



Crime Rate Analysis in Toronto, **Canada**

Supervisor: Prof. Hanan Swidan

Data Analytics for Business

Zekelman School of Information Technology

Group 3

Yen Nga Le

Srilakshmi Gummadidala

Tehsin Shaikh

Vinod Soloman Santhakumar

ABSTRACT

Crime, defined as the breach of established laws by individuals or groups, is a pervasive social issue with profound consequences for individuals, families, and communities. Addressing this challenge necessitates a holistic approach encompassing prevention, intervention, and victim support. This research delves into crime prevention strategies, aiming to comprehend crime patterns for the development of effective strategies in ensuring public safety, mitigating victimization, and enhancing community well-being.

The chosen topic gains significance in the global context due to the widespread impact of crime. Notably, incidents involving humans becoming targets of violent crimes underscore the urgency of a thorough analysis of crime trends. This research endeavors to contribute valuable insights by employing a comprehensive machine learning model. By integrating diverse data sources, including crime reports, demographic information, and environmental factors, the study seeks to provide a nuanced understanding of crime patterns and trends.

ACKNOWLEDGEMENT

We **express** our sincere gratitude to St Clair College for affording us the opportunity to undertake our capstone project, allowing us to showcase the skills and knowledge acquired throughout our course.

We would like to extend heartfelt thanks to my esteemed academic advisors, particularly Professor Hanan, whose unwavering guidance and support have been invaluable throughout the research process. Professor Hanan's expertise has illuminated the path of this study, and their insights have significantly contributed to the robustness of the research methodology.

TABLE OF CONTENTS

Introduction	4
Problem Statement	5
Dataset	5
Methodology.....	9
Data Analysis	10
Conclusion	19
Future Work	20
References.....	21
Appendix.....	21

INTRODUCTION

Crime is an unlawful act committed by an individual or a group that violates established laws and regulations within a given jurisdiction. It is a significant social issue with wide-ranging impacts on individuals, families, and communities.

Addressing crime requires a comprehensive approach that focuses on prevention, intervention, and support for victims. Therefore, understanding the nature of crime in communities is essential for law enforcement agencies and policymakers to develop effective crime prevention strategies. *(Government of Canada, 2023)*

The reason for selecting this topic is the pervasive nature of crime as a social issue that has profound effects on individuals and communities globally. By delving into crime prevention strategies and understanding crime patterns, we aim to tackle the issues concerning public safety, victimization, and community well-being.

As we see many cases in the news over the past years highlighting that international students are being targeted for fatal shootings or for violent crimes for unknown reasons and found dead without knowing the cause. These factors realize that it is important for a comprehensive analysis of crime trends involving different types of crime incidents. It would be essential to gather information from reliable sources and provide the appropriate valuable insights for the investigation agencies. *(The Times of India, 2023)*

The main objectives of the project are to conduct a thorough crime analysis in Toronto, identify crime factors and high-crime areas, assess the effectiveness of crime prevention measures, and provide actionable recommendations. Through these efforts, the project aims to contribute to the reduction of major crimes, enhance public safety, and create a safer community for all residents of Toronto.

Analysis Questions:

1. Have certain crime categories increased or decreased over time?
2. Are there specific areas or neighborhoods within Toronto that experience higher rates of major crimes?

3. How do the characteristics of each crime category (e.g., location, time of day) vary across different areas of the city?
4. Are there any seasonal patterns in major crime rates?
5. What are the top 5 offenses in each of the top 5 divisions with the highest overall offense counts, and how do they compare to each other in terms of frequency?
6. Based on latitude and longitude data, are there any areas of Toronto that consistently experience high levels of multiple major crimes?
7. Are there any significant occurrences of crime incidents over the Covid – 19 time? Whether the crime rate increased or decreased during this period?
8. What is the overall trend of major crimes in Toronto over the past several years?

PROBLEM STATEMENT

Develop a machine learning model for identifying and analyzing crime patterns that considers a wide range of data sources, including crime reports, demographic information, and environmental factors, to provide a more accurate and comprehensive understanding of crime patterns and trends.

This project aims to provide insights and actionable recommendations for law enforcement agencies, policymakers, and community organizations to reduce crime rates, improve community safety, and allocate resources more efficiently. In this way they can understand the root cause of crime and develop proactive strategies for preventing and reducing crime, particularly in areas that are most affected by crime. (*data.torontopolice.on.ca, n.d.*)

DATASET

1. Dataset source:

The dataset is provided by the Toronto Police Service and can be accessed through the Toronto Public Safety Portal. The portal aims to promote transparency and public awareness regarding crime in the city.

2. Data collection and funding:

Police open data includes any data collected or maintained by the Toronto Police Service unless certain data or data in its entirety is exempt for legal, privacy, security, and confidentiality or commercially sensitive reasons. The Toronto Police Service considers privacy and data quality to be of utmost importance. The Toronto Police Service is committed to the proactive provision of police open data while taking necessary measures to protect privacy, legal and confidential data.

Toronto Police Service provides open analytics to aid in visualizing and understanding police information. These interactive visualizations provide trend analysis and important information at a glance. (*data.torontopolice.on.ca, n.d.*)

3. Participants or Cases:

Datasets include millions of individual crime incidents, each representing a separate case. However, due to privacy concerns, certain details such as the victim's name and address are not included in the dataset.

