How is crime influenced by the surroundings?

DECEMBER 26

Authored by: Nutan

Contents

Introduction/Business Problem

Crime is a problem in any city and being prepared and planning to avert it is the key to dealing with it.
Crime occurs for various reasons and in various circumstances and could be localized based on the surroundings. It could be advantageous to understand the impact of the variables on the occurrence of crime to make it more decipherable in pattern and occurrence.
Toronto, the capital of the province of Ontario, is the most populous Canadian city. and will be the focal for us in this exercise

Problem

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

Target Audience

I am hoping that an analysis like this will help the police department in planning - 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 businesses, 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 acquisition and cleaning

Sources

Toronto city has taken steps in the last few years to democratize data and information and will be our source for this exercise.
The central data will be from the Toronto Police department on Major Crimes called the MCI data

Data on the location and other information about various venues in Toronto is from the Four square'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

	MCI	offence
0	Theft Over	Theft Over
1	Assault	Pointing A Firearm

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

	X	Y	Neighbourhood
0	-79.385193	43.659229	Bay Street Corridor (76)
1	-79.425400	43.777592	Newtonbrook West (36)

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.

	occurrencedate	occurrenceyear	occurrencemonth	occurrenceday	occurrencedayofyear	occurrencedayofweek	occurrencehour
0	2014-06-20T10:55:00.000Z	2014.0	June	20.0	171.0	Friday	10
1	2014-07-02T00:20:00.000Z	2014.0	July	2.0	183.0	Wednesday	0

	reporteddate	reportedyear	reportedmonth	reportedday	reporteddayofyear	reporteddayofweek	reportedhour
0	2014-06-20T13:20:00.000Z	2014	June	20	171	Friday	13
1	2014-07-02T02:58:00.000Z	2014	July	2	183	Wednesday	2

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

```
Torrono_venues.head()
```

```
(876412, 7)
```

:

	event_unique_id	MCI Latitude	MCI Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	GO-201633448	43.65432	-79.383202	Downtown Toronto	43.653232	-79.385296	Neighborhood
1	GO-201633448	43.65432	-79.383202	Indigo	43.653515	-79.380696	Bookstore

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

	event_unique_id	MCI Latitude	MCI Longitude
0	GO-201633448	43.65432	-79.383202
1	GO-201633448	43.65432	-79.383202
2	GO-201633448	43.65432	-79.383202
3	GO-201633448	43.65432	-79.383202
4	GO-201633448	43.65432	-79.383202

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

Venue	Venue Latitude	Venue Longitude	Venue Category
Downtown Toronto	43.653232	-79.385296	Neighborhood
Indigo	43.653515	-79.380696	Bookstore
Japango	43.655268	-79.385165	Sushi Restaurant
Nathan Phillips Square	43.652270	-79.383516	Plaza
Crepe Delicious	43.654536	-79.380889	Fast Food Restaurant

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

The data from MCI is primarily from 2014-2018. In this case we restricted our analysis to an years' worth of data from 2016 also cut down on some of the columns. This gave a dataset of 32K rows and 19 columns to explore.

Repeated columns like `[lat, long]` or unique ID columns `[ucr_code, ucr_ext, Division, hood-id, ObjectId ]` are dropped off for the rest of the analysis

```
MCI_2016 = MCI_raw[MCI_raw['reportedyear'] == 2016][['X', 'Y', 'event_unique_id', 'premisetype', 'offence', 'reportedyear',
                                                     'reportedmonth', 'reportedday', 'reporteddayofyear','reporteddayofweek',
                                                     'reportedhour', 'occurrenceyear', 'occurrencemonth', 'occurrenceday',
                                                     'occurrencedayofyear','occurrencedayofweek', 'occurrencehour', 'MCI',
                                                     'Neighbourhood']].copy()

print(MCI_2016.shape)
MCI_2016.head(2)
```

```
(32816, 19)
```

Each crime in the MCI data has an `[event_unique_id]` and each `[event_unique_id]` has geo coordinates in the form of `[X, Y]`

