Kimya Khakzad Shahandashti
219391267
kimya@yorku.ca

Zahra Hosseini
219867142
zhsn@yorku.ca

1 Introduction

In this report, we present the results of our analysis of Toronto robberies dataset using Python and Tableau. Our analysis involved applying various techniques to identify patterns in crime rates, including time series analysis, spatial analysis.

In order to evaluate the success of our dataset and analysis, we will assess the accuracy of the data, relevance of findings, robustness of methods, and reproducibility of results. Accuracy of data is the first step in this evaluation process, which involves identifying missing values, outliers, and other inconsistencies and making any necessary corrections. The relevance of our findings will be assessed to ensure that they are meaningful and useful for decision-making purposes. The robustness of our methods will be evaluated by testing the stability of results when using different analytical techniques or subsets of data. Finally, we will assess the reproducibility of our results to ensure that they can be replicated using different data sources or analytical methods. By evaluating our dataset and analysis according to these criteria, we can ensure that our findings are reliable and useful for future research and decision-making.

The analysis of robberies data is crucial in understanding robberies trends and identifying areas that require attention from law enforcement agencies. By identifying patterns in robberies rates, law enforcement agencies can allocate resources more effectively and implement policies that address specific robberies trends.

Our analysis involved the use of Python and Tableau, which are powerful tools for data analysis and visualization. With these tools, we were able to analyze large datasets, identify patterns, and visualize the results in a way that is easy to understand.

The report is structured as follows. First, we provide a brief overview of the data used in our analysis. We then describe the techniques we used to analyze the data, including time series analysis, spatial analysis, and seasonality analysis. We present the results of our analysis and evaluate their relevance and robustness. Finally, we conclude the report with a summary of our findings and recommendations for law enforcement agencies based on our analysis.

2 Dataset

The data set used in this analysis is the Major Crime Indicators (MCI) occurrences dataset provided by the Toronto Police Service. The dataset includes information on all MCI occurrences by reported date and related offenses. The MCI categories include Assault, Break and Enter, Auto Theft, Robbery, and Theft Over.

The dataset contains 27 fields, including unique identifiers for each occurrence and its corresponding offense, as well as information on the division where the offense occurred, the date of the offense, the location and premises type, the UCR (Uniform Crime Reporting) code and extension, the offense title, the reported year, month, day, day of the year and week, and hour, as well as the occurrence year, month, day, day of the year and week, and hour.

Additionally, the dataset includes information on the MCI category of occurrence, the identifier of the neighborhood, and the name of the neighborhood, as well as the longitude and latitude coordinates of the location of the offense.

It is important to note that the data provided is at the offense and/or victim level, and therefore, one occurrence number may have several records associated with it, related to the various MCIs used to categorize the occurrence. However, occurrences that have been deemed unfounded are not included in the dataset.

The success of this dataset was evaluated based on the accuracy of the data, the relevance of the findings, the robustness of methods, and the reproducibility of the results, as detailed in the introduction section of this report. The evaluation involved reviewing the data for missing values, outliers, and other inconsistencies and making any necessary corrections, as well as assessing the stability of the results when different methods are used or when

different subsets of the data are analyzed.

Table 1: Description of Fields in Toronto robberies Dataset[1]

Field	Field Name	Description
1	Index	Unique Identifier
2	event_unique_id	Offence Number
3	Division	Police Division where Offence Occurred
4	occurrence_date	Date of Offence
5	reporteddate	Date Offence was Reported
6	location_type	Location Type of Offence
7	premises_type	Premises Type of Offence
8	ucr_code	UCR Code for Offence
9	ucr_ext	UCR Extension for Offence
10	Offence	Title of Offence
11	reportedyear	Year Offence was Reported
12	reportedmonth	Month Offence was Reported
13	reportedday	Day of the Month Offence was Reported
14	reporteddayofyear	Day of the Year Offence was Reported
15	reporteddayofweek	Day of the Week Offence was Reported
16	reportedhour	Hour Offence was Reported
17	occurrenceyear	Year Offence Occurred
18	occurrencemonth	Month Offence Occurred
19	occurrenceday	Day of the Month Offence Occurred
20	occurrencedayofyear	Day of the Year Offence Occurred
21	occurrencedayofweek	Day of the Week Offence Occurred
22	occurrencehour	Hour Offence Occurred
23	MCI	MCI Category of Occurrence
24	Hood_ID	Identifier of Neighbourhood
25	Neighbourhood	Name of Neighbourhood
26	Long	Longitude Coordinates (Offset to nearest intersection)
27	Lat	Latitude Coordinates (Offset to nearest intersection)

3 Motivation

In this section, we will investigate the patterns and trends of robbery incidents in Toronto by answering four key questions.

