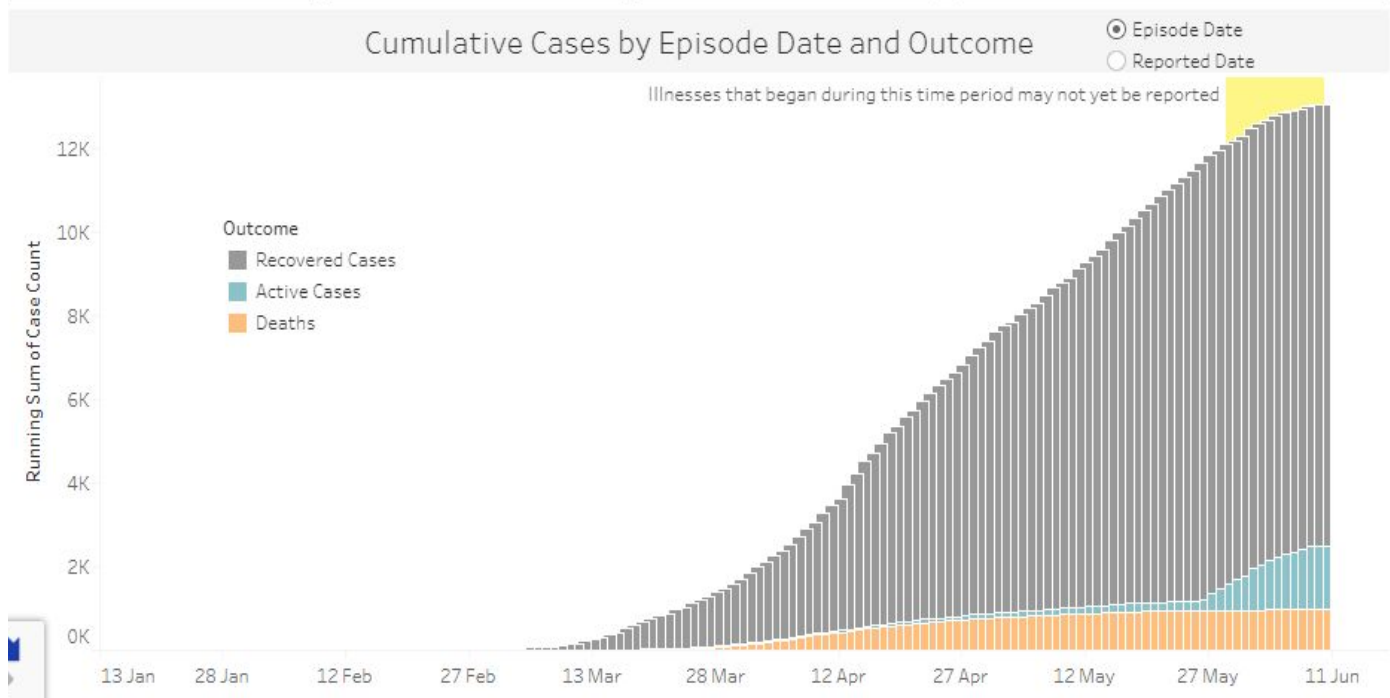


# Toronto Neighborhoods - Safe or not? (as of June 11, 2020)

By Aditya Kolluru

Cases*	Recovered Cases	Deaths	Cumulative Institutional Outbreaks +
13,063	10,558	965	155



## **Introduction:**

The COVID-19 pandemic in Canada is part of the ongoing worldwide pandemic of coronavirus disease 2019 (COVID-19) caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). Canada declared States of emergency in all its provinces. Due to this the bars, restaurants, cinemas, playgrounds, parks and other businesses and places where there are large social gatherings have been ordered closed by provinces, territories, and municipalities across the country. Initially, some jurisdictions allowed restaurants or bars to stay open with reduced capacity and social distancing.

## **Business Problem:**

With easing social restrictions of the bars, restaurants, playgrounds and other places, people are worried how safe it is to go outside. Using data science methodology and machine learning techniques like clustering, this project answers the question,

1. Which neighborhood would pose the least risk if someone wants to go out to a restaurant or a sports field to play in Toronto?, although it is always recommended to stay home.