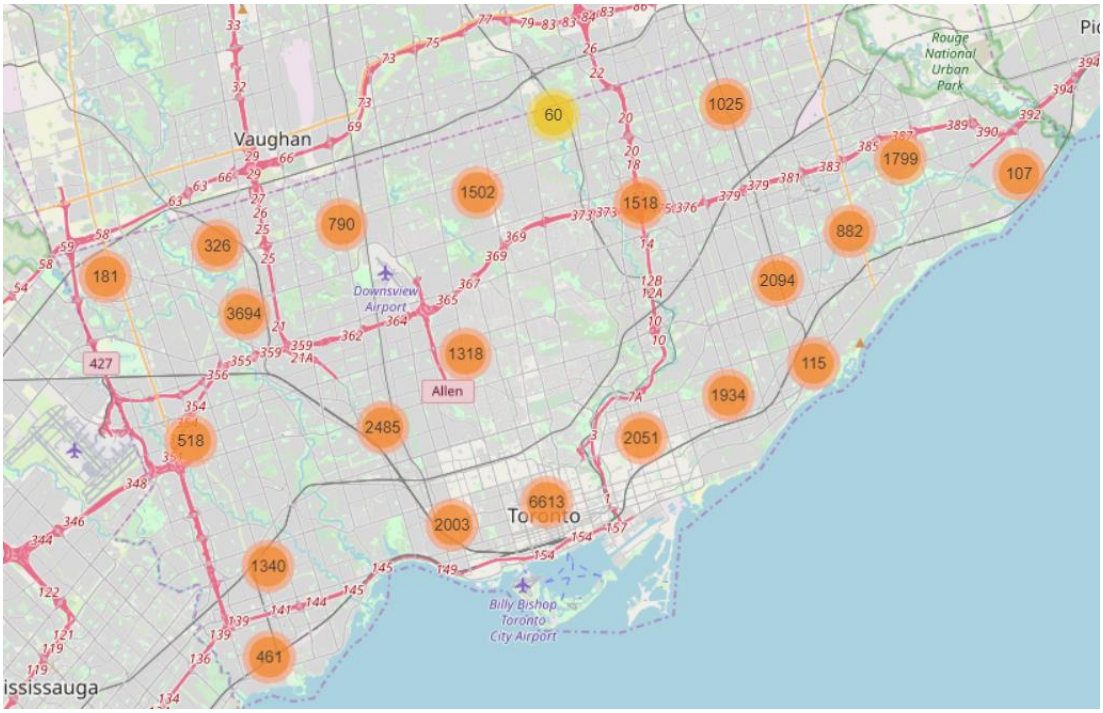
These coordinates are used to query the foursquare API to get the details of the venues within a 300m radius of the crime scene

Foursquare returns the results in the form of JSON . This JSON is then parsed to extract [`'Venue', 'Venue Latitude', 'Venue Longitude', 'Venue Category'`] details and these were then combined with the Crime scene event id [`event_unique_id`] and the coordinates of the crime scene `[X, Y]` to arrive the data set which contained all the venues within a 300 m radius of the crime scene

Exploratory Data Analysis

The goal of EDA was to better understand the data and the patterns in the data which could help interpret the final modelling results better

First the data was plotted to understand the distribution of the crimes across the city.
Downtown Toronto has a larger share of crimes than other neighbourhoods as can be seen in the map below

