Toronto Police Service

Major Crime Indicators

1 August, 2023

Jianan Sun: 214831226
Ranajay Senapati: 219851864
Shilpi Dixit: 219352301
Daniel Alex: 219938471

Toronto Police Service
Analytics & Innovation

The work we submitted as part of this assignment is original, and due credit is given to others where appropriate. We accept and acknowledge that that each member of our team is equally responsible for if the assignment found to be plagiarized in any way, and we will be subject to school's Academic Integrity policy

Agenda

- Executive Summary
- Introduction
- Analysis
- Modelling
- Conclusions
- References
- Appendix

Toronto Police Service
Analytics & Innovation

Executive Summary

Goal

Predict probability of crimes per Neighbourhood based on date and time of day

Solution

Predicts incident_score, which indicates the probability of crime happening given a particular neighbourhood, day of year, day of week and time of day.

Derived Features

- incident_score: Crime rate happening in a neighbourhood as per day of the year, day of week and time of day.
- recency: Prioritises recent crimes occurred in a neighbourhood as more likely to repeat

External Sources

Used Toronto Neighbourhood Profile Open Data to find average individual income, population, unemployment rate and low income population for every neighbourhood

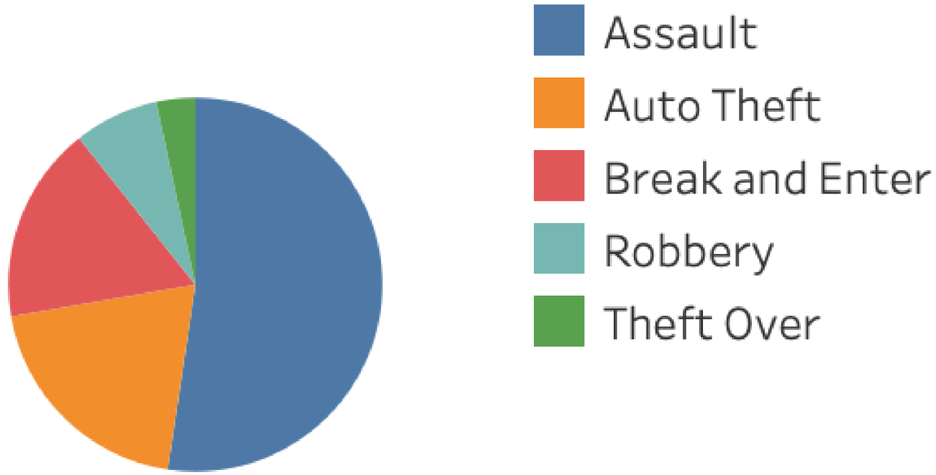
Method

CatBoost: Gradient boosting algorithm for categorical features.

