# Analytics Startup Plan

**Synopsis: *This document provides a high-level walkthrough of the activities required to guide completion of the analysis.***

|  |  |
| --- | --- |
| **Project** | *Policing responce Optimization* |
| **Requestor** | *Toronto Police Service* |
| **Date of Request** | *July 12th, 2022* |
| **Target Quarter for Delivery** | *August 12th, 2022* |
| **Epic Link(s)** | *N/A* |
| **Business Impact** | *Improving the police response in identifying the possible crime type* |

## 1.0 Business Opportunity Brief

|  |  |
| --- | --- |
|  | Clearly articulated business statement of the Ask, opportunity, or problem you are trying to solve for. An important step is to understand the nature of the business, system or process and the desired problems to be addressed. This will be communicated back to All stakeholders for alignment. |

There is a situation that a 911 dispatcher can be unsure of the crime type when the caller unable to describe what is happening, and a police officer cannot clarify the crime type when they hear a call for help. To optimize the police response, they need to know the crime tendency in the city and most probable crime type corresponding to corelated features, such as location, time, and premise information.

**The specific ask:**

**I am going to develop the crime type prediction system using historical crime data publically published by Toronto Police Service**

*Clearly articulate the specific task you will be conducting to help achieve the opportunity*

References:

Park, N. (2021, May 29). Crime type classification using Neural Networks: A brief walkthrough. Medium. <https://nicksypark.medium.com/crime-type-classification-using-neural-networks-a-brief-walkthrough-841b273f9afe>

Walczak, S. (2021, October 7). Predicting crime and other uses of neural networks in police decision making. Frontiers. <https://www.frontiersin.org/articles/10.3389/fpsyg.2021.587943/full#ref20>

## 1.1 Supporting Insights

|  |  |
| --- | --- |
|  | Define any supporting insights, trends and research findings. Where relevant, list key competitors in the market. What are their key messages, products & services? What is their share of market, nationally and regionally? |

Crime has adversely affected people and their society all over the world, and Toronto is no exception. Toronto is the fastest-growing city in North America, and its population has been tremendously grown, resulting in demand rising for policing service. Therefore, Toronto Police Service requires more robust responses to serve community safety. Especially, in the case of emergency, life and death depend on even a seconds-delay in response. Hence, being aware of crime tendency in the city and identifying the crime type immediately is vital for a speedy and well-prepared 911/police officer response. Machine learning has been applied in police decision making and optimization in many countries, and Toronto Police Service also has engaged in it and shared open data source regarding crimes.

References:

McKenzie, A. (2019, April 4). Calling 911: Improving safety and response times for emergency workers. Fleet Complete Blog. <https://blog.fleetcomplete.com/improving-safety-response-times-for-emergency-workers#:~:text=Complete%2C%20can%20help.-,Urgency,literally%20putting%20out%20a%20fire>.

Toronto Police Service – 2022 Operating Budget Request. (n.d.). Toronto.ca. <https://www.toronto.ca/legdocs/mmis/2022/ex/bgrd/backgroundfile-175060.pdf>

## 1.2 Project Gains

|  |  |
| --- | --- |
|  | *Describe any revenue gains, quality improvements, cost and time savings (as applicable). What will you do differently and why would our customers care. What are the implications if we do nothing? This section is particularly key for prioritization against company goals and KPI’s.* |

I am going to help make a 911 response and a police officer response more efficient, which could contribute the safety in the city by conducting this analysis. Using crime type prediction, I am going to improve both quantitative and qualitative factors, which are emergency response time and the survey results regarding response preparedness by police officers and 911 dispatchers. If this project is not implemented, it can be possible that more victims may not be rescued properly, and police officers may not be well-prepared and put in danger.