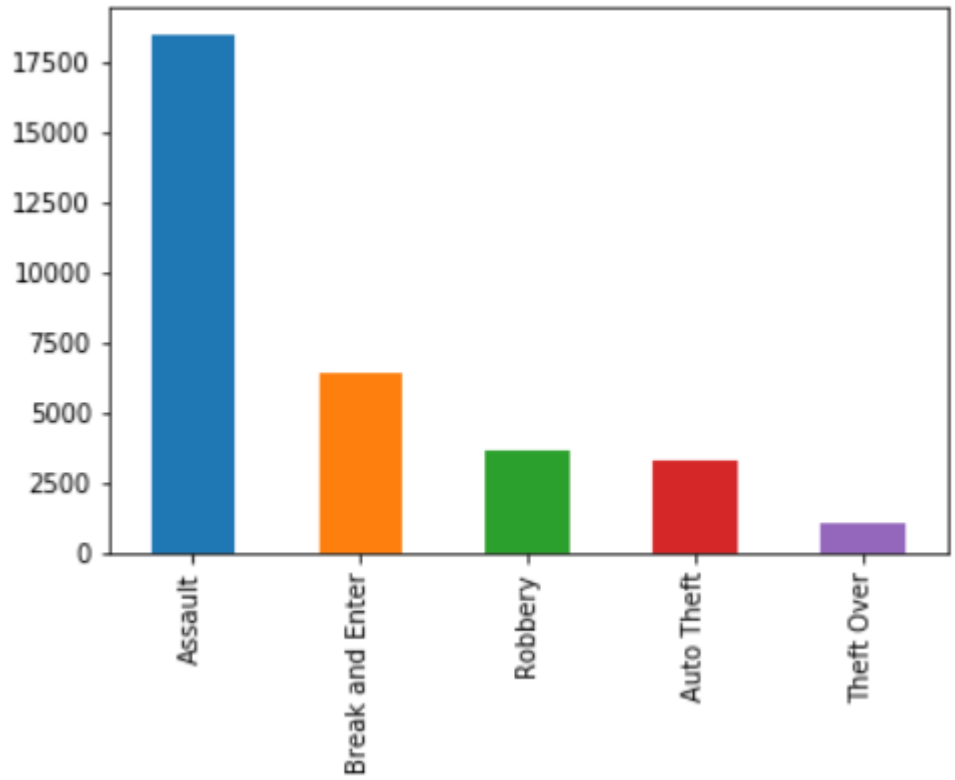


The second highest number of crimes are in the north western part of the city

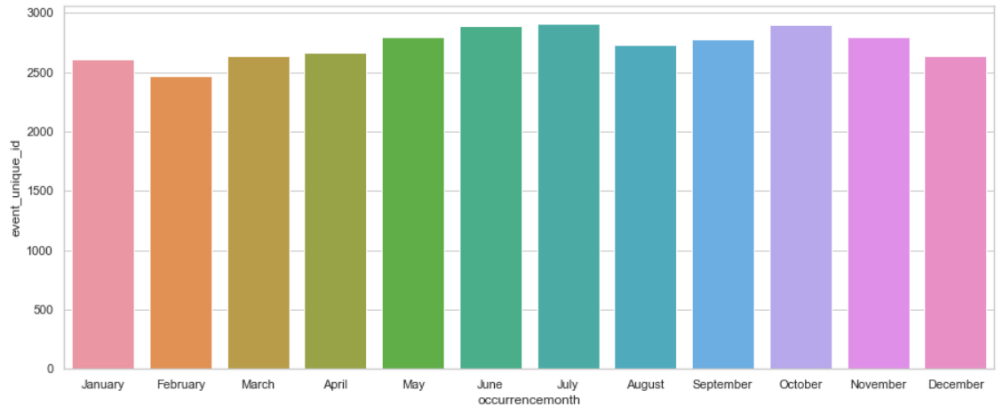
Similar story comes from the neighbourhood list as well. Church-Yonge and Water front are communities in downtown Toronto have the highest number of major crimes

Neighbourhood	event_unique_id
Church-Yonge Corridor (75)	1151
Waterfront Communities-The Island (77)	1141
West Humber-Clairville (1)	818
Moss Park (73)	725
Bay Street Corridor (76)	667
York University Heights (27)	646

A chart on the type of crime frequency indicates that Assault is a very prevalent crime and is almost 3X more frequent than the next frequent crime – Break and Enter.

