1. Introduction/Business problem

When you want to explore a big city there are 100s and even 1000s of places that you can look at on foursquare and get a list of a lot of places to go, to eat etc. But if you’re planning to take a trip to a big city and you look up venues that might be attractive to travellers, that same venue at that very day might also be riddled with car thefts, assaults, and even in some cases homicides. So, my idea for this project is to scrape the foursquare website for top venues in the city of Toronto and analyse all the criminal activities that take place in those venues on a daily, weekly and monthly basis. And make a list of neighbourhoods that you should probably want to avoid if you are planning a big trip, because being mugged and having your wallet or car stolen is the last you want on your trip.

To do that I will be looking up for Major crime indicators in the city of Toronto and give the user the list of neighbourhoods he/she might want to avoid in Toronto.

I will then use Foursquare API to look for places/venues people find interesting so that the user can have the best experience possible of the city of Toronto and at the same time avoid the places in the city where most crimes occur.

2. Data Acquisition and cleaning

2.1 Data Acquisition

For this project I will be using Data from 3 sources:

1). For my crime dataset I will be using the data published by Toronto police and I will be specifically looking for major crime indicators.

Here is a link to the crime dataset that i will be using (<http://data.torontopolice.on.ca/datasets/mci-2014-to-2018>.)

The crime dataset originally contained the following columns:

* **Index\_**
* **event\_unique\_id**
* **occurrencedate**
* **reporteddate**
* **premisetype**
* **ucr\_code**
* **ucr\_ext**
* **offence**
* **reportedyear**
* **reportedmonth**
* **occurrencedayofyear**
* **occurrencedayofweek**
* **occurrencehour**
* **MCI**
* **Division**
* **Hood\_ID**
* **Neighbourhood**
* **Long**
* **Lat**

2).The second source will be the Wikipedia page from which I will gather the information about the different boroughs of Canada (more specifically Toronto, I couldn’t find one for Toronto that’s why I used the Canada’s borough list)

Here is the link to the Wikipedia page (<https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M>)

This dataset originally contained the following columns:

* **Postalcode**
* **Borough**
* **Neighborhood**

However I will also need to add Latitude and Longitude values in my second data frame as well and I will do so with the help of Foursquare API.

2.2 Data Cleaning

For the crime dataset I had to drop most of the columns which were of no use for the analysis, for example (Index\_ , event\_unique\_id , occurrencedate, reporteddate, premisetype, ucr\_code, ucr\_ext, reportedyear,reportedmonth, occurrencedayofyear, occurrencedayofweek, occurrencehour, Hood\_ID)(see Fig 2.1)

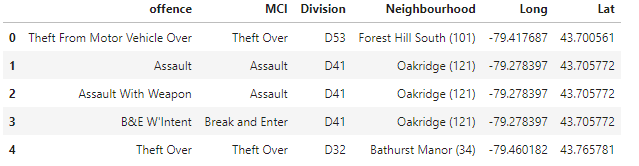


Fig 2.1 Toronto crime data after pre-processing.

The second data is scraped from a Wikipedia page using the Beautiful Soup library in python. Using this library we can extract the data in the tabular format as shown in the website. After the web scraping, string manipulation is required to get the names of the boroughs in the correct form (see fig 2.2).

