

# Analysis on Break and enter Crime in Toronto from 2014 to 2019

Coursera IBM Data Science Final Project



# 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

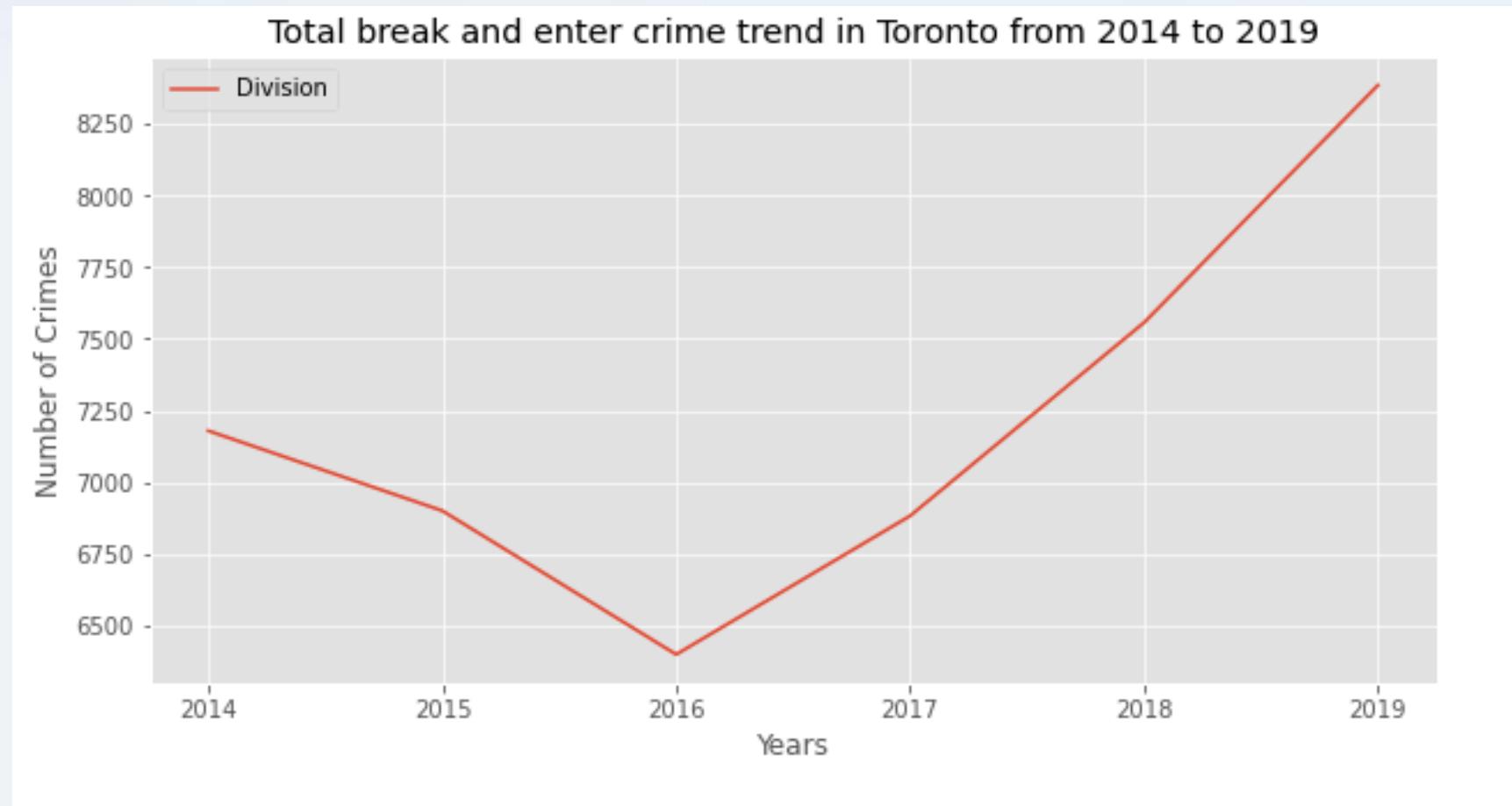
# Data

- 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 Tronto Police Service website: A dataset consisting of the crime statistics of each Neighbourhood in Toronto along with location, recorded year, premise type.  
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.