## **Data:**

To solve the problem, we will need the following data

1. Covid-19 cases from all the neighborhoods in Toronto obtained from [Toronto Covid-19 data](#).
2. List of all neighborhoods and postal codes in Toronto extracted from [Toronto Postal Codes](#) Wikipedia page by web-scraping techniques.
3. Latitude and Longitude coordinates obtained by using the Python Geocoder package.
4. Venue data for those neighborhoods by using Foursquare API.

## **Methodology:**

The goal of this project is to provide correct information to people on which neighborhoods are safe to explore and which are to be avoided, given the venue.

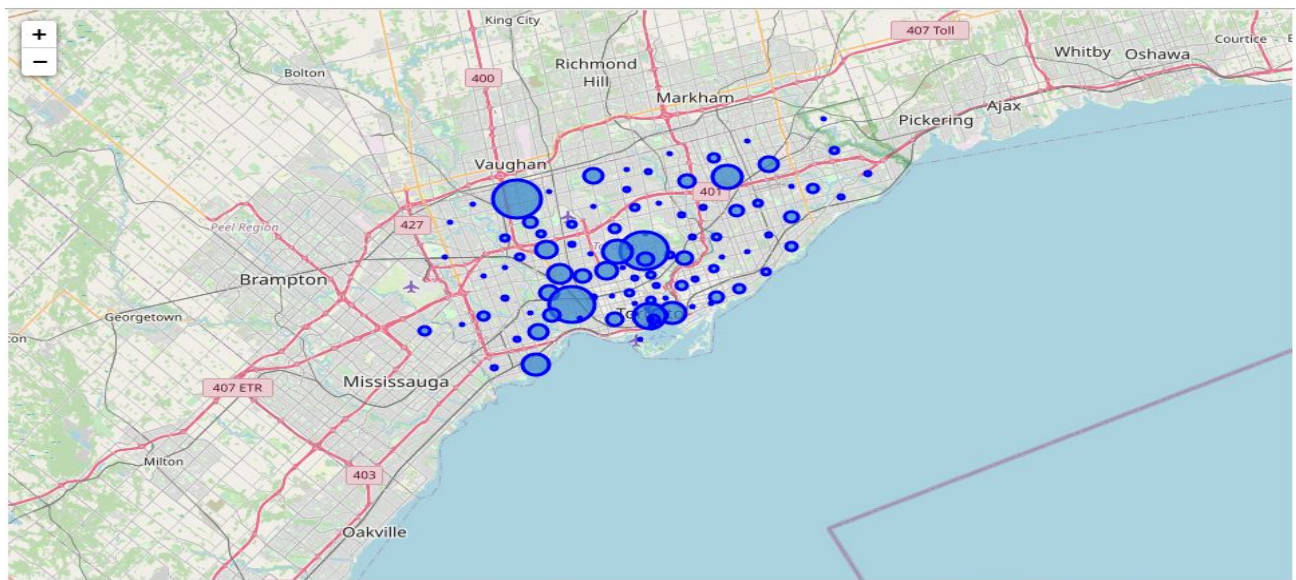
## **Analytic Approach:**

The city of Toronto has registered cases in 102 neighborhoods. To make the cluster analysis possible:

1. Initially, extract the Covid-19 case data from the [Toronto Covid-19 data](#) website and read it as an excel spreadsheet and convert it into a DataFrame. IBM watson has an in-built setting to build the data frame.
2. To get the postal codes of those neighborhoods, Python's BeautifulSoup technique is used for web-scraping from [Toronto Postal Codes](#), and the resultant data frame is joined with the original data frame to get the neighborhood names, cases and postal codes into one data frame.

	Neighborhood	Cases	Postal Code
0	Agincourt	57	M1B
1	Agincourt South-Malvern West	48	M1C
2	Alderwood	32	M1E
3	Annex	76	M1G
4	Banbury-Don Mills	29	M1H

- Foursquare API uses latitude and longitude to give the desired output. Python's geocoder package is used to get the data and now it is populated into a new data frame and this is used to visualize a map in folium using size of markers as the number of cases.



- The next step is to get the top 30 venues which are close to 500mts of these neighborhoods which will be done using Foursquare API.
- These top 30 venues are weighed using mean function and the dataframe with the 10 most common venues in each neighborhood is returned.