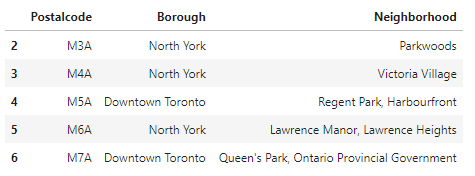
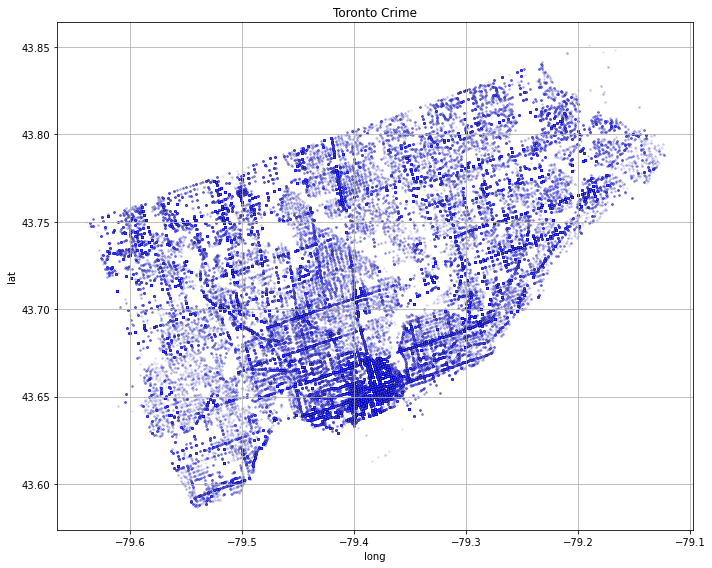


Fig 2.2 List of Canada Boroughs

3. Methodology

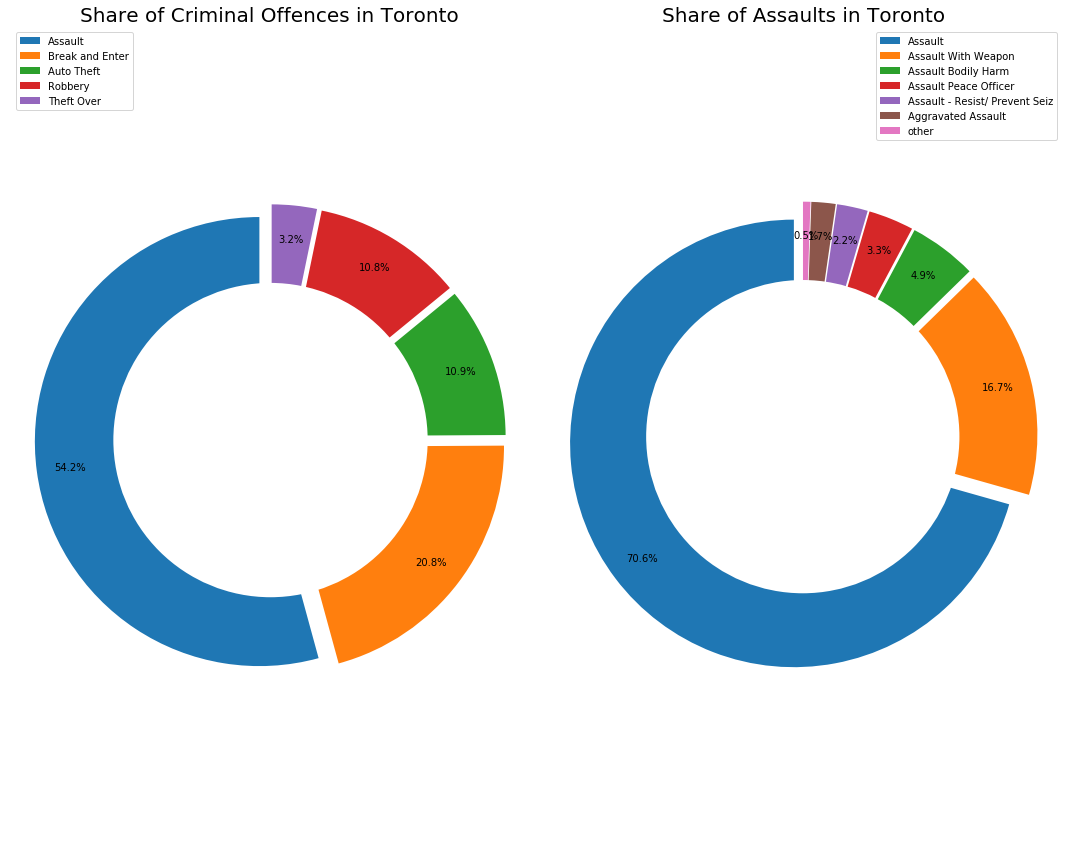
3.1 Exploratory data analysis

I wanted to get a general idea of the crimes in other areas close to Toronto before I wanted to explore the crimes in Toronto and as expected majority of the crimes in this map happen in Toronto as you can clearly see in this map, the south side (downtown) of Toronto .