3.1 What is the trend of robbery incidents over time?

Firstly, we will examine the trend of robbery incidents over time to identify any significant changes in the number of incidents and possibly correlate them with events or changes in the city's demographic or economic landscape.
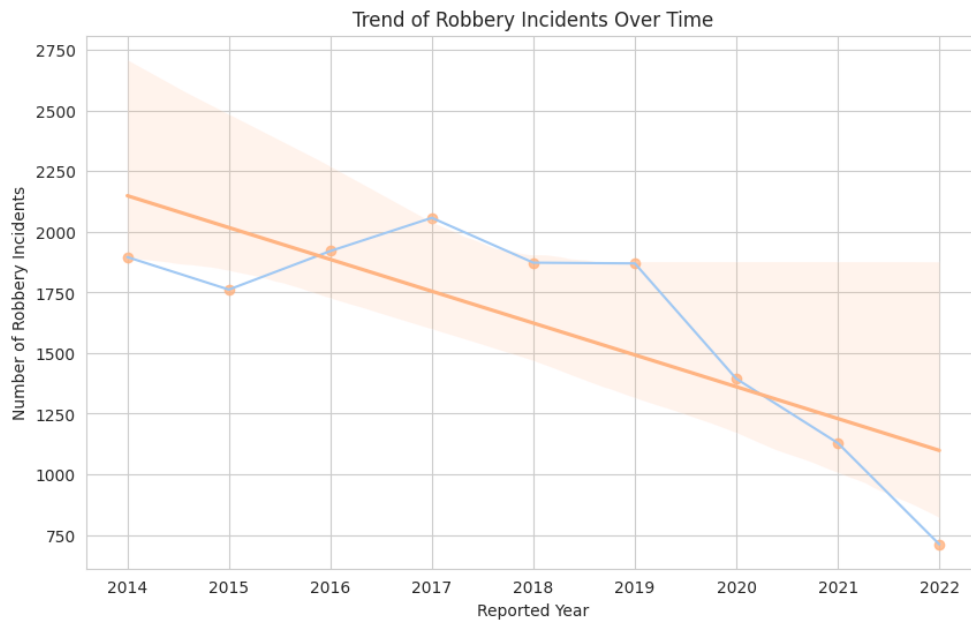
Figure 1: Histogram showing the trend of robbery incidents in Toronto from 2014 to 2022-06-30. The x-axis represents the years, and the y-axis represents the number of reported robbery incidents.

3.2 Are there any seasonality patterns in the occurrence of robbery incidents?

Secondly, we will explore whether there are any seasonality patterns in the occurrence of robbery incidents, which can provide valuable insights for law enforcement agencies to allocate resources effectively and focus on specific times of the year when the likelihood of robbery incidents is higher.
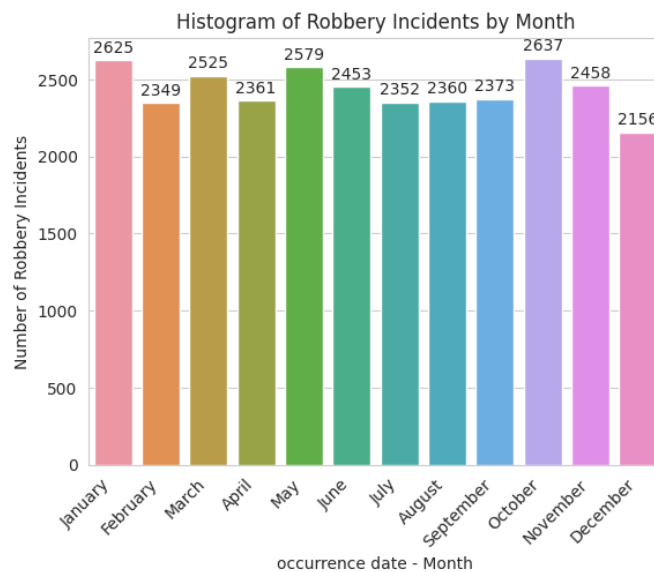

Figure 2: Histogram showing the distribution of robbery incidents by month in Toronto. The x-axis represents the months of the year, and the y-axis represents the number of reported robbery incidents. The figure illustrates whether there are any seasonality patterns in the occurrence of robbery incidents, indicating any months with a higher likelihood of robbery incidents.