4. Variables of Interest:

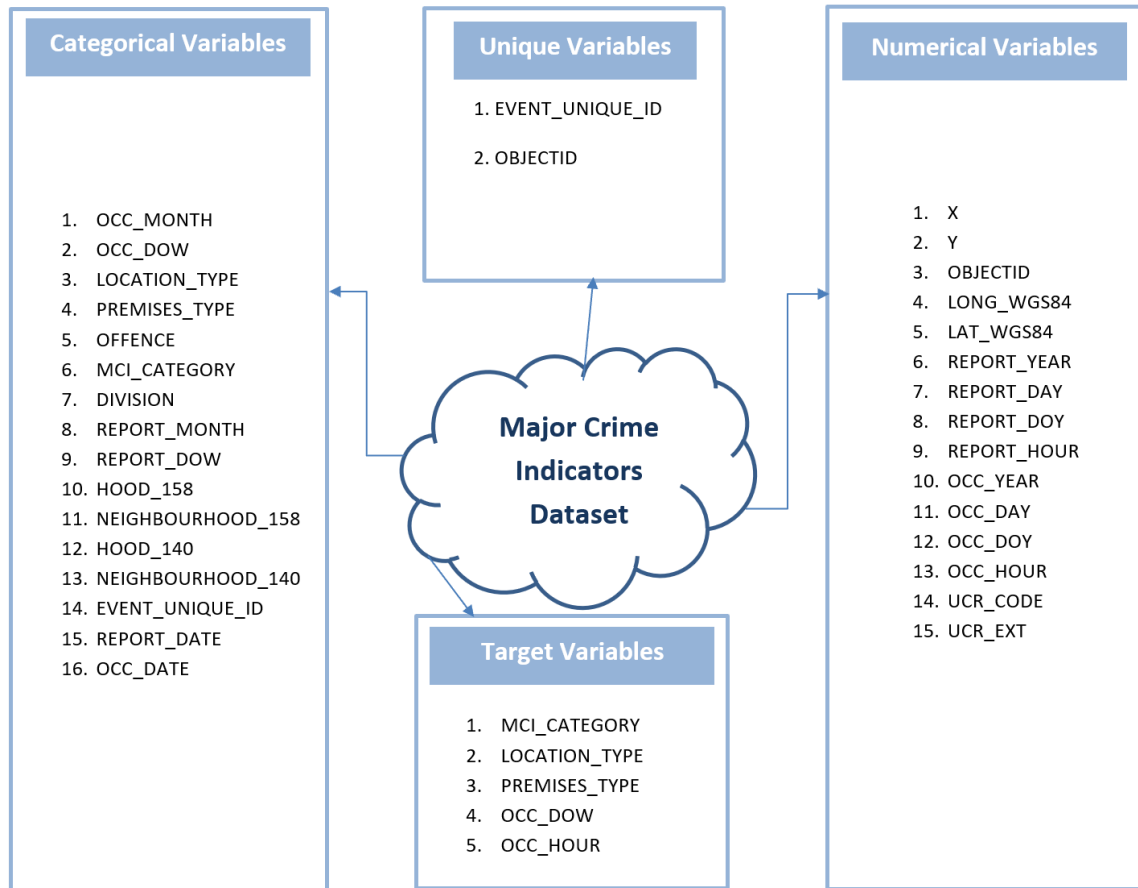
The dataset includes various variables that provide information about each crime incident such as date and related offences, categories include Assault, Break and Enter, Auto Theft, Robbery and Theft Over, Premises type, location, neighborhood, latitude, and longitude, etc.

5. Data Field Descriptions:

Field	Field Name	Description
1	EVENT_UNIQUE_ID	Offence Number
2	REPORT_DATE	Date Offence was Reported (time is displayed in UTC format when downloaded as a CSV)
3	OCC_DATE	Date Offence Occurred (time is displayed in UTC format when downloaded as a CSV)
4	REPORT_YEAR	Year Offence was Reported
5	REPORT_MONTH	Month Offence was Reported
6	REPORT_DAY	Day of the Month Offence was Reported
7	REPORT_DOY	Day of the Year Offence was Reported
8	REPORT_DOW	Day of the Week Offence was Reported
9	REPORT_HOUR	Hour Offence was Reported
10	OCC_YEAR	Year Offence Occurred

11	OCC_MONTH	Month Offence Occurred
12	OCC_DAY	Day of the Month Offence Occurred
13	OCC_DOY	Day of the Year Offence Occurred
14	OCC_DOW	Day of the Week Offence Occurred
15	OCC_HOUR	Hour Offence Occurred
16	DIVISION	Police Division where Offence Occurred
17	LOCATION_TYPE	Location Type of Offence
18	PREMISES_TYPE	Premises Type of Offence
19	UCR_CODE	UCR Code for Offence
20	UCR_EXT	UCR Extension for Offence
21	OFFENCE	Title of Offence
22	MCI_CATEGORY	MCI Category of Occurrence
23	HOOD_158	Identifier of Neighbourhood using City of Toronto's new 158 neighbourhood structure
24	NEIGHBOURHOOD_158	Name of Neighbourhood using City of Toronto's new 158 neighbourhood structure
25	HOOD_140	Identifier of Neighbourhood using City of Toronto's old 140 neighbourhood structure
26	NEIGHBOURHOOD_140	Name of Neighbourhood using City of Toronto's old 140 neighbourhood structure
27	LONG_WGS84	Longitude Coordinates (Offset to nearest intersection)
28	LAT_WGS84	Latitude Coordinates (Offset to nearest intersection)

6. Database Schema:



Here X and Y are the same actual longitude and Latitude values.

7. Dataset size:

The dataset includes data from 2014 to 2022 with categories include Assault, Break and Enter, Auto Theft, Robbery and Theft Over (Excludes Sexual Violations).

The dataset contains overall 360,076 cases with 29 variables.

8. Included locations and graphing capabilities:

This dataset provides the information about Toronto city, Ontario, Canada., major crime records based on the below geographic information.

We can make the related graphs based on the city and neighborhood divisions, latitude, and longitude areas.