I thought it would be effective to find places with high criminal density, the neighbourhoods around also exhibit high criminal activity and there are also other spots suggesting criminal neighbourhoods can be clustered. Contrary, the white spots are green - park - areas. You can also infer that crime appears mostly alongside the major roads. Technically, you can observe every major street, avenue, and road in the City. Additionally, you can clearly see, the Young Street heading North from Downtown.

3.2 Visualising Crime

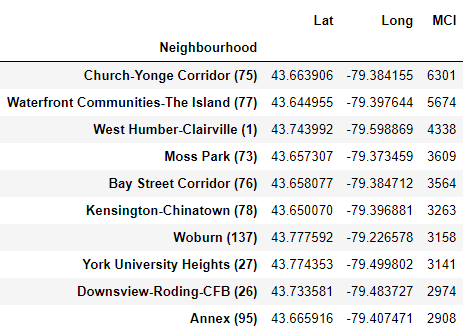
Now that we have a general idea of where most crimes happen in Toronto I thought it’d be nice to look at actually what kind of crimes occur the most , so for that I had to make a separate table out of my crime dataset that originally had over 12 columns.



As we can see from the pie charts that out of all the crimes that take place in the city of Toronto 54.2% percent of them are Assaults! , so in my crime dataset the offence assault had many different possible entries, so I wanted to know out of all the assaults that happen in Toronto what other kinds of assaults also take place, I found out that out of all the Assaults that happen 70.6% of assaults were regular assaults ( which can be anywhere from fighting to mugging) but 16.7 % of assaults actually happen with a weapon! A weapon doesn’t necessarily mean a gun, it can be anything from a gun to a knife or anything that can-do bodily harm to another person.

3.3 Boroughs that one should avoid

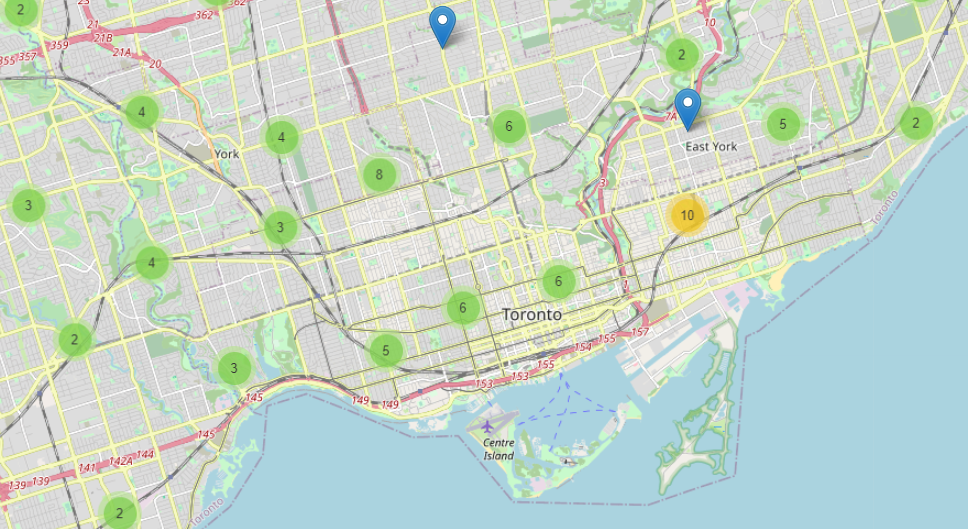
Now I wanted to cluster neighbourhoods and count MCI (major crime indicators) for each neighbourhood so that I can later plot that on a map and this is what I found,



