

# Analysis on break and enter crime in Toronto from 2014 to 2019

## Introduction

The aim of this project is to find a safe and secure location for living or opening a business in Toronto, Canada. Specifically, this report will be targeted to the break and enter crime situation over the past 5 years in different neighbourhood in Toronto.

We will make use of our data science tools to analyse data and focus on the trend of each borough from 2014 to 2019, and will focus on the 2019 data specifically to choose the safest and most dangerous boroughs by analysing crime data and short listing the neighbourhood.

## Data

Based on definition of our problem, factors that will influence our decision are:

finding the most dangerous borough based on crime statistics in 2019

finding the trend of break and enter crime for each borough

We will be using the geographical coordinates of Toronto to plot neighbourhoods in a borough that is dangerous and in the city's vicinity, and finally cluster our neighborhoods and present our findings.

Following data sources will be needed to extract/generate the required information:

Part 1: Using a real world data set from Toronto Police Service website: A dataset consisting of the crime statistics of each Neighbourhood in Toronto along with location, recorded year, premise type.

<https://data.torontopolice.on.ca/datasets/break-and-enter-2014-to-2019/data> [https://prod-hub-indexer.s3.amazonaws.com/files/d9b3dd6402454c379ba57994230aabea/0/full/3857/d9b3dd6402454c379ba57994230aabea\\_0\\_full\\_3857.csv](https://prod-hub-indexer.s3.amazonaws.com/files/d9b3dd6402454c379ba57994230aabea/0/full/3857/d9b3dd6402454c379ba57994230aabea_0_full_3857.csv)

Part 2: Gathering additional information of the list of officially categorized boroughs in Toronto: Borough information will be used to map the existing data where each neighbourhood can be assigned with the right borough.

Part 3: Creating a new consolidated dataset of the Neighborhoods, along with their boroughs, crime data and the respective Neighbourhood's co-ordinates.: This data will be fetched using OpenCage Geocoder to find the safest and most dangerous borough and explore the neighbourhood by plotting it on maps using Folium and perform exploratory data analysis.

## Methodology

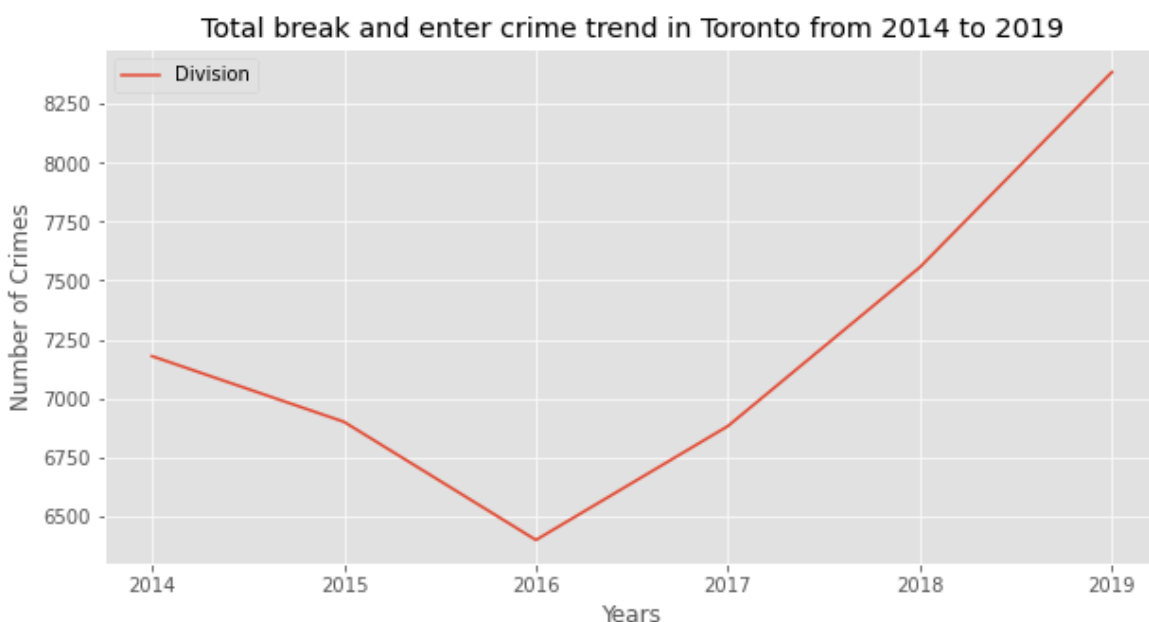
Categorized the methodology section into two parts:

Exploratory Data Analysis: Visualise the crime reports in different Toronto district to identify the 10 most dangerous area. Also to see the trend of each year

Modelling: To help stakeholders choose the right neighborhood within a borough we will be clustering similar neighborhoods using K - means clustering which is a form of unsupervised machine learning algorithm that clusters data based on predefined cluster size.