Conclusions

Build a Dashboard with model and and private TPS data for police deployment

Crime Breakdown

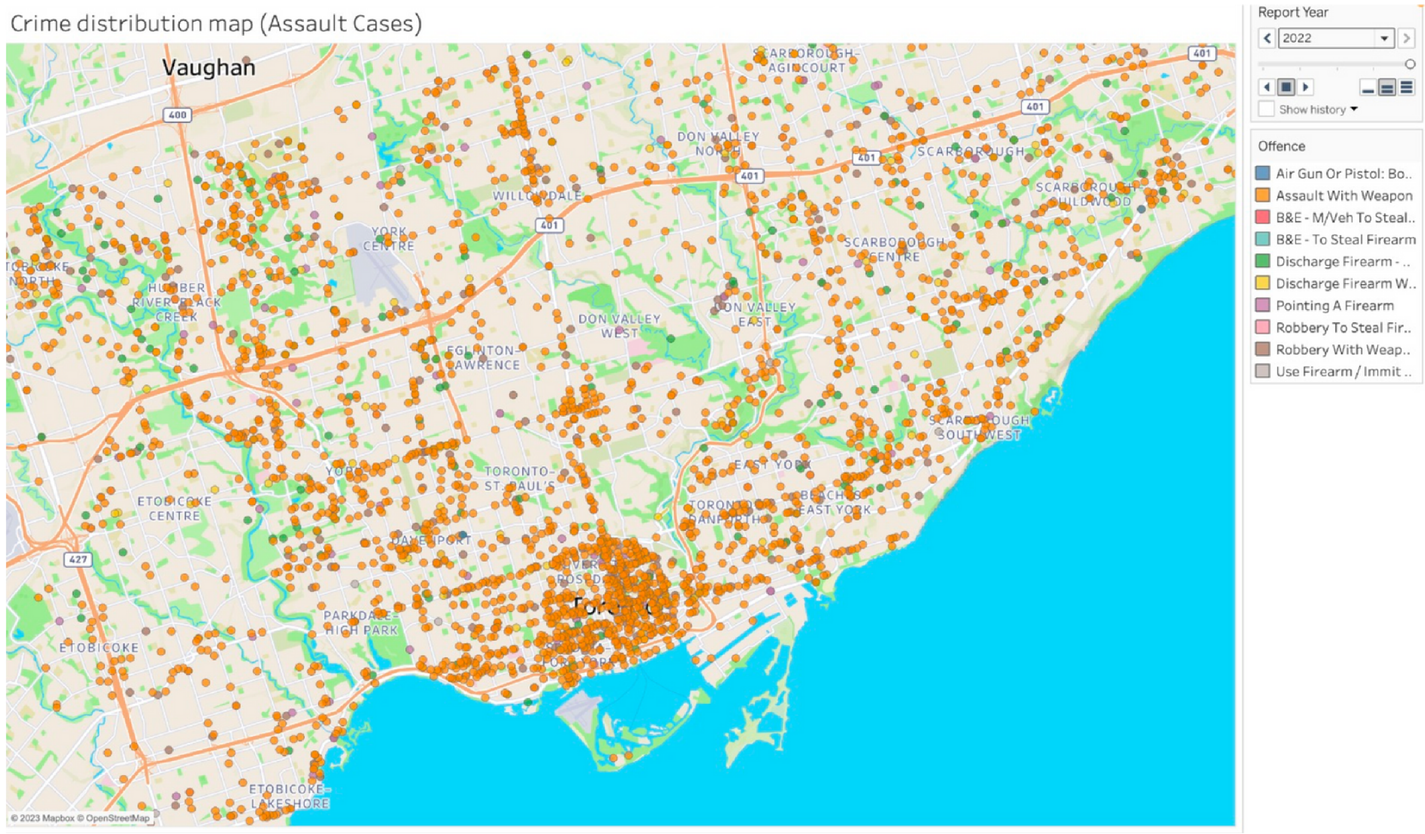


Toronto Police Service
Analytics & Innovation

Introduction

We decided to primarily focus on assault cases since it has the most amount of crimes as shown in the Executive Summary. Our aim is to:

- Pinpoint neighbourhoods most likely to experience an assault related threat given significant and urgent trends.
- Discover predictive patterns in potential hotbeds for assault related crimes.
- Consider significant determinants and trends that influence crime location, date and time of day.



Toronto Police Service
Analytics & Innovation

Introduction (contd..)