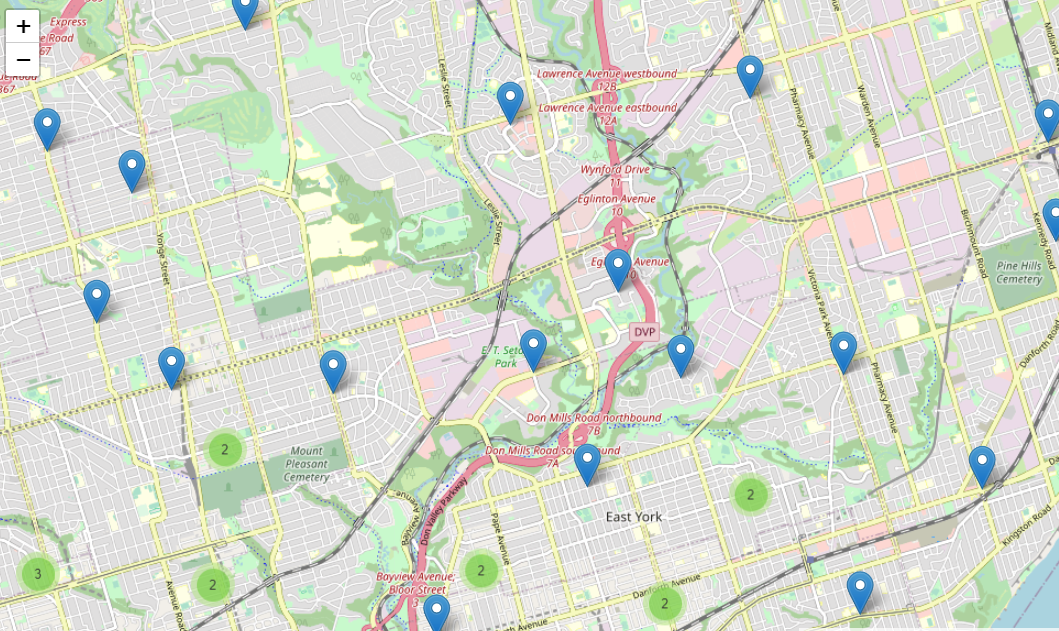
As we can see that Church-Yonge Corridor has an MCI of 6301, we’ll stay clear of that area.

Now that we know where most crimes take place lets get a better of that on a map to give to the user.

3.4 Crime Map of Toronto



All these clusters consist of crimes that actually take place in Toronto, unfortunately we can’t zoom in to see more areas further but that can be done easily on the notebook that I will also be providing.



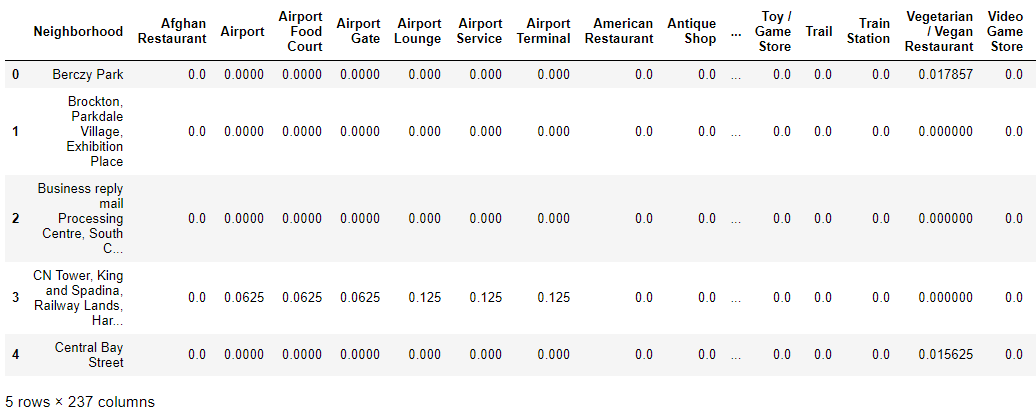
That should give user the general idea of what the map is going to look like after zooming in further.