Geography Information:

- ❖ West longitude: -79.89358291913355
- ❖ East longitude: -78.93736375220695
- ❖ North latitude: 43.90544544158975
- ❖ South latitude: 43.46884134594142

9. Dataset Limitations:

Reporting Bias: The dataset relies on reported crimes to the Toronto Police. However, not all crimes may be reported, resulting in potential underrepresentation or discrepancies in the actual crime rates.

Contextual Factors: The dataset may not capture important contextual factors that can influence crime rates, such as socio-economic conditions, demographic changes, or specific events.

Data Interpretation: The dataset provides information about crime incidents but may not provide sufficient context or details to fully understand the circumstances or underlying causes of the crimes.

Population Changes: The dataset does not account for changes in population size or demographics over time.

METHODOLOGY

Data Preprocessing using Python Libraries:

Python libraries played a crucial role in refining our raw datasets for analysis. Leveraging Pandas, we ensured data quality by addressing duplicates, missing values, and inconsistencies in crime reports, demographic information, and environmental datasets. These preprocessing steps were pivotal in preparing a clean and standardized dataset, laying the foundation for our subsequent machine learning model.

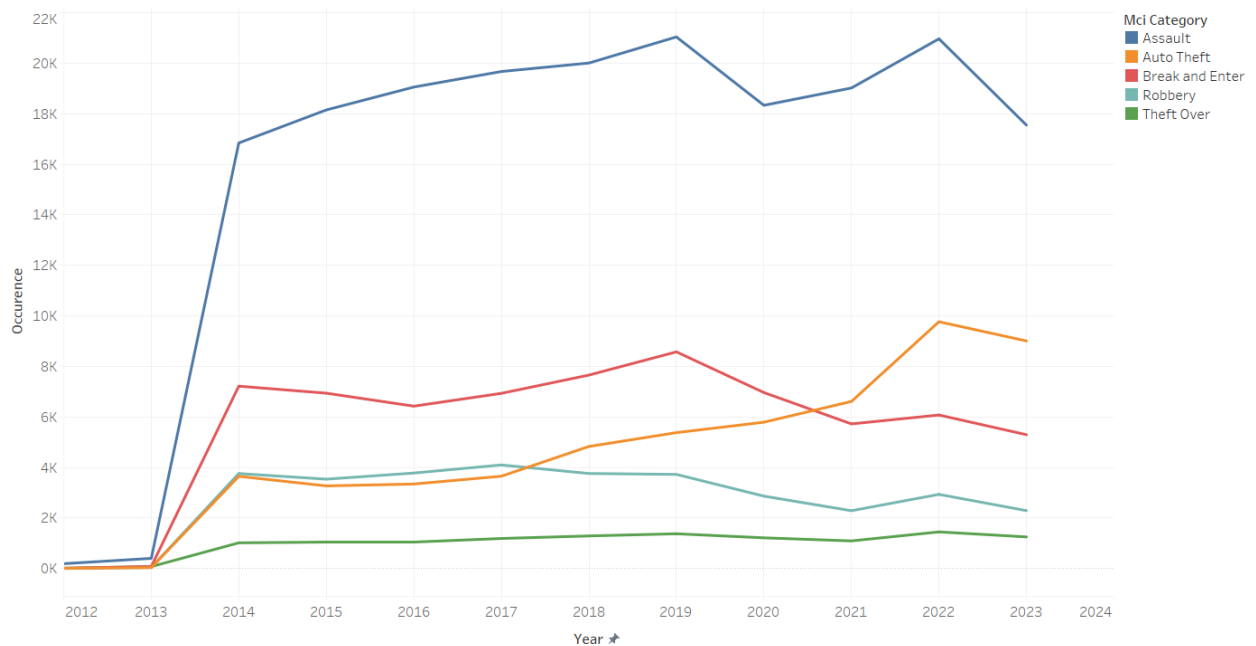
Data Visualization with Tableau:

Tableau emerged as a robust tool for creating insightful visualizations to aid in the interpretation and communication of findings. Using Tableau, we crafted interactive visualizations showcasing crime trends over time and space. Demographic insights and the exploration of environmental factors were seamlessly portrayed, providing a comprehensive understanding of the contextual elements influencing criminal activity. This approach not only streamlined the analytical process but also ensured effective communication of insights through visually engaging narratives.