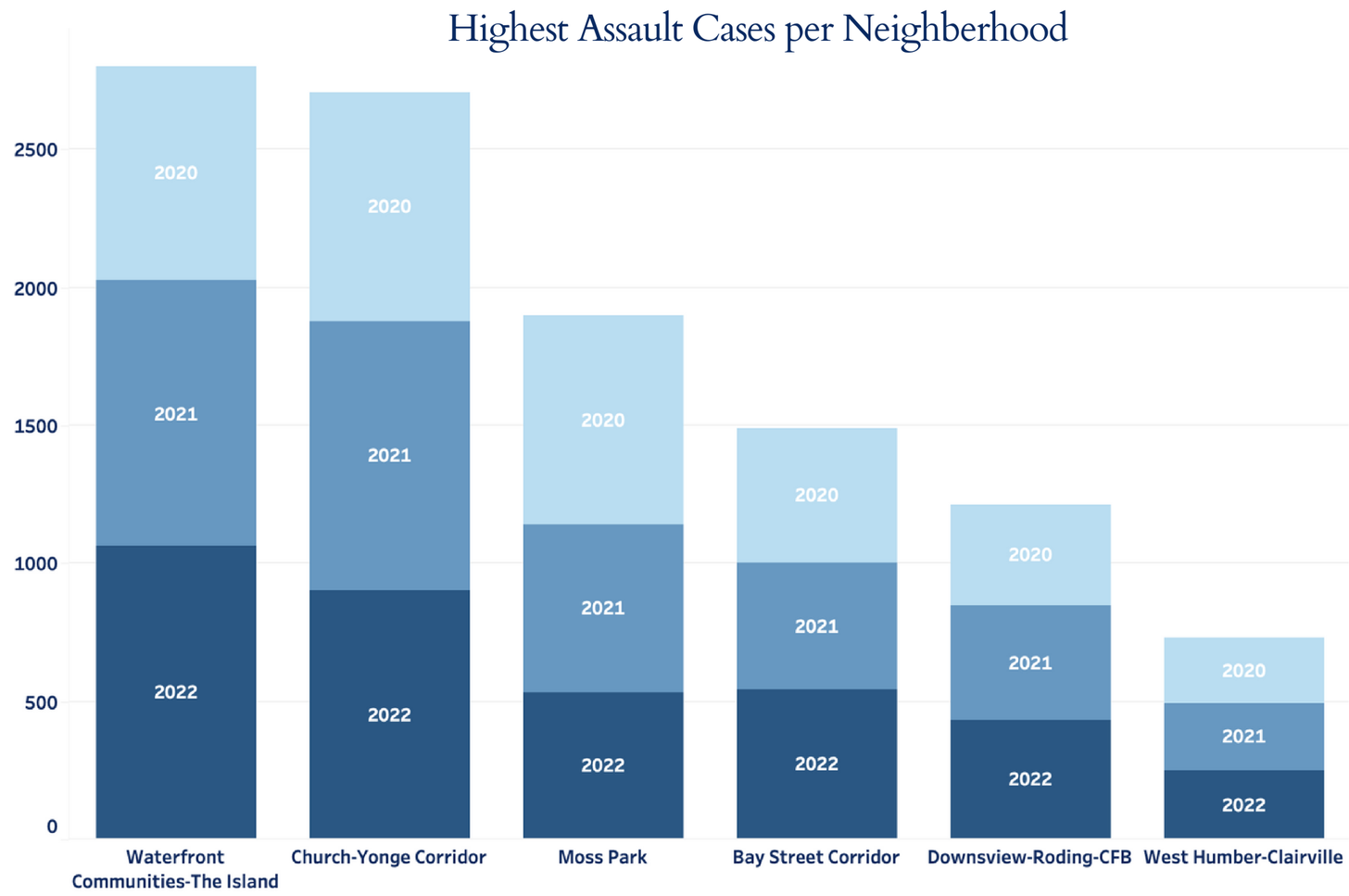
Utilises open publicly available datasets

- Major Crime Indicators TPS Crime Statistics
- Toronto Neighbourhood Profiles

Objective:

Our intention is the overall reduction of assault related crime in Toronto in the near future, and to assist the TPS in their goal to serve and protect its communities.

Given the time and resource constraints of the TPS, we want to provide a scaleable AI solution that identifies neighbourhoods most likely to experience an assault related threat at a certain time of the day.



Toronto Police Service
Analytics & Innovation

Basic Analysis (Descriptive Analytics)