3.5 Modelling

Using the final dataset containing the neighbourhoods in Toronto along with that neighbourhood’s latitude and longitude Values, we can find all the venues within a 500-meter radius of each neighbourhood by connecting to the Foursquare API. This returns a json file containing all the venues in each neighbourhood which is converted to a pandas data frame. This data frame contains all the venues along with their coordinates and the which category they fall into.

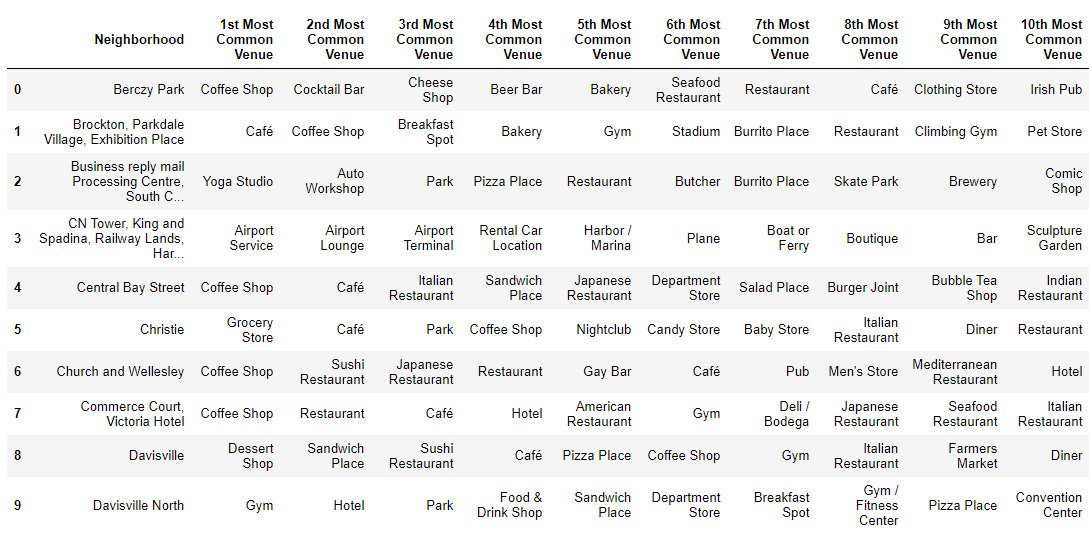


After One hot encoding was performed on this data by which categorical variables are converted into a form that could be provided to ML algos to do a better job of predicting Venues. The data is then grouped by each neighbourhood that we see here and the mean of those neighbourhoods are calculated thus giving us a top 10 common venues that people visit in Toronto.



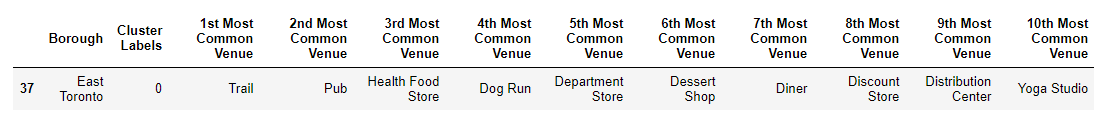
4. Results

After running the K-means clustering we can access each cluster created to see which neighbourhoods were assigned the 1’st to 10th most common venues.