DATA ANALYSIS

1. Have certain crime categories increased or decreased over time?

Overall Trend of Crimes over the Years



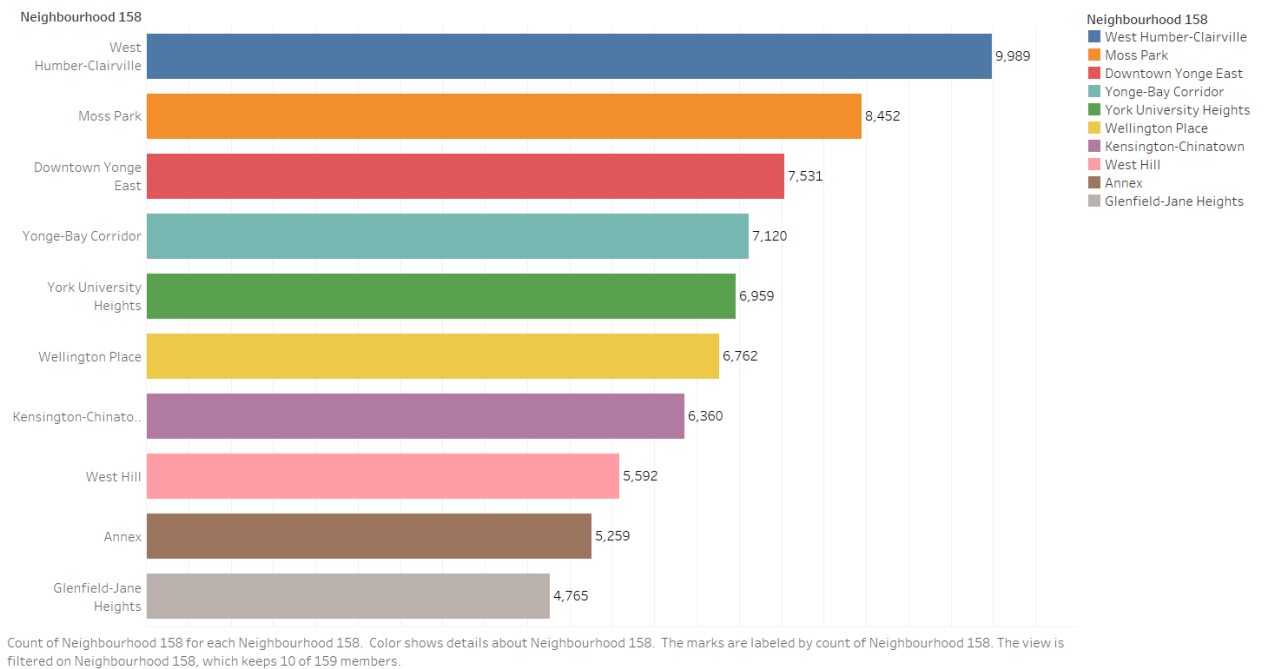
The trend of count of Major_Crime_Indicators_Open_Data.csv for Occ Year. Color shows details about Mci Category. The view is filtered on Occ Year, which keeps non-Null values only.

Description:

- According to the graph, Assault crime category is the popular crime category among all other categories.
- However, the Theft Over crime type has the lowest crime rate over the years.

2. Are there specific areas or neighborhoods within Toronto that experience higher rates of major crimes?

Top 10 Neighbourhood areas with highest crime rate

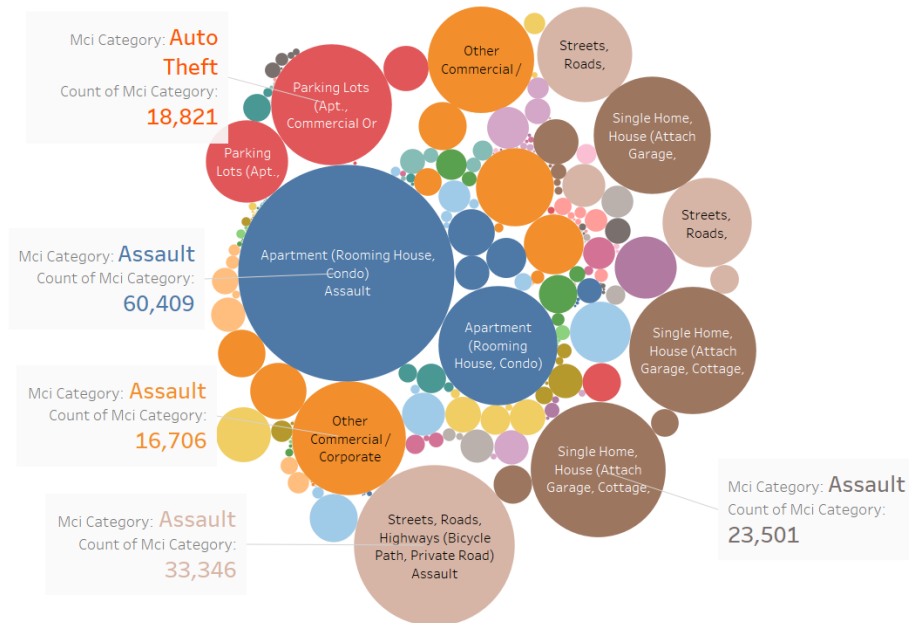


Description:

- The bar chart displays the top 10 neighborhoods within Toronto with the highest rates of crime.
- West Humber-Clairville has the highest crime rate followed by Moss Park among these neighborhoods, while Annex has the lowest crime rate.