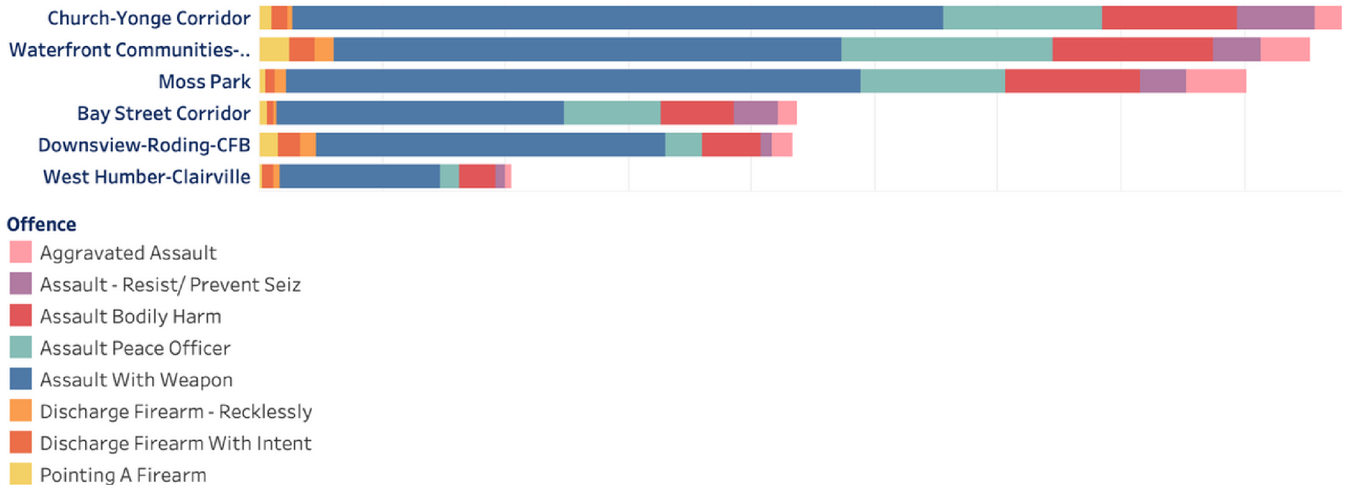
Using 3 years of data (2020-2022), we decided to categorize time of crime occurrence hours into the below:

- Day: 05:00 AM to 05:00 PM
- Night: 05:01 PM to 11:00 PM
- Midnight: 11:01 PM to 04:59 AM

Neighbourhood 140	Day	Night	Midnight
Waterfront Communities-The Island	1,179	889	731
Church-Yonge Corridor	1,426	852	430
Moss Park	1,021	560	319
Bay Street Corridor	894	399	197
Kensington-Chinatown	613	381	223
Downsview-Roding-CFB	593	401	220
York University Heights	507	293	205
West Humber-Clairville	382	233	118
Islington-City Centre West	360	228	117
Woburn	532	350	173

Most Crimes done during daytime hours and least at Midnight.

Assault Types per Neighbourhood



We can observe that:

- The most assault related crimes occur in Waterfront Communities, Church-Yonge, Moss Park, Bay Street.
- Most common crime: Assault with Weapons.

Toronto Police Service
Analytics & Innovation

Modelling

Models Used:

CatBoost Regressor

- Ensemble method designed to efficiently handle categorical variables.
- Combination of gradient boosting and decision trees.
- Incorporates regularization to prevent overfitting in One Hot Encoding.
- Uses Ordered Boosting to handle categorical variables rather than One Hot Encoding.
- Ordered Boosting assigns categorical variables scores based on target variable.
- We used a CatBoost Regressor since our dependent variables are both numerical and categorical.

Time Series Analysis

- Analyze crimes for every neighbourhood across the years (per 1000 population)

Model Analysis

Notable Input Features:

- We have created a derived feature 'incident_score' where we take into calculate crime rate happening in a neighbourhood as per day of the year, day of week and time of day. This is useful in taking into account the pattern of crime rates happening across past years.
- We have introduced a new feature for 'recency'. It indicates more recent a crime has happened in a neighbourhood higher the chance of happening it sooner. This feature helps in giving more importance to the neighbourhood which has growing crime rates in the recent past.
- We have also tried to capture socio-economic factors by including average individual income, population, unemployment rate and low income population for every neighbourhood