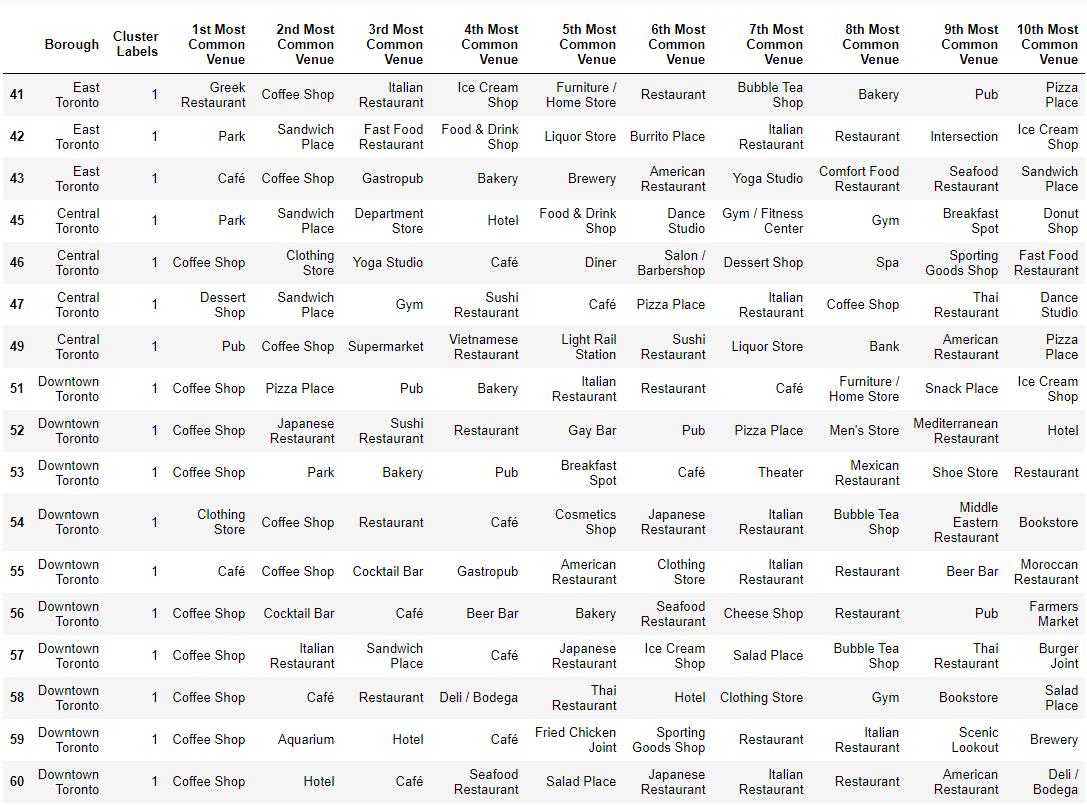
And now we will look for all the cluster labels from 0-4 to see what the most common venue is for that cluster label making it easier for the user to later look at what Cluster label he/she wants to go visit and then get an idea of actually where to go.

Let’s Look at common venues for cluster label 0 first.



This cluster is probably the smallest cluster out of all the other clusters, this is because of the unique venues in each of the neighbourhoods, hence they couldn't be clustered into similar neighbourhoods.

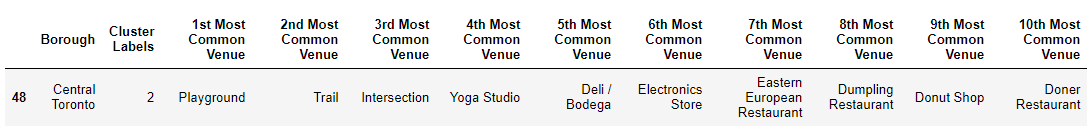
Now lets look at common venues in cluster label 1.



The cluster one is the biggest cluster with 9 of the 15 neighbourhoods in Toronto. Upon closely examining these neighbourhoods we can see that the most common venues in these neighbourhoods are Restaurants, Cafe, Supermarkets, and stores.

Looking into the neighbourhoods in the second, third and fourth clusters, we can see these clusters have only one neighbourhood in each. This is because of the unique venues in each of the neighbourhoods, hence they couldn't be clustered into similar neighbourhoods.

The cluster 2 has one neighbourhood which consists of venues such as Playgrounds, yoga studio, Bodegas, Eastern European Restaurant, etc.