3.3 Which areas in Toronto are most affected by robberies?

Thirdly, we will look at which areas in Toronto are most affected by robberies. By identifying the hotspots where extra attention is needed to deter crime, we can help law enforcement agencies and local governments allocate their resources effectively and ensure public safety.
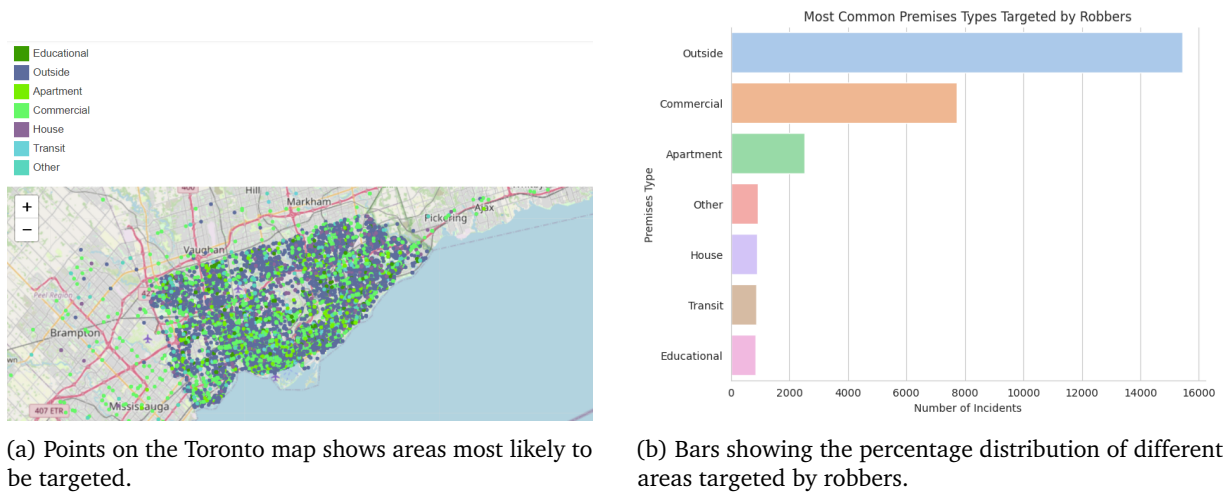
(a) Points on the Toronto map shows areas most likely to be targeted.

(b) Bars showing the percentage distribution of different areas targeted by robbers.

Figure 3: Comparison of areas targeted by robbers in Toronto.

3.4 What types of premises are most likely to be targeted by robbers?

Finally, we will analyze the types of premises that are most likely to be targeted by robbers. This information can help businesses and individuals take preventative measures to reduce the risk of becoming a victim of robbery.

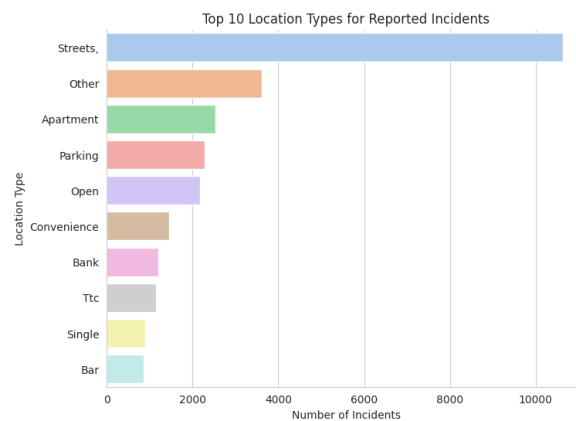(a) Points on the Toronto map shows the types of location most likely to be targeted.

Figure 4: Comparison of areas targeted by robbers in Toronto.

