrieved from <https://doi.org/10.1186/s12889-022-13807-4>

Toronto Police Service. (2023). *Annual Statistical Report 2023.* Retrieved from <https://data.torontopolice.on.ca/documents/17ebb34ad86744a2a58803f3563768ef/about>

Wang, L., Lee, G., & Williams, I. (2019). The spatial and social patterning of property and violent crime in Toronto neighborhoods: A spatial-quantitative approach. *ISPRS International Journal of Geo-Information*, *8*(1), 51. Retrieved from <https://www.mdpi.com/2220-9964/8/1/51>