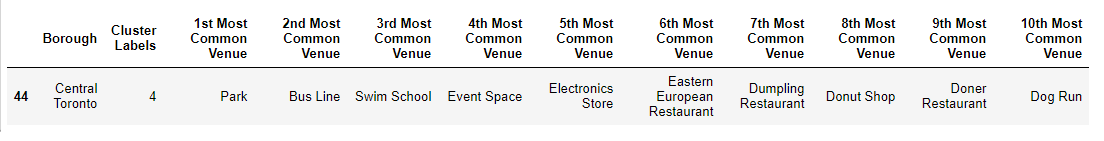
Cluster 2

The cluster 3 has one neighbourhood which consists of venues such as park, cycle studio, donut shop, etc.



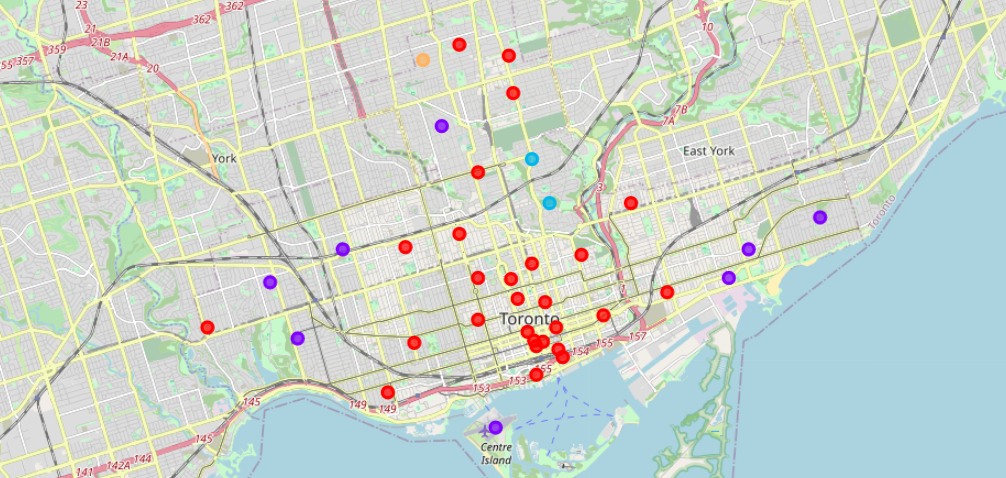
Cluster 3

The cluster 4 has one neighbourhood which consists of venues such as swim school, Dumping Restaurant, etc.



Cluster 4

Now that we have successfully clustered venues together lets plot them on a map for the user to see and get a better idea of the most common venues in the city of Toronto.



Visualising the clustered neighbourhoods on a map using the folium library.

Each cluster is colour coded for the ease of presentation; we can see that majority of the neighbourhood falls in the red cluster which is the first cluster(cluster label 0 ). Three neighbourhoods have their own cluster (Blue, Purple and orange), these are clusters two three and fourth.

5. Discussion

The aim of this project is to help people who are looking to have a safe trip to Toronto, an expat can choose a neighbourhood to visit by referring to the common venues in that neighbourhood. For example if a person is looking to play around with their kids and just want to have sort of a picnic he can look up the common venues in cluster 3 which shows that the most common venue in that cluster is actually the park. If a person is looking for a neighbourhood with stores and restaurants in a close proximity then the neighbourhoods in the first cluster is suitable. For a family I feel that the neighbourhoods in Cluster 4 are more suitable dues to the common venues in that cluster, these neighbourhoods have common venues such as Parks, Gym/Fitness centres, Bus Stops, Restaurants, Electronics Stores and Soccer fields which is ideal for a family. The choices of neighbourhoods may vary from person to person.

6.Conclusion

This project helps a person get a better understanding of the neighbourhoods with respect to the most common venues in that neighbourhood. I’ve just taken safety as a primary concern to shortlist the safest borough of Toronto with the lowest crimes.