3. Weekly Count of the Crime Occurrences by Premises.

Crime Rate based on Premises

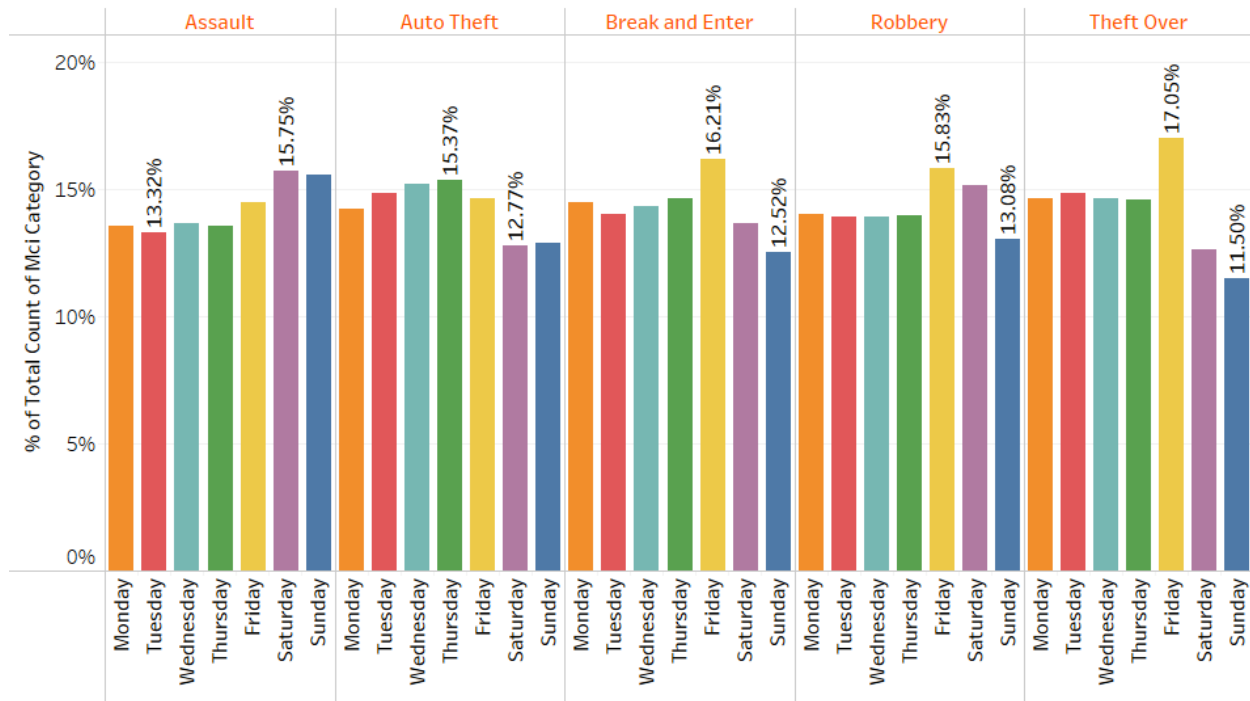


Location Type and MCI Category. Color shows details about Location Type. Size shows count of MCI Category. The marks are labeled by Location Type and MCI Category. The view is filtered on Location Type and MCI Category. The Location Type filter excludes Unknown. The MCI Category filter excludes Null.

Description:

- In this tree map, we can observe that the majority of the reported crime incidents have taken place in the Apartments, followed by occurrences in Streets, Roads, Highways.
- Particularly during the time frame of Fridays to Sundays.

4. Are there any seasonal patterns in major crime rates?



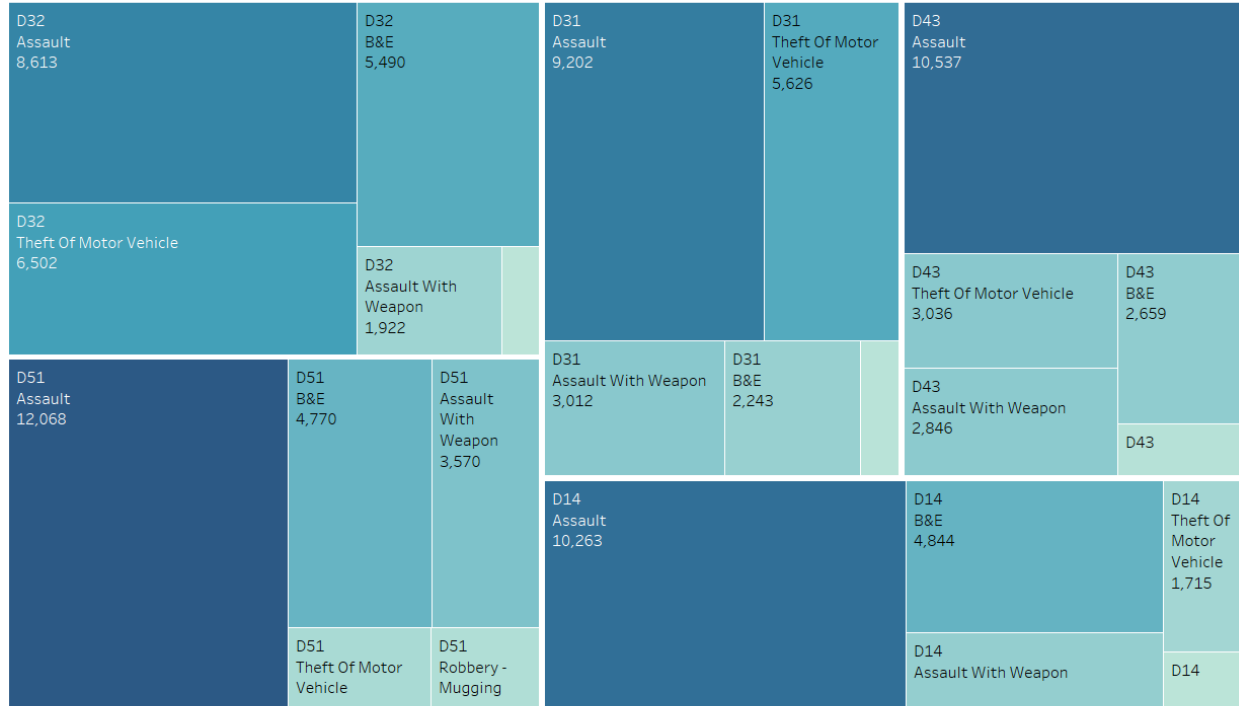
% of Total Count of MCI Category for each Occ Dow broken down by MCI Category. Color shows details about Occ Dow. The data is filtered on Occ Month and Occ Year. The Occ Month filter excludes Null. The Occ Year filter ranges from 2000 to 2023. The view is filtered on MCI Category, which excludes Null.

Description:

- As per the bar graph we are showcasing the seasonal patterns in terms of Weekdays of Crime rates.
- As shown Particularly during the time frame of Fridays to Sundays are the peak time of Crime.

5. What are the top 5 offenses in each of the top 5 divisions with the highest overall offense counts, and how do they compare to each other in terms of frequency?