## *Note: Completion of the following sections is possible only after a careful assessment and triage of the Ask. This is required to determine scope, resource, time, priority and data availability.*

## 2.0 Analytics Objective

|  |  |
| --- | --- |
|  | List the key questions, assumptions and define the hypotheses. Often the deliverable may not just be an analysis output, however a recommended operating model or blueprint for a pilot etc.  Note: Asking the right questions and truly understanding the problem will lead to the right data, right mathematics, and right techniques to be employed. |

1. Revealing the crime tendency for each crime type and location in Toronto.

[Questions]

* Which crime happens most frequentely?
* Which area is a hotspot for any crime type?
* Which area is a hot spot for the specific crime type?
* Which time window are crimes most likely to occur in?
* Which day of week are crimes most likely to occur on?

[Assumptions]

* Crime occurrence has a certain tendency

[Hypothesis]

* N/A (Descriptive analyses do not require hypothesis)

1. Developing the prediction model that classifies crime type using location, time, and premise type information.

[Questions]

* Which crime type will occur when location, time, and premise type information are given?

[Assumptions]

* Crime prediction model could categorize the crime types with necessary information

[Hypothesis]

* N/A (Predictions do not require hypothesis)

## 2.1 Other related questions and Assumptions:

|  |  |
| --- | --- |
|  | List any assumptions that may affect the analysis |

Although all the data required to analyze the crime tendency and predict a crime type are in the dataset, there is one assumption that the crime reporting and data collection were correctly done by police officers and data analysts in Toronto Police Service.

## 2.2 Success measures/metrics

|  |  |
| --- | --- |
|  | What does success look like? Define the key performance indicators (success definition/indicators, drivers and key metrics) against which the objectives will be analyzed. These should be drawn from the interlock meeting with key stakeholders and will inform the approach and methodology for the analysis. |

**KPIs for Analytics objective 1**

1. Showing the following results

* The crime that happens most frequentely
* The area that is a hotspot for any crime type
* The area that is a hot spot for the specific crime type
* The time window that crimes are most likely to occur in
* The day of week that crimes are most likely to occur on

1. Having the suggestion for the most important crimes, areas, and time for a patrol

**KPIs for Analytics objective 2**

1. Achieving the model accuracy greater than 70%
2. Showcasing the simulation using the developed prediction model

**KPIs for the entire project**

1. Demonstrating a significant difference in response time of a 911 dispatcher and police officer between before and after introducing the crime type prediction system when a crime type is unknown.
2. Obtaining the favorable survey results regarding response preparedness by police officers and 911 dispatchers.

References:

Barkved, K. (2022, March 9). How to know if your machine learning model has good performance. Data Science without Code. <https://www.obviously.ai/post/machine-learning-model-performance>

## 2.3 Methodology and Approach

|  |  |
| --- | --- |
|  | Now that you have a good understanding of the Ask and deliverable, detail the recommended approach/methodology. |

**Type of Analysis:** Discriptive analysis and Newral Networks

The first approach will be discriptive analysis to identify the crime tendency in Toronto, such as the most frequent crime, the hotspots for any and each crime type, the time window that crimes are most likely to occur in, and the date of week that the crimes are most likely to occur on. The second approach will be to use neural networks to determine which crime will occur in the specific neighbourhood, time of day, day of week, and location.

**Methodology:** Key questions from ‘Analytics objective’ will be tackled in ascending order as outlined in ‘5.0 Timelines and deliverable section’.

I will start by identifying the useful variables for inputs and data pre-processing. I will then reduce the number of crime types by combining similar classes to overcome the data imbalance. After data partition, I will built the newral networks, such as MLP, CNN, and RNN. Major inputs are neighbourfood, longitude, latitude, time, day of week, and location. Backpropagation and RBF will be used for train. I will also define the performance metrics, such as loss function, accuracy, F1 score, and AUC score.

**Output:** The output will be a set of insights and strategic recommendations that will help us to optimize a 911 call and a police officer responses.