# Methodology

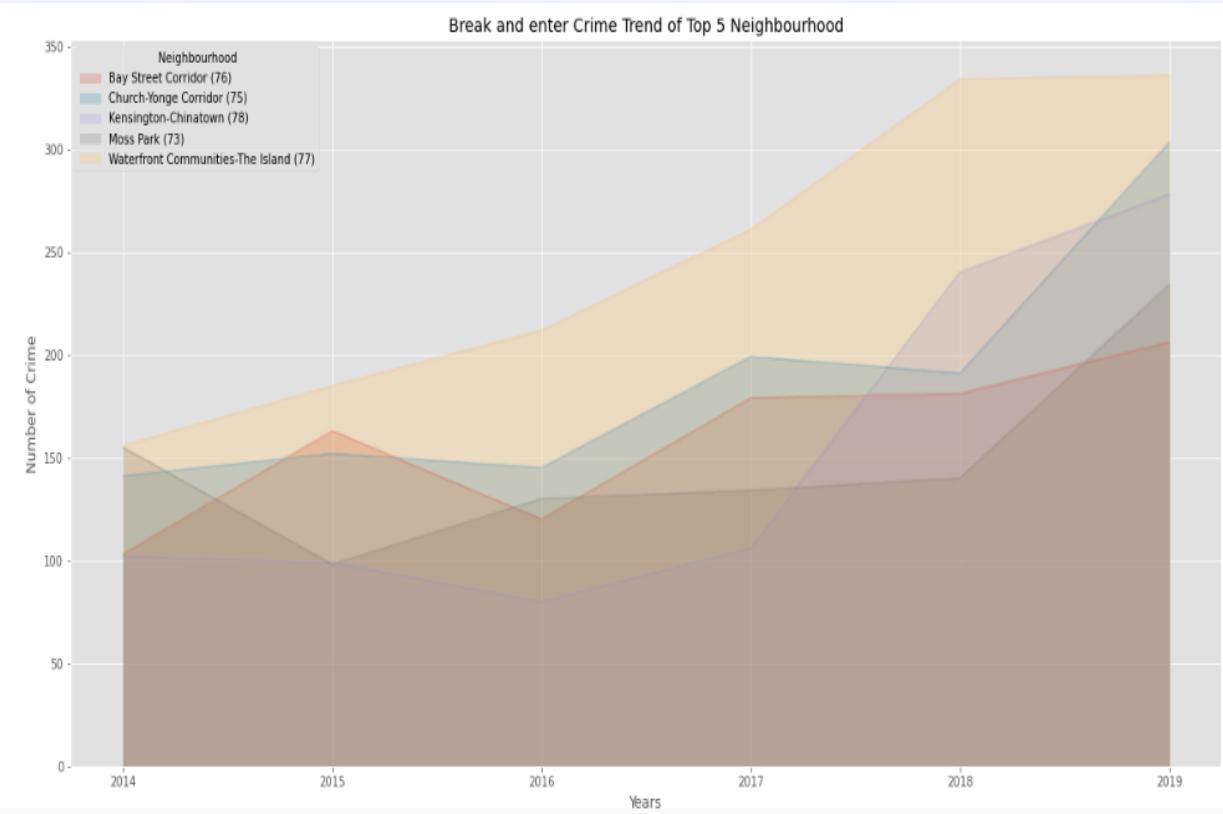
- Categorized the methodology section into two parts:  
Exploratory Data Analysis: Visualise the crime reports in different Toronto district to identify the 10 most dangerous area. Also to see the trend of each year  
Modelling: To help stakeholders choose the right neighborhood within a borough we will 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.