Top 5 Offences by Division



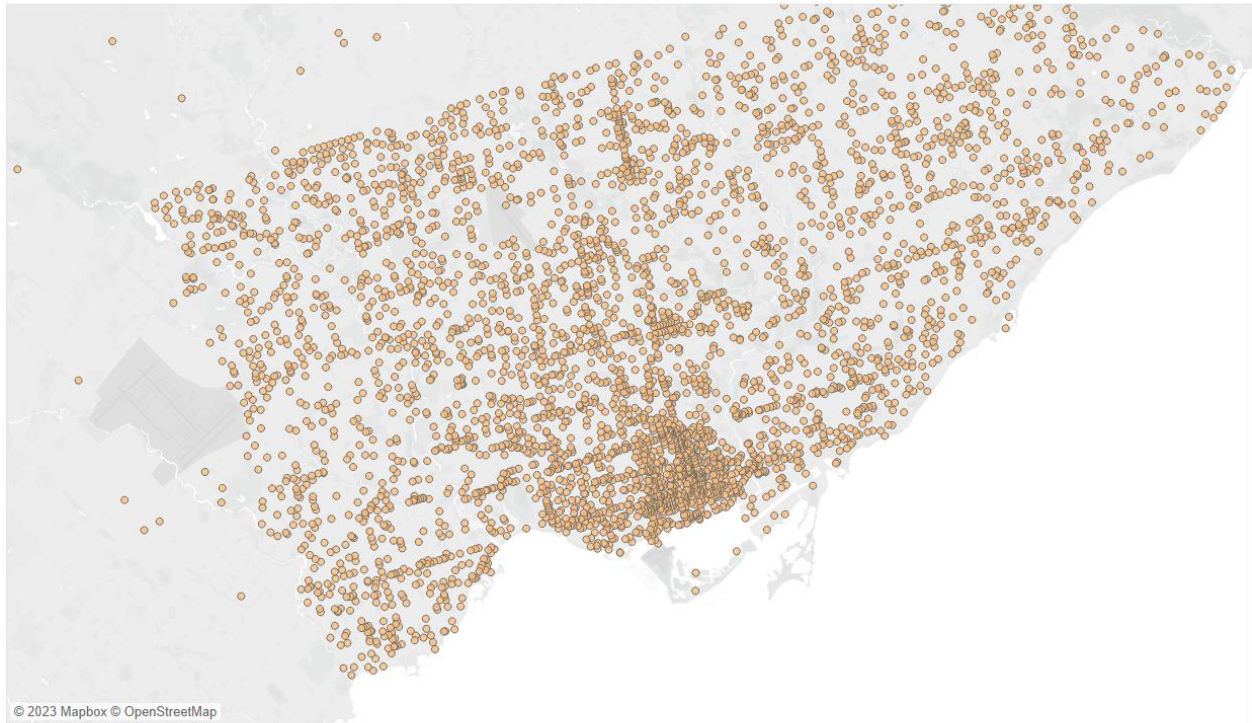
Division, Offence and count of Event Unique Id. Color shows count of Offence. Size shows count of Event Unique Id. The marks are labeled by Division, Offence and count of Event Unique Id. The view is filtered on Offence and Division. The Offence filter has multiple members selected. The Division filter has multiple members selected.

Description:

- As per the bubble chart Police Division D51 has the highest number of offense count incidents in the Assault crime category, closely followed by D43 Division.
- Assault being popular in the top 5 offence and division category.

6. Based on latitude and longitude data, are there any areas of Toronto that consistently experience high levels of multiple major crimes?

Offense Rate by Area - Theft Over



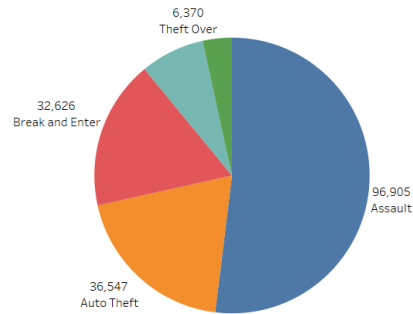
Map based on Long Wgs84 and Lat Wgs84. Color shows details about Offence. The view is filtered on Lat Wgs84 and Offence. The Lat Wgs84 filter includes values greater than or equal to 1.000. The Offence filter has multiple members selected.

Description:

- The neighborhood of West Humber Clair exhibits the highest incidence of theft, comprising over 37.3% of the total reported crimes. Within this category, 56.5% (507 cases) are classified as commercial theft, 21.8% (492 cases) occur in residential areas, and 37.3% (482 cases) are attributed to other locations.

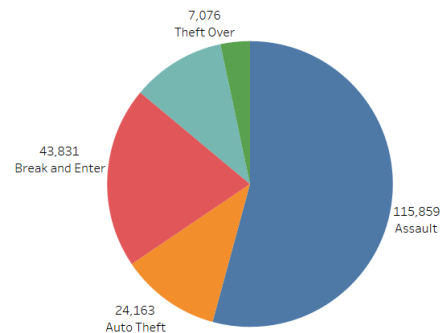
**7. Are there any significant occurrences of crime incidents over the Covid – 19 time?
Whether the crime rate increased or decreased during this period?**

Crime Count After Covid- 19



Count of Event Unique Id and Mci Category. Color shows details about Mci Category. Size shows count of Event Unique Id. The marks are labeled by count of Event Unique Id and Mci Category. The data is filtered on Occ Year, which ranges from 2019 to 2023.