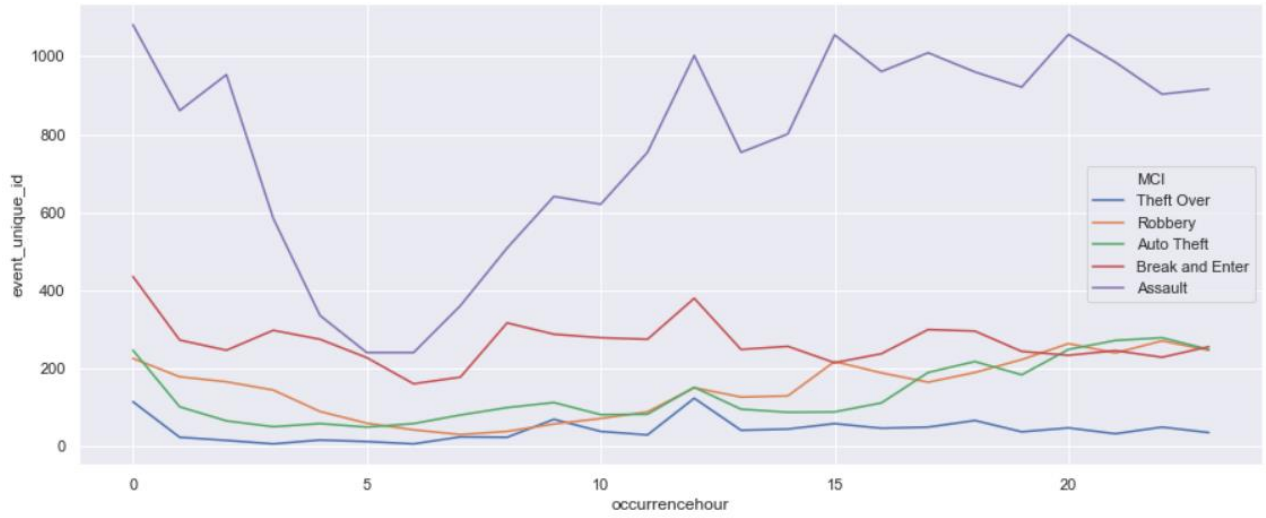


Summer months , with the exception of October have a higher crime rate than the winter months



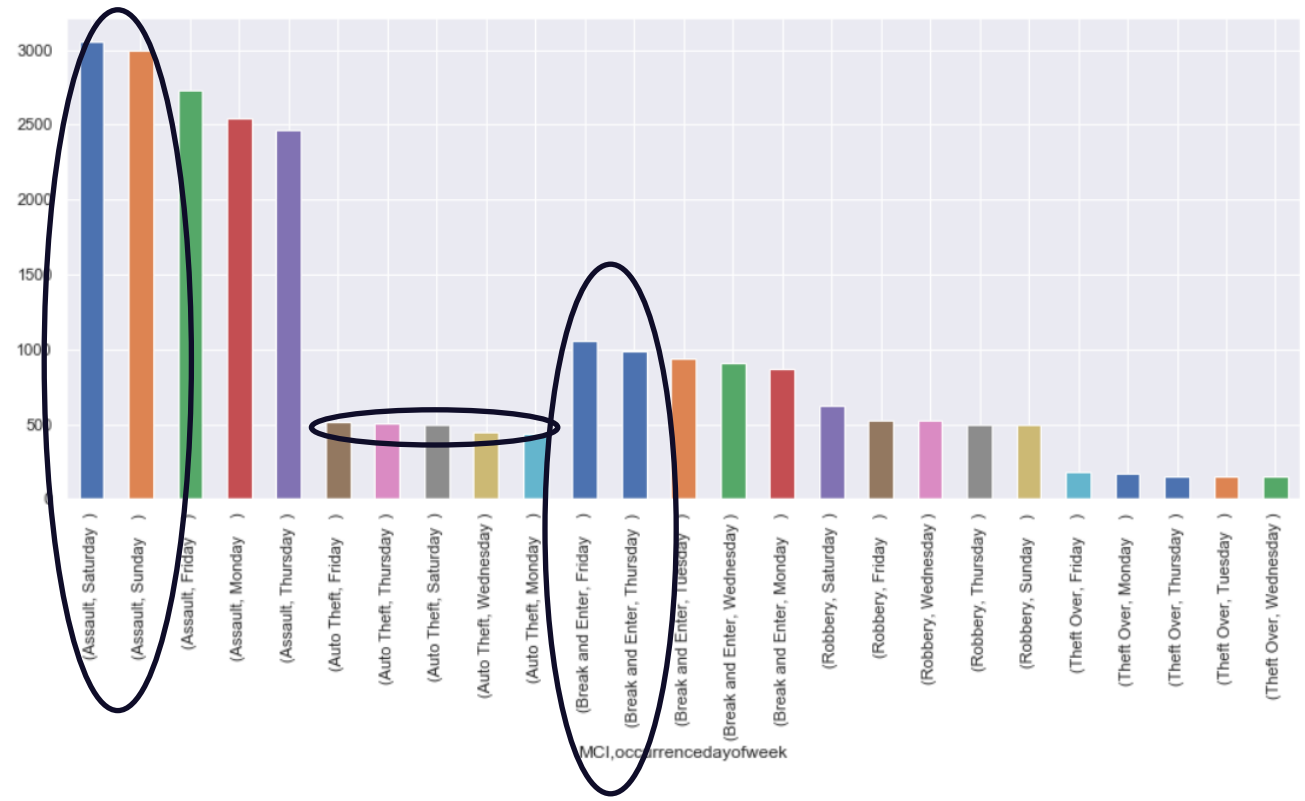
A pivot on the occurrence hour of crime by the type of crime , shows that