the trend for the break and enter from 2014 to 2019

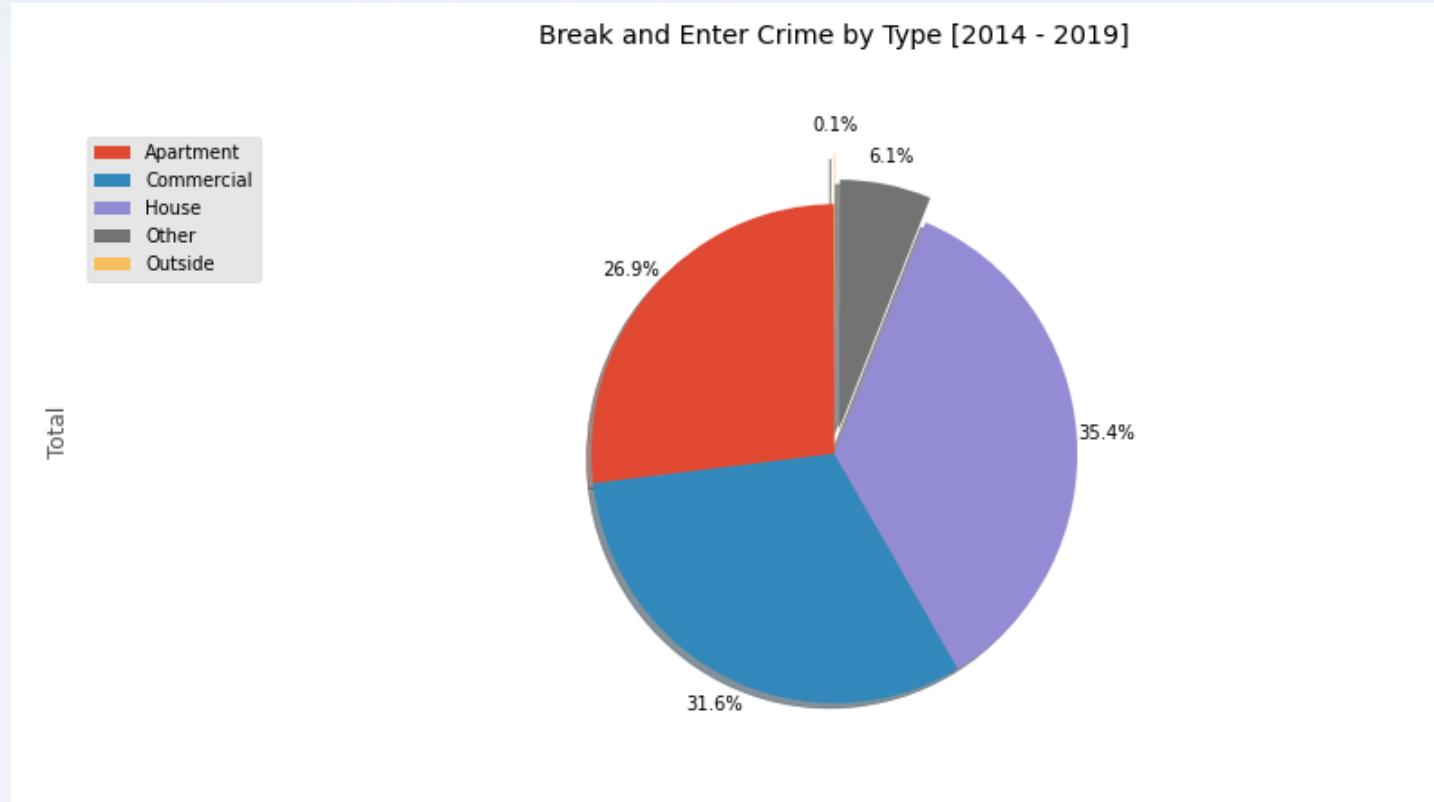


the 5 most dangerous areas trend over these 5 years

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	