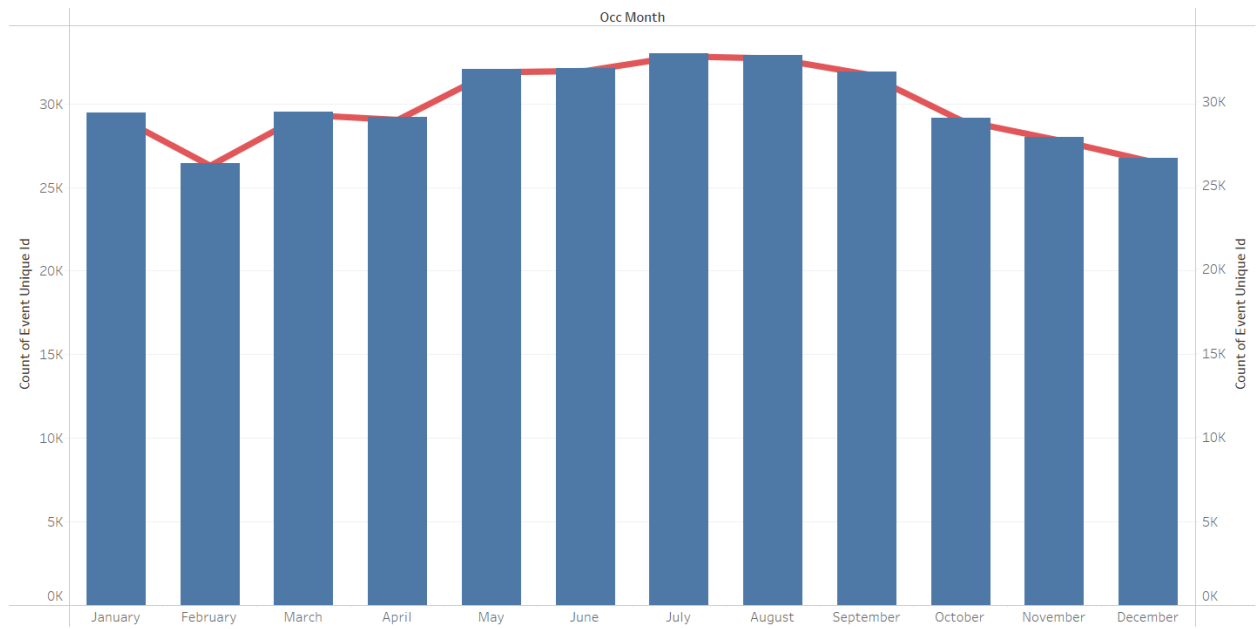
Crime Count Before Covid- 19



Count of Event Unique Id and Mci Category. Color shows details about Mci Category. Size shows count of Event Unique Id. The marks are labeled by count of Event Unique Id and Mci Category. The data is filtered on Occ Year, which ranges from 2000 to 2019.

- As per the above maps based on the different crime categories, there is little difference in the crime rate before and after the Covid – 19 time.
- According to the observation of each crime category, Assault has been highest before and after Covid – 19 followed by Auto Theft, Break & Enter and Theft Over.

8. What is the overall trend of major crimes in Toronto over the months?



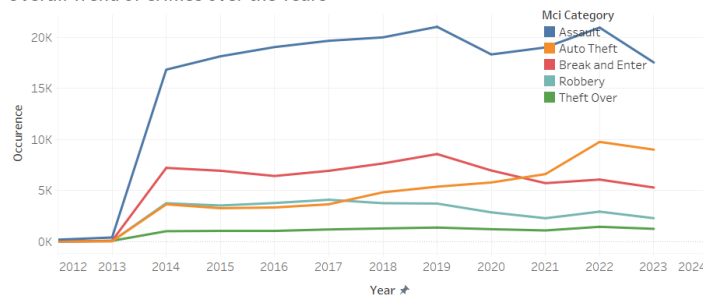
The trends of count of Event Unique Id and count of Event Unique Id for Occ Month. The data is filtered on Occ Year, which ranges from 2014 to 2023.

Description:

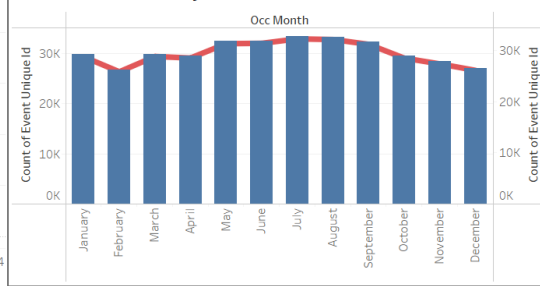
- By observing this graph, we can conclude that the trend of crime gradually increased between May to September and decreased from October onwards.

Dashboard #1:

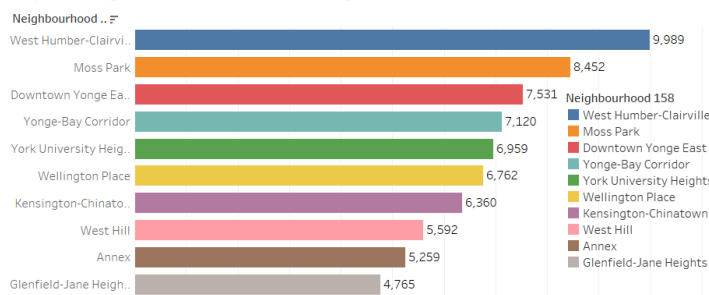
Overall Trend of Crimes over the Years



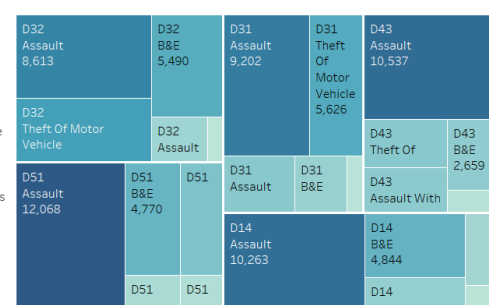
Overall Trend of Major Crimes in Toronto Over Months