We will use K-Means clustering to address this problem so as to group data based on existing venues which will help in the decision making process.

1. see the trend for the break and enter from 2014 to 2019



we can observe from the chart, the break and enter crime decrease to the bottom in 2016, and from then on, it has increased sharply.

2. Let's continue to see the trend for the 5 most dangerous areas trend over these 5 years.

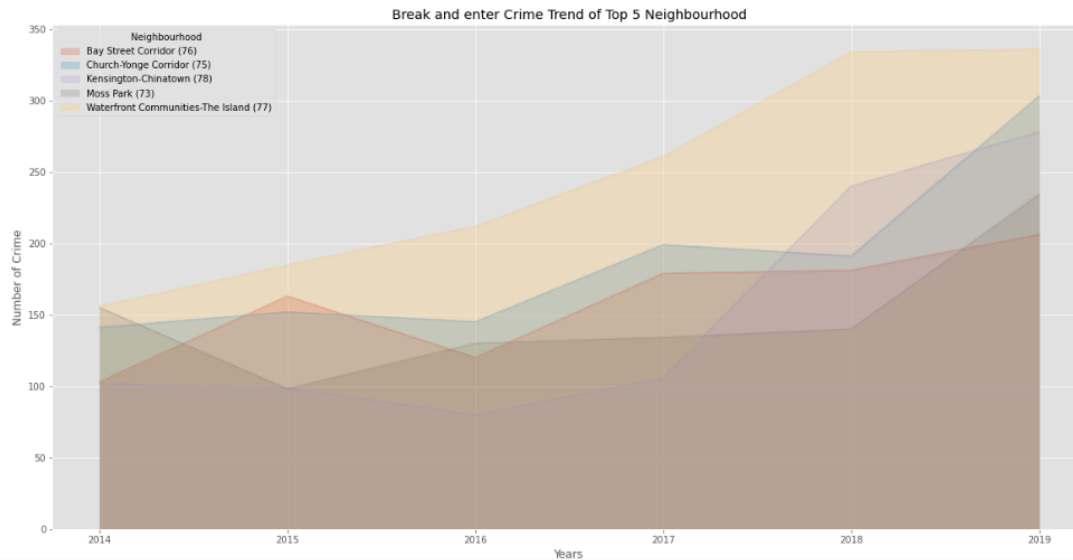
```

Waterfront Communities-The Island (77)    1484
Church-Yonge Corridor (75)                1131
Bay Street Corridor (76)                  952
Kensington-Chinatown (78)                905
Moss Park (73)                           891
...
Caledonia-Fairbank (109)                  87
Guildwood (140)                           87
Maple Leaf (29)                           76
Lambton Baby Point (114)                   75
Elms-Old Rexdale (5)                       63
Name: Neighbourhood, Length: 140, dtype: int64

```

they are Waterfront Communities-The Island (77), Church-Yonge Corridor (75), Bay Street Corridor (76), Kensington-Chinatown (78), and Moss Park (73):