- Assault is highest at midnight, but is also higher in the evenings & nights
- Break & Enter has a different pattern, its relatively flat all through out the day but has a high during midnight(possibly as lesser number of people would be around)
- Auto Theft, Robbery are more frequent during evenings and nights



To profile crimes better, they are compared against each other based on the occurrence day of the week.

- Assault seems to occur equally frequently during any day , Weekends seem be relatively higher
- Friday for Break and enter criminal and almost any day for an auto thief seem to be good days. Noticeably Break and enter seems the least frequent during Saturdays and Sundays . The auto theft graph is relatively flat all through the week



Methodology & Algorithm used

The approach towards understanding the relationship between venues and crimes was through clustering analysis. Intention is to use the venues near each crime scene and then put them through a clustering algorithm so the unsupervised algorithm can identify a relationship between the crime and the venues nearby.

The venues data from foursquare was transformed into one hot encoded data based on venue categories. i.e. the rows representing a crime scene , the columns the venues close to the crime scene and the values in each cell the average frequency with which the venue is close to the crime scene

This data was put through a 5 cluster KNN algorithm

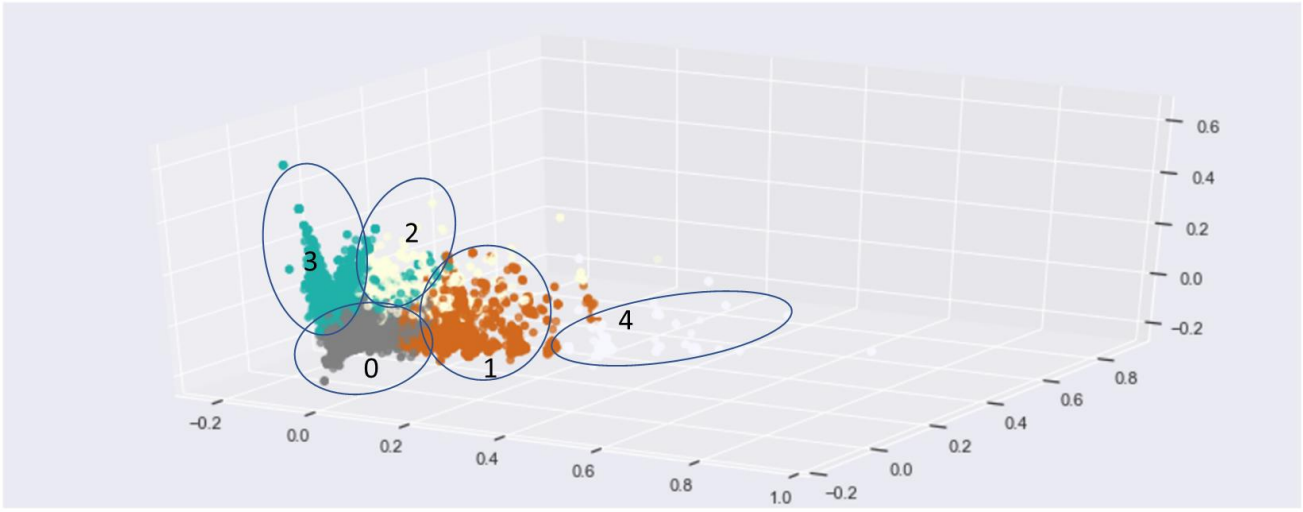
The results of the clustered data were merged with the crime scene data and the venue data. This data frame now allows for an end to end analysis – Crime data -> Venues close to the Crime Scene -> Cluster results

