

|  |
| --- |
| How is crime influenced by the surroundings? |
|  |
| December 24  Authored by: Nutan |

Contents

[Introduction/Business Problem 3](#_Toc28092651)

[Target Audience 3](#_Toc28092652)

[Data 3](#_Toc28092653)

[Sources 3](#_Toc28092654)

[Data Description 3](#_Toc28092655)

[Linking data to problem 5](#_Toc28092656)

[Methodology 5](#_Toc28092657)

[Data Preparation & Transformation 5](#_Toc28092658)

[Methodology 5](#_Toc28092659)

[Results 6](#_Toc28092660)

[Result Discussion 6](#_Toc28092661)

[Conclusion 6](#_Toc28092662)

[Further next steps 6](#_Toc28092663)

# Introduction/Business Problem

Crime is a problem in any city and being prepared and planning to avert them is the key to dealing with it.

In this data analysis exercise , I am aiming to explore and identify quantitatively if there is a pattern to crime i.e. what areas does crime occur in and do the surrounding locations have an impact on the number and nature of the crime that happens in an area.

Toronto, the capital of the province of Ontario, is the most populous Canadian city. and will be the focal for us in this exercise

## Target Audience

I am hoping that an analysis like this will help in planning for the police department - where and how should they organize themselves to better thwart crime

It will also help citizens plan their way around the city and be cautious by knowing that certain surroundings and locations are more likely targets for crime than others.

It might also help business who are willing to set shop pick areas to make it safe place to work for employees and for the business.

There might be some intuitions or hypothesis on areas and crime but a quantitative lens offers a more useful & concrete way to evaluate those and will also help plan and act correctly.

# Data

## Sources

Toronto city has taken steps in the last few years to democratize data and information and will be our source in this exercise.

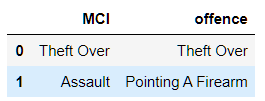
The central data will be from the Toronto Police department on Major Crimes called the [MCI data](http://data.torontopolice.on.ca/datasets/mci-2014-to-2018)

Data on the location and other information about various venues in Toronto is from the Foursquare’s explore API

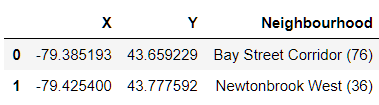
## Data Description

Lets go through some key columns in MCI data.

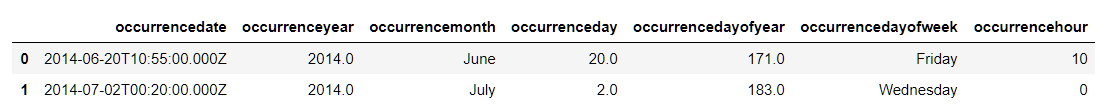
The data has the nature of the crime in the MCI and Offence columns

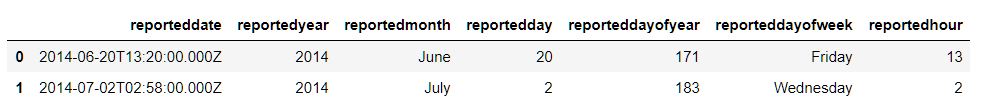


It has latitude [column named Y] and longitude data [column named X] and the name of the neighbourhood



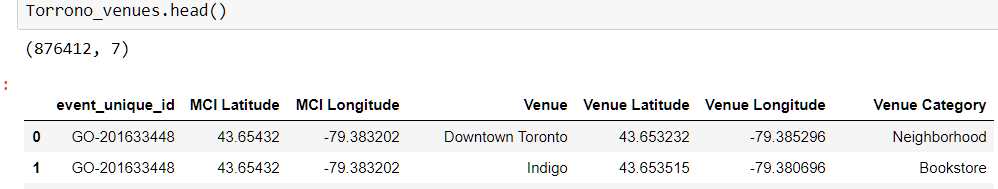
It has occurrence date when the crime has occurred and reported date i.e. when the crime was reported. There are also a number of columns which break down the dates into days, years, months, weekdays etc.



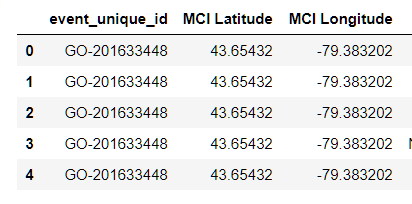


The Foursquare API is expected to give us venue details close to each crime geo coordinates

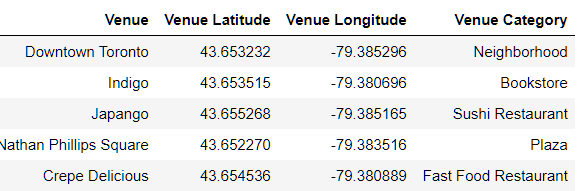
The details of how this data was extracted will be elaborated in the data preparation and transformation section of the report but he final data structure is as below



i.e. for each event id , where the crime occurred



The details of the venues which are within 300 m of the crime scene are listed along with Venue Name, Venue category and the location of the venue



## Linking data to problem

With the MCI and foursquare data we have

* Type of the Crime – [MCI] column in the data
* Details of the crime
  + When the crime occurred - [occurrencedate, 'occurrenceyear', 'occurrencemonth', 'occurrenceday', 'occurrencedayofyear', 'occurrencedayofweek', 'occurrencehour']
  + Where the crime occurred - ['X', 'Y','Neighbourhood']
* Venues which are close to the location of crime scene
  + ['Venue', 'Venue Latitude', 'Venue Longitude', 'Venue Category']

This data now allows for understanding the relationship between the crime scene and the venues surrounding the crime scene and if a particular type of crime scene is related to the surrounding venues

# Methodology

## Data Preparation & Transformation

Exploratory Data Analysis

## Methodology

section which represents the main component of the report where you discuss and describe any exploratory data analysis that you did, any inferential statistical testing that you performed, if any, and what machine learnings were used and why

Foursquare API is very useful online application used my many developers & other applications like Uber etc. In this project I have used it to retrieve information about the places present in the neighbourhoods of Toronto. The API returns a JSON file and we need to turn that into a data-frame. Here I’ve chosen 100 popular spots for each neighborhood within a radius of 300 m.

# Results

# Result Discussion

# Conclusion

## Further next steps