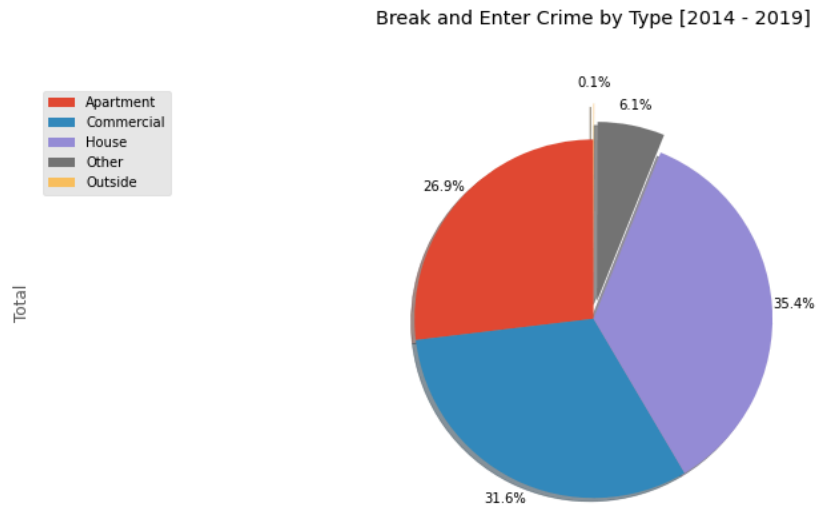
Neighbourhood	Bay Street Corridor (76)	Church-Yonge Corridor (75)	Kensington-Chinatown (78)	Moss Park (73)	Waterfront Communities-The Island (77)
reportedyear					
2014	103	141	102	155	156
2015	163	152	99	98	185
2016	120	145	80	130	212
2017	179	199	106	134	261
2018	181	191	240	140	334
2019	206	303	278	234	336



from the chart, all five areas the crime are on the trend, and the water front community (77) are higher than the other area.

### 3. The premise type analysis

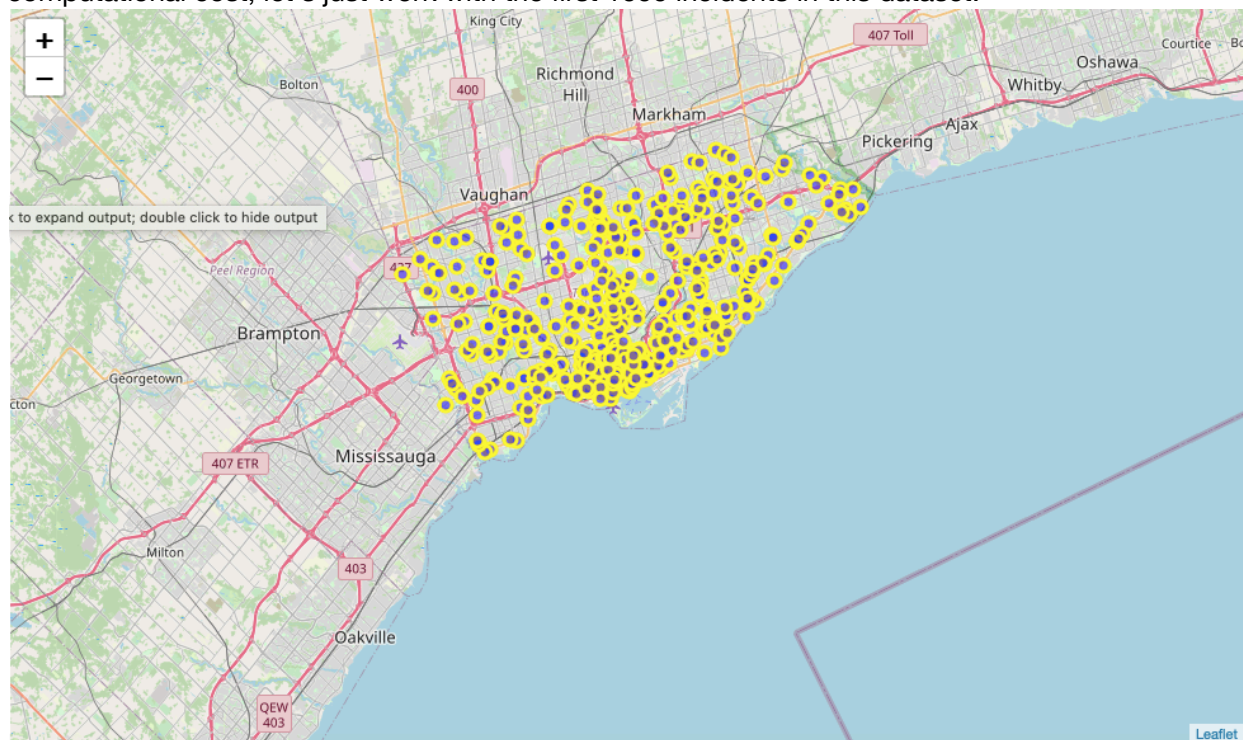
reportedyear	2014	2015	2016	2017	2018	2019	Total
premisetype							
Apartment	1865.0	1982.0	1599.0	1818.0	1986.0	2398.0	11648.0
Commercial	1916.0	1933.0	1900.0	2118.0	2552.0	3263.0	13682.0
House	2941.0	2637.0	2495.0	2575.0	2505.0	2169.0	15322.0
Other	457.0	346.0	402.0	369.0	511.0	541.0	2626.0
Outside	NaN	2.0	2.0	2.0	3.0	15.0	24.0