By answering these four questions, we can gain a comprehensive understanding of the patterns and trends of robbery incidents in Toronto. This will enable us to develop more effective strategies to prevent crime, improve public safety, and help ensure that Toronto remains a safe and secure place for all its citizens.

By analyzing this dataset, we can answer several important questions including:

1. What is the trend of robbery incidents over time?

2. Are there any seasonality patterns in the occurrence of robbery incidents?

3. Which areas in Toronto are most affected by robberies?

4. What types of premises are most likely to be targeted by robbers?

The analysis of the Toronto Robberies dataset can provide valuable information for law enforcement agencies and policymakers. Understanding the trend and patterns of robbery incidents over time can identify high-risk areas and times for robberies, and law enforcement resources can be deployed accordingly. The analysis can also provide insights into the demographics of the victims and perpetrators, which can inform the development of targeted interventions. Furthermore, the insights gained from the analysis can be used to inform the development of evidence-based crime prevention policies and strategies, with the ultimate goal of reducing the incidence of crime and improving community safety.

4 Methodology

The Toronto Robberies dataset offers rich information that can be analyzed using various statistical data analytics methods such as time-series analysis and data visualization. Other types of data analysis that can also be performed are spatial analysis and descriptive statistics. These methods can uncover patterns and trends in the data, identify high-risk areas, examine the demographics of victims and perpetrators, and inform the development of evidence-based crime prevention policies. The steps for performing data analytics on the Toronto Robberies dataset can be summarized as follows:

1. Data Cleaning: Handle missing values, remove duplicates, and correct any errors in the data.

2. Data Exploration: Perform exploratory data analysis to gain an understanding of the data and its distribution.

3. Data Visualization: Create visualizations such as histograms, scatter plots, and heat maps to explore relationships and patterns in the data.

4. Data Transformation: Transform the data into a format that is suitable for analysis, such as aggregating the data by year or location.

5. Time Series Analysis: Analyze time-series data to identify trends and patterns, such as the trend in the number of robberies over time.

6. Data Interpretation: Interpret the results of the data analysis and draw insights and conclusions about the data.

Technical problems that can arise in the Toronto Robberies dataset are missing values and outliers. These can lead to incorrect results during analysis and can be handled by removing rows, filling in missing values, using imputation techniques, transforming data, using robust methods, or identifying outliers through visualization techniques.

5 Evaluation

The success of this dataset will be evaluated based on the following criteria:

- Accuracy of data: The first step in evaluating the success of this analysis will be to assess the accuracy of the data. This will involve reviewing the data for missing values, outliers, and other inconsistencies and making any necessary corrections [2]. One of the metrics for evaluations is checking the accuracy of data, for preprocessing our data, we did find and remove missing values, inconsistencies and everything.