References:

Park, N. (2021, May 29). Crime type classification using Neural Networks: A brief walkthrough. Medium. <https://nicksypark.medium.com/crime-type-classification-using-neural-networks-a-brief-walkthrough-841b273f9afe>

Walczak, S. (2021, October 7). Predicting crime and other uses of neural networks in police decision making. Frontiers. <https://www.frontiersin.org/articles/10.3389/fpsyg.2021.587943/full#ref20>

## 3.0 Population, Variable Selection, considerations

|  |  |
| --- | --- |
|  | Capture learning about the data available today location, structure, and reliability; this would include data in operational systems including dealer sourced, data warehouse and any CRM or email marketing systems available today. |

**Audience/population selection:** Civilians, Police officers, Data analysts

**Observation window:** 2014/01/01 – 2021/12/31

**Inclusions:** All the rows of 281692

**Exclusions:** The columns of X, Y, Index, event\_unique\_id, Devision, reporteddate, premise, ucr\_code, ucr\_ext, offence, reportedyear, reportedmonth, reporteddayofyear, reportedhour, reporteddayofweek, occurrenceyear, occurrencemonth, occurrenceday, occurrencedayofyear, Hood\_ID, ObjectId

**Data Sources:** Major Crime Indicators <https://data.torontopolice.on.ca/search?q=crime>

**Audience Level:** N/A

**Variable Selection:** The columns of occurrencedate, location\_type, occurrencedayofweek, occurrencehour, MCI, Neighbourhood, Long, Lat

**Derived Variables:** MCI, location\_type

**Assumptions and data limitations:** Please refer to 4.0 Dependencies and Risks

## 4.0 Dependencies and Risks

|  |  |
| --- | --- |
|  | Identification of key factors that may influence the outcome of the project and likelihood of it happening: |

|  |  |  |  |
| --- | --- | --- | --- |
| **Risk** | **Likelihood (based on historical data)** | **Delay (based on historical data)** | **Impact** |
| ・Reported crimes do not cover all the crimes that actually occurred | Medium |  | This analysis is based on the calls for help by victims, so it will not be a big issue. |
| ・Police officers’ reports are not always accurate | Medium |  | It could affect the analysis, but it could be a very small impact. |

## 5.0 Deliverable Timelines

|  |  |
| --- | --- |
|  | List key dates and timelines as a work-back schedule. Activate line items based on complexity and line-of-sight required. Will set the stakeholder expectations for the process. |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Item** | **Major Events / Milestones** | **Description** | **Scope** | **Days** | **Date** |
| 1. | Kick-off / Formal Request | Meeting with my advisor, and check if the datase, and analytic methodology and approach are feasible | Approval of dataset and propose the analytics plan | 1 | July 12th, 2022 |
| 2. | Assessment / Triage | Getting some advise on the analytic plan and making it better | Improvement of the analytic plan | 5 | July 17th, 2022 |
| 3. | Prioritization | Finalize the analytic plan | Approval of the analytic plan | 3 | July 20th, 2022 |
| 4. | Data Exploration & Analysis   * Issues with duplicates * Issues with Spend data | Performing exploratory data analysis and descriptive analysis, and developing the predictive models | Completion of EDA and predictive models | 16 | August 5th, 2022 |
| 5. | Governance | Impementing risk management to identify possible risks | Making a final change in modeling | 2 | August 7th, 2022 |
| 6. | Documentation | Putting all the analysis into the word document | Making the document for the analysis | 5 | August 12th, 2022 |
| 7. | Internal team Presentation | Presenting to the internal team | Performing the presentation | 5 | August 12th, 2022 |
| 8. | Go/No Go | Meeting the advisor and getting an final approval | Approval for the entire project | 7 | August 19th, 2022 |
| 9. | Presentation | Presenting to the management | Performing the presentation | 1 | August 22-24th, 2022 |
| 10. | Pilot | Testing the model in the field | Testing the model | 3 | August 29-31st, 2022 |
| 11. | Delivery & sign-off | Initiating the predictive model in the real field | Project delivery | 1 | September 1st, 2022 |