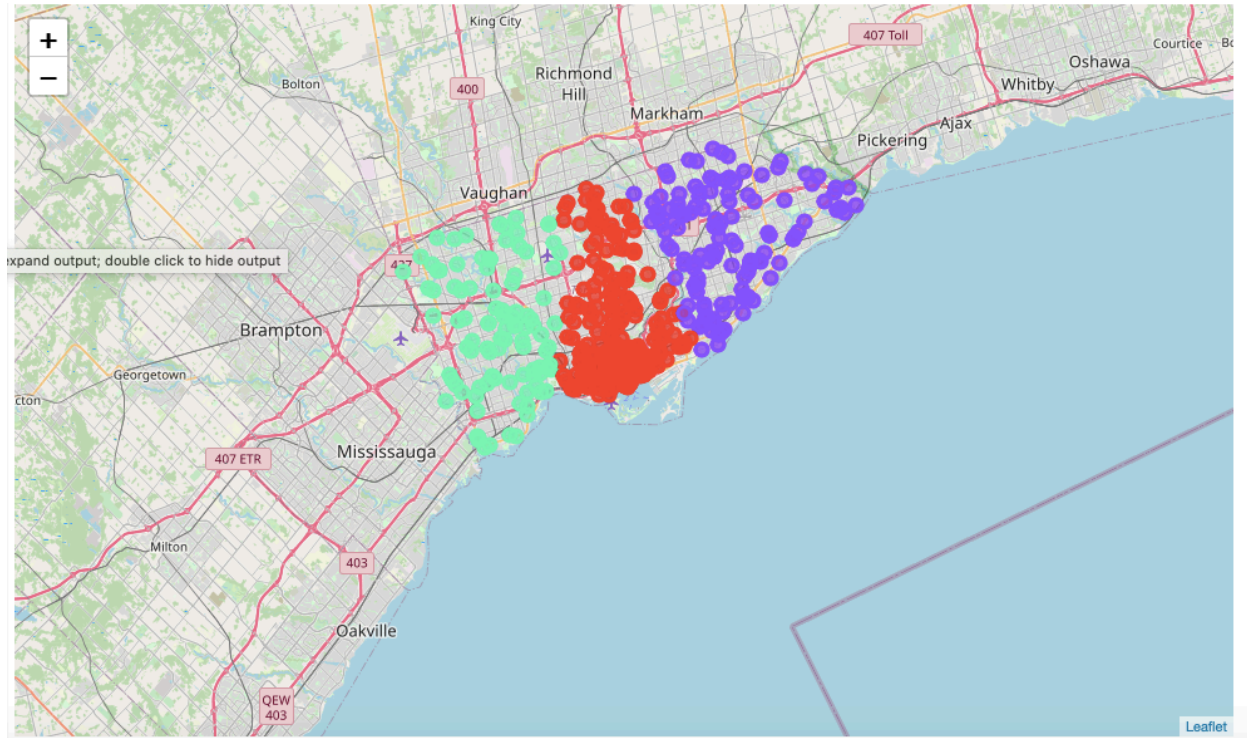
Houses and commercial estate take up most of the break and Enter Crime type.

#### 4. Visualization and k-means modeling

The dataframe consists of 8386 crimes, which took place in the year 2019. In order to reduce computational cost, let's just work with the first 1000 incidents in this dataset.



We use  $k=3$  to divide them into 3 clusters:



## Result and Discussion

The objective of the problem was to help stakeholders identify the most dangerous borough in Toronto to purchase a house or commercial estate. This has been achieved by first making use of Toronto crime data. We achieved this by grouping the neighborhoods into clusters to assist the police or the government to decide different area to govern by providing them with relevant data.

## Conclusion

We have explored the crime data to understand break and enter crimes in all neighborhoods of Toronto and later categorized them into different boroughs, this helped us group the neighborhoods into boroughs and choose the dangerous borough first. We can see from the chart that the crime trend is on the rise, and the break and enter crime into house is the most common type. The most dangerous areas are concentrated in the downtown area. The whole crime map can be divided into 3 clusters to let police or government to manage.