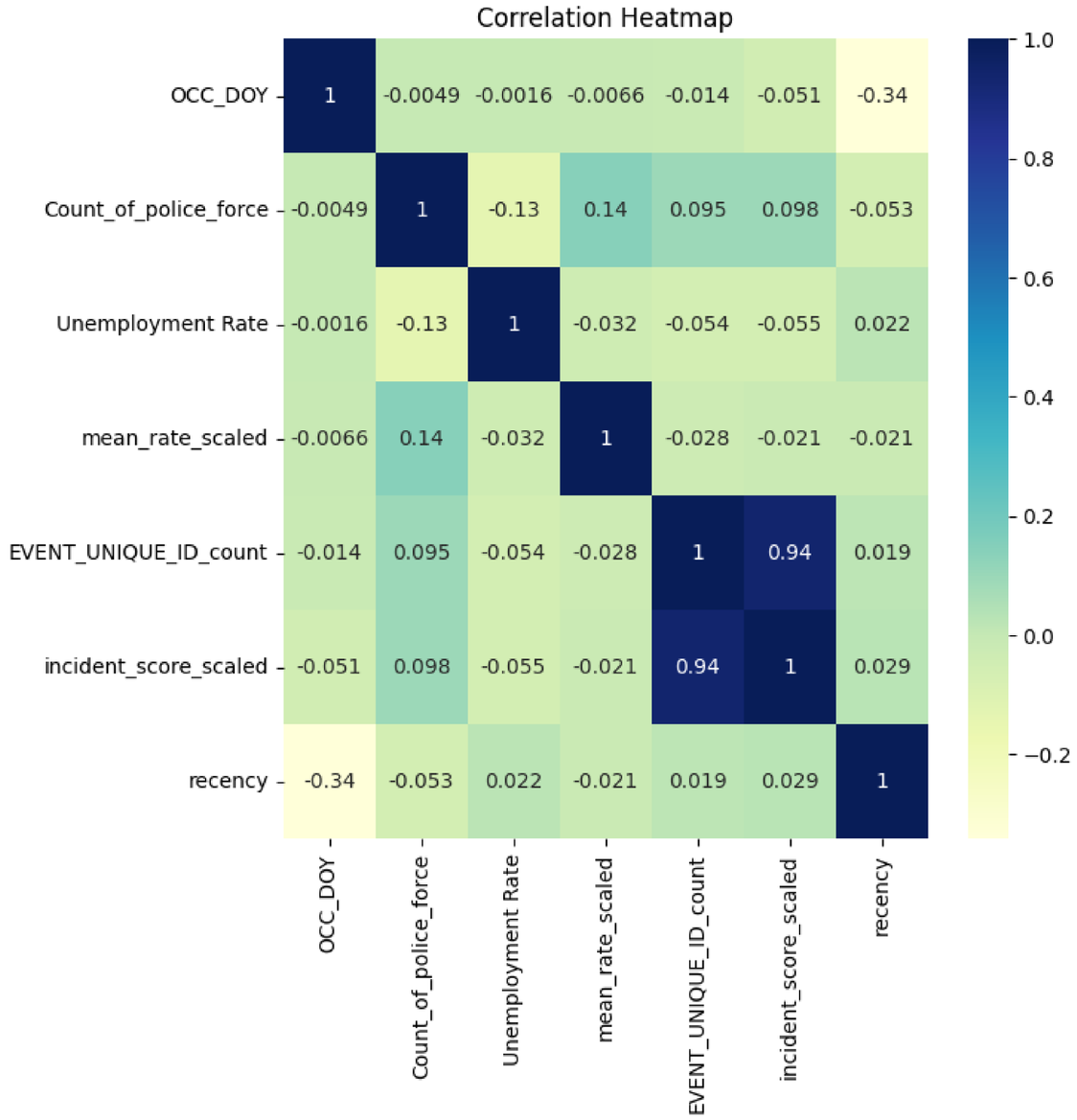
Final Outcome: The final CatBoost model built predicts the incident_score, which indicates the probability of crime happening given a particular neighbourhood, day of year, day of week and time of day.

Model Analysis: Feature Importance

Feature Name	Feature Importance as per catBoost Regressor
EVENT_UNIQUE_ID_count (count of incidents per neighbourhood, day of year)	94.37
recency (Recency Score)	4.42
time_of_day	0.44
mean_rate_scaled (Crime rate per population)	0.35
Unemployment Rate	0.15
Count_of_police_force	0.12
OCC_DOW (Occurrence day of week)	0.10
Low Income population(LIM-AT)	0.05

Toronto Police Service
Analytics & Innovation

Model Analysis: Feature Importance Correlation Heatmap



Examines Correlation between independent variables

Toronto Police Service
Analytics & Innovation

Modelling Results: Metrics

R-Squared for training dataset: 0.85
R-Squared for testing dataset: 0.84
MSE: 0.0012

Mean Square Error is calculated by taking the average of the squared differences between the predicted and actual values. A low MSE generally mean that the model is better at predicting the target variable.

The model's MSE is desirable since it indicates that the model is better at predicting the target variable.

R-squared determines the proportion of variance in the target variable explained by the model.

The model's R-squared is desirable since it indicates that the model is a good fit for the data and explains a large percentage of variation in the target variable.