	Crime Meta data					Crime Details		Cluster Labels	Venue Details	
	X	Y	event_unique_id	premisetype	offence	occurrenceyear	occurrencemonth	Cluster Labels	1st Most Common Venue	2nd Most Common Venue
5055	-79.383202	43.654320	GO-201633448	Outside	Robbery - Swarming	2016.0	January	0.0	Coffee Shop	Clothing Store
5288	-79.544701	43.632122	GO-201630054	Commercial	B&E	2016.0	January	3.0	Fast Food Restaurant	Vietnamese Restaurant
5329	-79.433281	43.637745	GO-201629453	Apartment	Assault With Weapon	2016.0	January	0.0	Café	Pizza Place
5330	-79.374352	43.662918	GO-201629375	Apartment	B&E	2016.0	January	0.0	Japanese Restaurant	Coffee Shop
5331	-79.403870	43.666660	GO-201634311	Outside	Robbery - Mugging	2016.0	January	0.0	Café	Restaurant

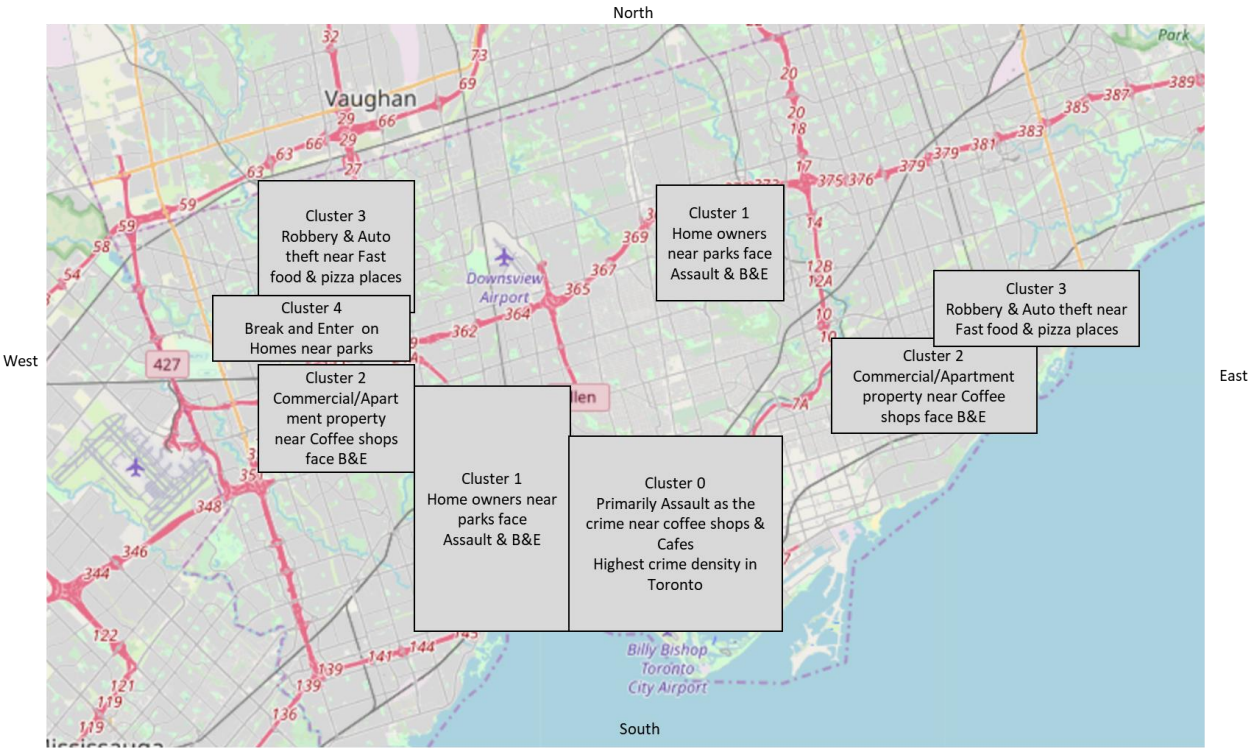
Results

For visualizing the clusters, the attributes are reduced to 3 dimensions using PCA and are plotted in a 3D scatter chart

- Cluster 4 seems to be evidently different from the rest of the clusters, towards the right bottom of the chart (pink coloured dots) , It also has some outlier
- Rest of the clusters seem to be closely knit
- Clusters 0 & 3 are towards the left of the chart and seems to have a distinct identity
- Clusters 1 & 2 are close to the clusters 0 & 3 , with cluster 2 almost overlapping the others around



Analysing each cluster with the crime, venue and premise lens gives a perspective of the nature of the clusters . The notebook has detail analysis but a summary of the clusters for the ease of consumption is in the below image.



Cluster 0 – Centered around down town Toronto , Almost all MCI Crime types are higher in absolute numbers in this region ,crime scenes are largely around coffee shops and Cafes , possibly as the down town itself has a larger share of them

Cluster 1 – Assault, Break and Enter are relatively higher with the risk majorly around homes near parks

Cluster 2 – Comparable to cluster 1 in some ways but different in the targets of crime primarily. Cluster 2 is split on either side of the city centre i.e. towards the Western and Eastern parts of the city. While Cluster 1 was around homes, cluster 2 focusses on Apartments and Commercial spaces with Coffee shops in the neighbourhood. Residents here risk facing Break and Enter, as it is in relatively higher proportion.