Top 10 Neighbourhood areas with highest crime rate



Top 5 Offences by Division

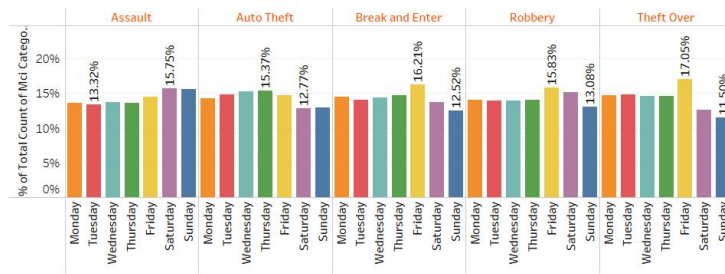


Description:

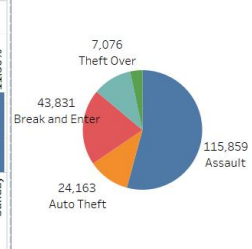
- Within this dashboard, we present a comprehensive overview of the crime rate. We analyze the overall trend of major crimes over the years, examining the incidents that have occurred within specific MCI categories, including Assault, Break & Enter, Auto Theft, Robbery, and Theft Over.
- In addition, we have assessed the top 10 neighborhood areas with high crime rates. Through this analysis, we identify the areas that have experienced a significant level of criminal activity.
- As per the treemap, Police Division D51 has the highest number of offenses count incidents in the Assault crime category, closely followed by D43 Division. Assault being popular in the top 5 offence and division category.
- By observing this graph, we can conclude that the trend of crime gradually increased between May to September and decreased from October onwards.

Dashboard #2:

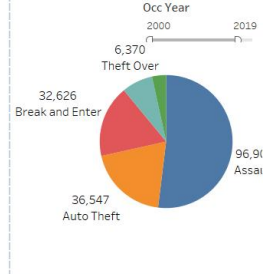
Major Crime Rate by Day of Week



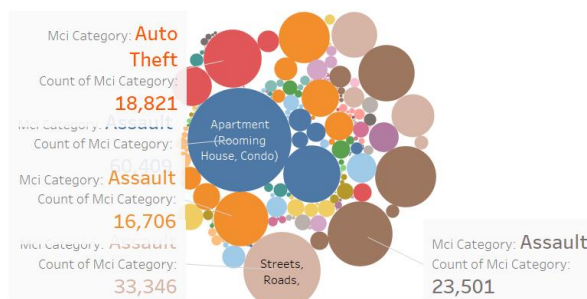
Crime Count Before Covid-19



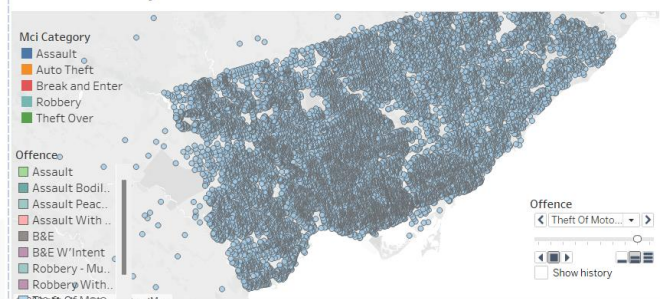
Crime Count After Covid-19



Crime Rate based on Premises



Offense Rate by Area - Theft Of Motor Vehicle



Description:

- In this second dashboard, we have determined the weekly crime count occurrences, revealing that weekends, particularly starting from Friday, exhibit a high crime rate and the primary target for crime incidents is apartments.
- Lastly, we have identified crime rate insights before and after the Covid-19 situation, specifically examining the popularity of each MCI category.

Conclusions

Firstly, we observed a constant increase in the overall number of crimes in Toronto over the past few years, with a slight decrease during Covid – 19 times.

The top five most common crime categories remained consistent over the years, with Assault, Break and Enter, Auto Theft, Robbery, and Theft Over being the most prevalent.

Neighborhoods, including West Humber-Clairville, Mose Park, Downtown Yonge East areas are popular among others in highest major crimes.

Seasonal pattern in major crime rates, with higher crime counts starting from summer months until December. However, lower crime counts can be seen in the first quarter of the year.

Assault and Break and Enter are the major concerns in all the top 5 divisions with high overall offense counts. Auto Theft stands out as a significant issue in D32 and D31, while Robbery and Theft Over incidents show a relatively consistent frequency across all divisions.

Lack of available demographic information, such as age, gender, and socioeconomic status, is a limitation of the dataset.

Future Work

➤ **Combining Demographic Data with Major Crime Indicator (MCI) Data**

- Combining demographic data of Toronto with Major Crime Indicator (MCI) data for Toronto analyze the factors influencing crime patterns.
- Identify and list socio-economic variables in demographics and crime-related variables in MCI, emphasizing common identifiers like neighborhoods for merging.
- Perform cleaning, feature engineering and exploratory analysis to capture socio-economic and crime-related information in the integrated dataset.

➤ **Machine Learning Algorithm Development**

- Develop a machine learning algorithm that is suitable for predicting major crime indicators and influencing factors based on the integrated dataset.
- Train machine learning algorithm, optimize its parameters, and validate its performance using cross-validation techniques.

References

1. data.torontopolice.on.ca. (n.d.). *data.torontopolice.on.ca*. Retrieved from data.torontopolice.on.ca: <https://data.torontopolice.on.ca/datasets/TorontoPS::major-crime-indicators-open-data/explore>
2. Government of Canada. (2023, Dec 15). *Government of Canada*. Retrieved from Government of Canada: https://www150.statcan.gc.ca/n1/en/subjects/crime_and_justice/crimes_and_offences
3. The Times of India. (2023, May 14). *The Times of India*. Retrieved from The Times of India: <https://timesofindia.indiatimes.com/nri/us-canada-news/gujarat-student-found-dead-in-toronto/articleshow/100220123.cms?from=mdr>

Appendix

- Python notebook file with EDA Analysis.