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Agincourt	Fast Food Restaurant	Dessert Shop	Falafel Restaurant	Event Space	Ethiopian Restaurant	Electronics Store	Eastern European Restaurant	Drugstore	Donut Shop	Dog Run
1	Agincourt South-Malvern West	Bar	Women's Store	Dim Sum Restaurant	Falafel Restaurant	Event Space	Ethiopian Restaurant	Electronics Store	Eastern European Restaurant	Drugstore	Donut Shop
2	Alderwood	Mexican Restaurant	Breakfast Spot	Electronics Store	Bank	Intersection	Medical Center	Rental Car Location	Women's Store	Ethiopian Restaurant	Eastern European Restaurant
3	Annex	Coffee Shop	Pharmacy	Korean Restaurant	Soccer Field	Dessert Shop	Event Space	Ethiopian Restaurant	Electronics Store	Eastern European Restaurant	Drugstore
4	Banbury-Don Mills	Athletics & Sports	Gas Station	Bakery	Thai Restaurant	Caribbean Restaurant	Bank	Fried Chicken Joint	Hakka Restaurant	Drugstore	Donut Shop

- K-cluster analysis is done on this dataframe which clusters each of the neighborhoods into 3 unique clusters.

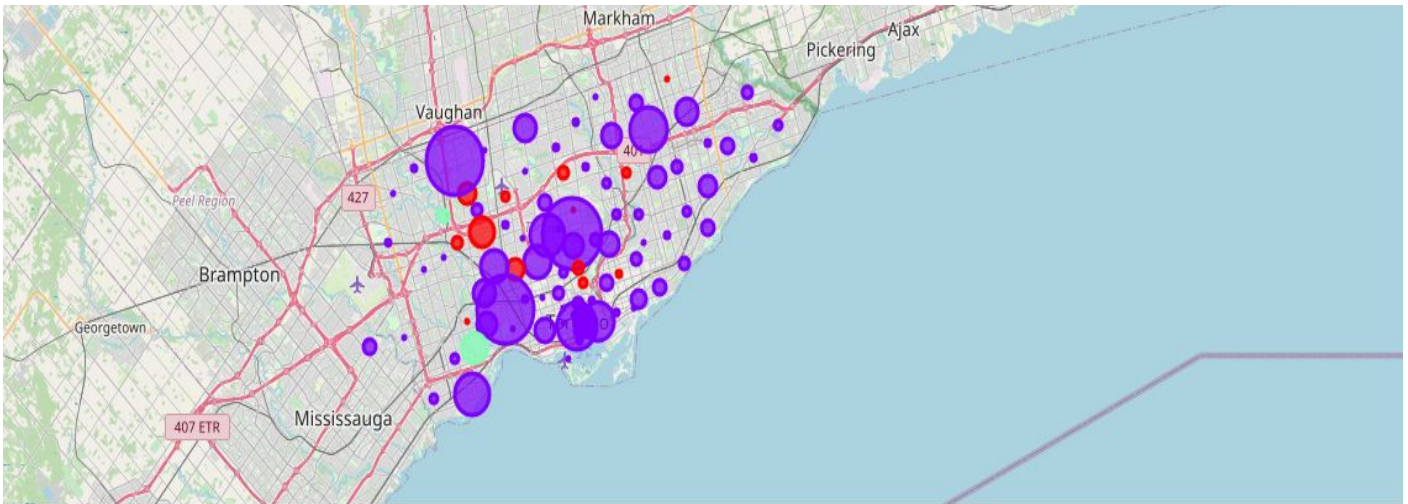
## Results:

The results give 5 clusters which are the following:

Cluster 0: Playground dominant neighborhoods with moderate shopping malls (red color)

Cluster 1: Restaurant dominant neighborhoods (purple color)

Cluster 2: Sports field dominant cluster (Mint green)



## Discussion and Observations:

From the above map, For a family/person willing to go to

1. Restaurants, coffee shops, the safest neighborhood would be the East York region, Weston area and more importantly they should avoid going to Downtown Toronto and Etobicoke region.
2. Parks, shopping malls, the safest neighborhoods are the North York and Scarborough region.
3. Play sports and convenience stores, the safest neighborhood is Mississauga and Malton region.

## Conclusion:

This project identifies the business problem, the data of Covid-19 cases in each neighborhood is extracted, cleaned and venues of each of these neighborhoods are prepared for segregating into 3 clusters by using machine learning techniques. These clusters provide us information about how safe each neighborhood is based on the venue selected.