The model also doesn't suffer from overfitting due to both training and test data having almost similar R2 score.

Model Limitations

General Model Crime Volatility

Due to the unpredictability of crime, and the specific nature of the offence, or circumstances leading to the crime, it wouldn't be possible to pinpoint exact location or premises for the next potential offence.

Creating a model that predicts a specific time of the day and neighbourhood, as well as a specific location and premises within the neighbourhood would be both infeasible and unrealistic

Limited Data

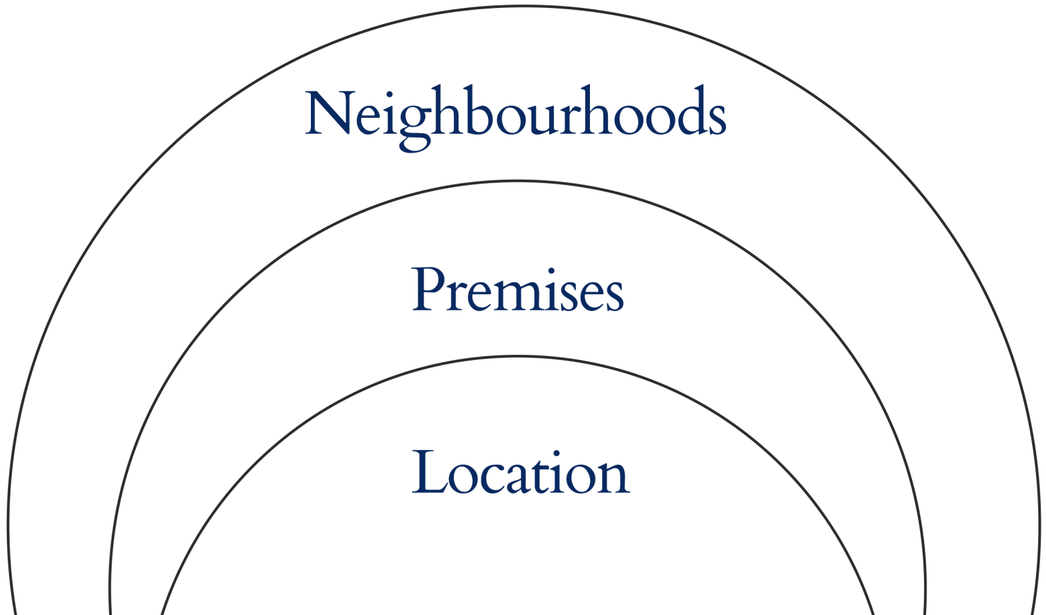
The existing dataset provides trends and inference variables for Assault based crimes, but Limited availability of data that are crime predictors for a neighbourhood (homelessness, unemployment etc)

Black-box Model

The CatBoost algorithm, like other ensemble models, can be hard to interpret. Specifically, the calculations on which the model forms its predictions. This can make it challenging to identify the reasoning behind model decisions.

External Factors Domain-specific

Our model is incapable of considering systemic and socio-economic factors influencing crime, so domain knowledge, career experience, personal judgement and empathy by the TPS is of vital importance



Toronto Police Service
Analytics & Innovation

Conclusions and next steps...

Dashboard built in Model

Using the Final Model and integrating the predictions provided by real-time crime data into a dashboard for easy interpretability and better allocation of resources.

Pipeline Real time data

Build pipeline from recorded data to Model to Dashboard for quick, accurate decision making for TPS

Merge Data

Utilise both publicly available recorded crime statistics with private data of police force deployment to be able to adapt to crime trends within hoods and districts

Thank you

Toronto Police Services and Schulich School of Business

Service at our Core

Do the right thing

Connect with Compassion

Reflect and Grow

References

Toronto Police Public Safety Data Portal: Major Crime Indicators Dataset

- https://data.torontopolice.on.ca/pages/major-crime-indicators

Toronto Open Data: Neighbourhood Profiles Dataset

- https://open.toronto.ca/dataset/neighbourhood-profiles/

Toronto Police Services Website

- https://www.tps.ca/mission-vision-values/

Appendix (code)

1. Kindly refer to attached Part-1 python notebook file for calculation of crime rate using time series.
2. Kindly refer to attached Part-2 python notebook for final model building using CatBoost regressor.