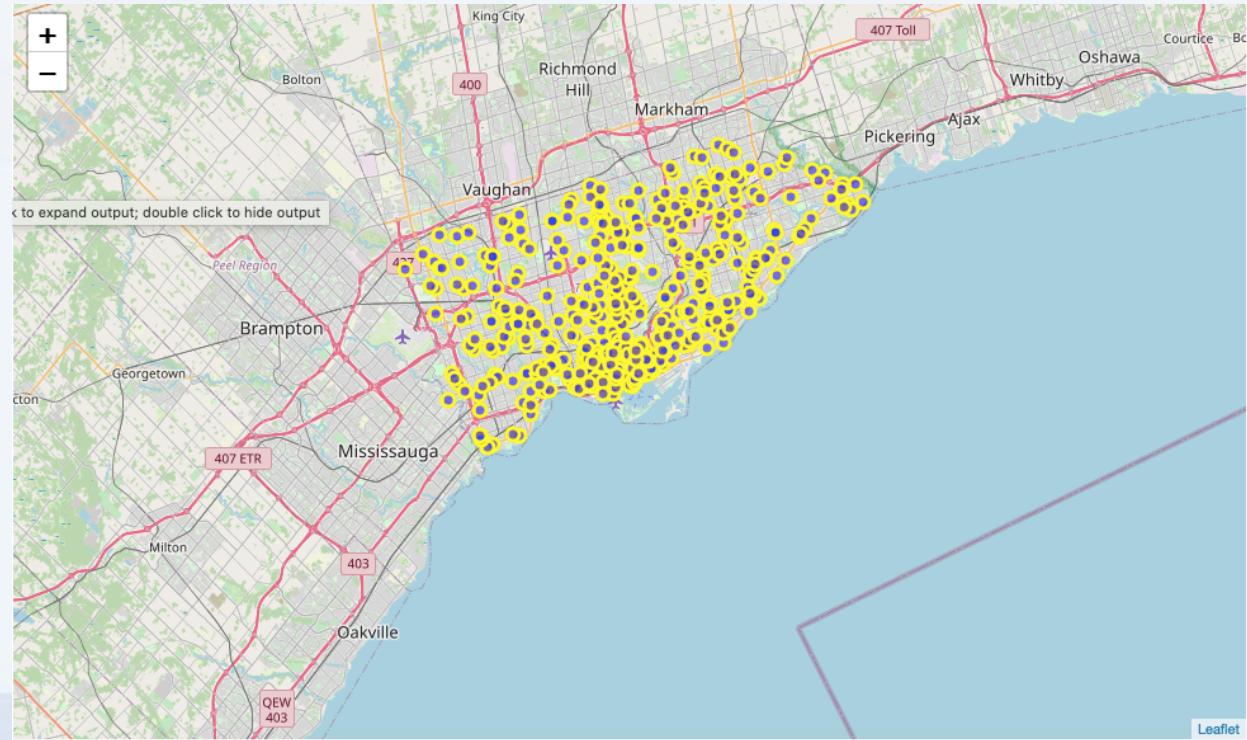


# premise type analysis

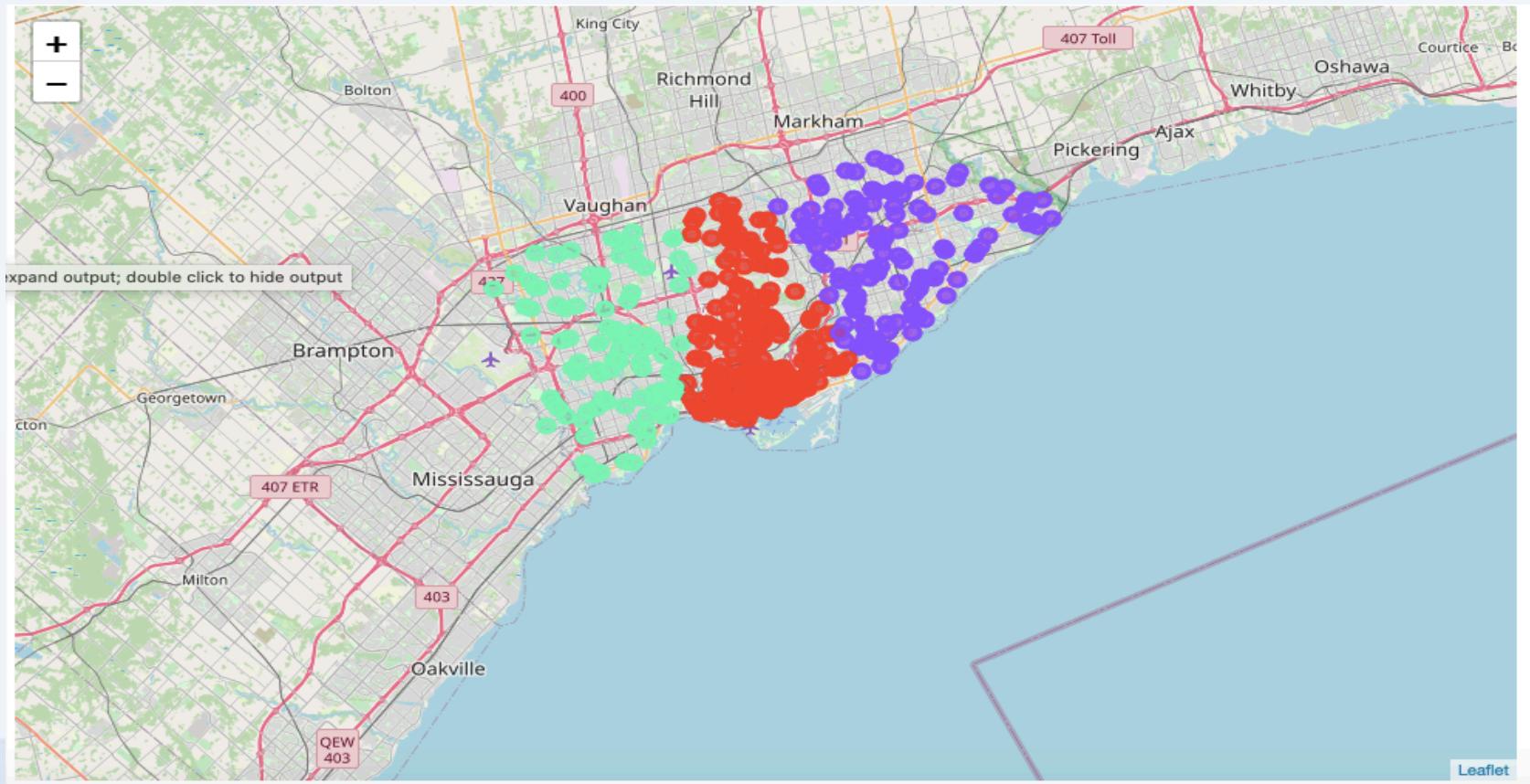


# Visualization and k-means modeling

- 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.