Cluster 3 –Also a split cluster, with similar proportion in the north western and eastern parts of Toronto. Robbery and Auto Theft seems to be relatively higher surrounding Fast Food joints and Pizza places.

Cluster 4 – This has largely Break and enter crimes on homes in the western suburbs of Toronto. Parks seem to be closest to these crime scenes

Result Discussion

A majority of the crime in general is centred around downtown , also has the highest density of venues as well. Though Toronto itself is multicultural , Downtown Toronto is particularly an amalgam with tourists, office spaces and some residents living in that area.
If you are new resident and looking out for a house then that is not the place to look for definitely. Should be an easy guess for the police to plan well around that area too.
If you are a business owner then Down Town Toronto might be offering a large foot fall so knowing the times

when crime is most likely to occur [Fridays and Saturdays when Robberies & break Enters are more likely] & dealing with the issues in a safer way might still be a workable option.

In the north western & eastern areas , homes near a park seem more attractive for break and enter crime. Especially around the neighbourhoods of West Humber-Clairville, Scarborough Village.

While most of the cluster and crimes seem to be around the city like a u chain, The safest places seem to in the central parts of the city around York Mills, Mt Pleasant, Bay view village etc

Conclusion

In this exploration, I analysed the crime in Toronto with the lens of what premises and what kind of surrounding venues impact the nature of the crime. With an unsupervised clustering algorithm, patterns and relationships between the type of crime and the surrounding areas could be identified.

This hopefully would help choose a place to live or to set shop and would give a better perspective on how neighbourhoods and locations in that neighbourhood impact the local lives.

Further next steps

This was a very interesting analysis and can be extended with more attributes from the Toronto Open data portal.

Simple addition of demographics data to the existing Neighbourhood data and then clustering will help us explore more use cases like do senior resident areas have different crime vs areas with working parents.

Economic and Well being indicators can also be added to further analyse and understand the reasons for the crime and how it can be fundamentally addressed.