  For a step further, We used k-means clustering algorithm to find out if there is any outlier in our data, as we can see there is just 117 datapoint int other cluster and all of the other datapoint about 30 thousands points are in another cluster.
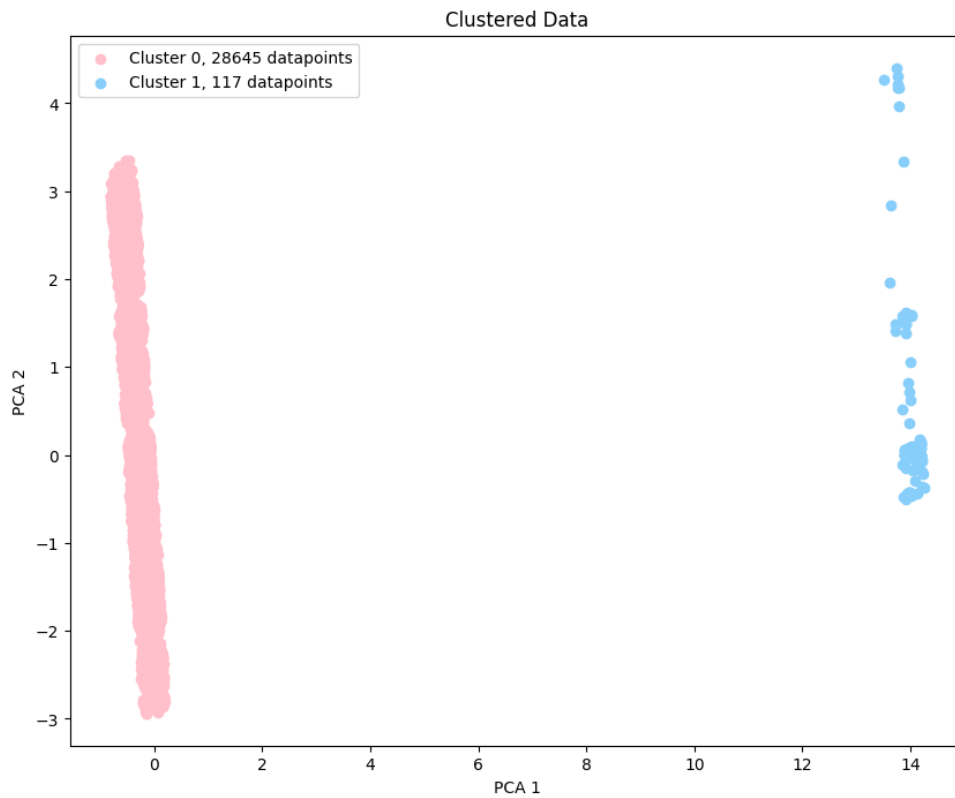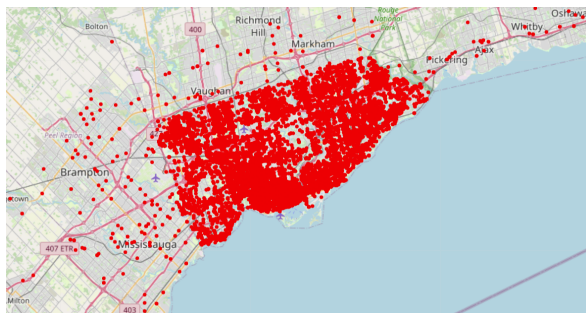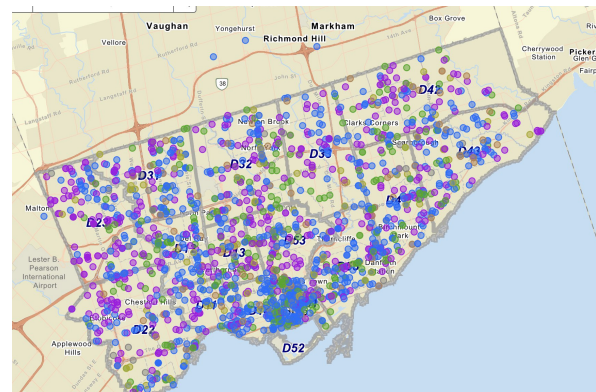
Figure 5

- Relevance of findings: The next step in evaluating the success of this analysis will be to assess the relevance of the findings. We can see here that division 51 has more crimes(overall- I mean any kind of crime) by the toronto crim app. Also our result shows that D51 has most robberies which we can realize using two different dataset there is same output.

(a) Points on the Toronto map shows the types of location most likely to be targeted.

(b) Green circles on the map represent incidents of Break and Enter, while purple circles indicate incidents of Auto Theft.

Figure 6: Comparison of areas targeted by robbers in Toronto.

- Robustness of methods: This will involve evaluating the stability of the results when different methods are used or when different subsets of the data are analyzed.

- Reproducibility: This will involve evaluating the extent to which the results can be reproduced using different data sources or using different analytical methods.

In summary, the success of this analysis will be evaluated based on the accuracy of the data, the relevance of the findings, the robustness of the methods, the practical utility of the insights gained, and the reproducibility of the results.

6 Conclusion and Future work

Robbery-Income Relationship: We can combine this dataset with the dataset containing the average income of different neighborhoods of Toronto to find the trend of crime rate in relation to income.
Comparative Analysis: Future work can include a comparative analysis of robbery incidents across different cities in Canada or globally.
Community Policing: Include an analysis of the impact of community policing initiatives, such as security cameras, on the frequency and severity of robbery incidents in Toronto.

References

[1] Toronto Robberies 2014 to 2022-06-30. https://www.kaggle.com/datasets/hayawi/toronto-robbery-2014-to-20220630.

[2] AGRESTI, A., AND FINLAY, B. *Statistical methods for the social sciences*, 4th ed. Prentice Hall, Upper Saddle River, NJ, 2009.