Capstone Project: Battle of Neighbourhoods

COVID-19 High alert zones

1.Introduction:

Toronto is the capital city of Canada with a huge population. Currently, the world is facing a pandemic caused by the novel Corona Virus. The citizens are requested to stay indoors and the majorly busy venues like Restaurants, Cinema, Malls, etc. are to remain closed. Despite the restrictions from the Toronto Police Department, there have been breaches of rules in some areas where people are still walking on the streets without a necessary reason and are not practicing social distancing. This will lead to the massive spread of virus in the community.

1.1Problem description:

The Toronto Police Department has to manage the forces they have wisely in order to avoid the people from breaking the rules in the high alert zones. A clear analysis on the high alert zones will help the Department to assess how much Force has to be assigned in the area depending on how vast the area population is and how busy the area is.

1.2Target Audience:

The target audience here is the Toronto Police Department which is seeking for information on the areas in Toronto. This will help them distribute their resources effectively.

1.3Success criteria

The aim of this project is to give a clear picture to the Toronto Police Department on the zones that are highly prone to a huge spread of the virus based on 2 factors: Population in the Neighbourhood and the measure of how busy the area is.

2. Data Description:

The main aim of the project is to explore the neighbourhoods in the city of Toronto, hence Toronto neighbourhoods data is essential. The structured format of the data is not available on the internet, hence we need to scrap it through an existing Wikipedia page that has all the required information. The coordinates of the neighbourhoods is important in the data. The information like the most common venues, population, Borough are also required and should be present in the final report. The data needs to be cleaned up and should look like the following:

	PostalCode	Borough	Neighbourhood	Latitude	Longitude
0	M1B	Scarborough	Rouge, Malvern	43.8067	-79.1944
1	M1C	Scarborough	Highland Creek, Rouge Hill, Port Union	43.7845	-79.1605
2	M1E	Scarborough	Guildwood]], Morningside, West Hill	43.7636	-79.1887
3	M1G	Scarborough	Woburn	43.771	-79.2169
4	M1H	Scarborough	Cedarbrae	43.7731	-79.2395
5	M1J	Scarborough	Scarborough Village	43.7447	-79.2395
6	M1K	Scarborough	East Birchmount Park, Ionview, Kennedy Park	43.7279	-79.262
7	M1L	Scarborough	Clairlea, Golden Mile, Oakridge	43.6915	-79.2866
8	M1M	Scarborough	Cliffcrest, Cliffside, Scarborough Village West	43.7163	-79.2395
9	M1N	Scarborough	Birch Cliff, Cliffside West	43.6927	-79.2648

2.1 Data features:

We will be using the information provider, Foursquare.com, to explore the various venues in each neighbourhood. The information on how busy the neighbourhood is, is obtained by understanding the trends of theses venues in the respective neighbourhood. The information obtained are:

1. Neighbourhood
2. Neighbourhood Latitude
3. Neighbourhood Longitude

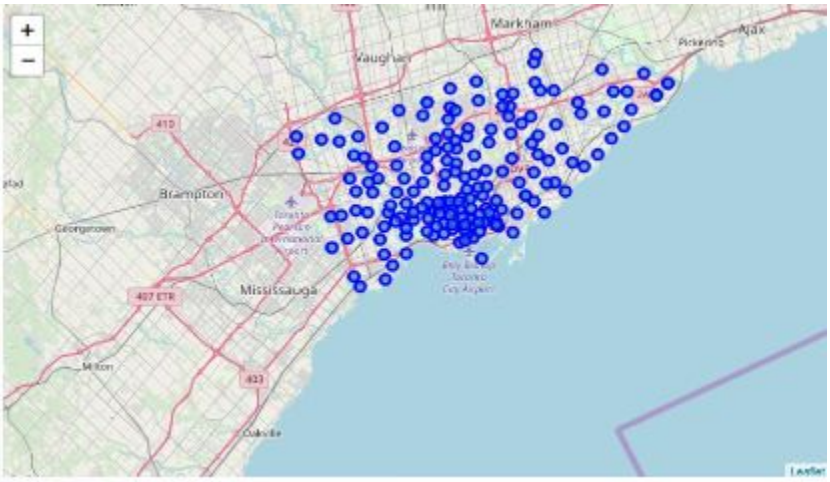
4. Venue Name
5. Venue Category
6. Venue Latitude
7. Venue Longitude

	Neighbourhood	Neighbourhood Latitude	Neighbourhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Rouge, Malvern	43.806686	-79.194353	Wendy's	43.807448	-79.199056	Fast Food Restaurant
1	Highland Creek, Rouge Hill, Port Union	43.784535	-79.160497	Royal Canadian Legion	43.782533	-79.163085	Bar
2	Highland Creek, Rouge Hill, Port Union	43.784535	-79.160497	Affordable Toronto Movers	43.787919	-79.162977	Moving Target
3	Guildwood]], Morningside, West Hill	43.763573	-79.188711	Swiss Chalet Rotisserie & Grill	43.767697	-79.189914	Pizza Place
4	Guildwood]], Morningside, West Hill	43.763573	-79.188711	G & G Electronics	43.765309	-79.191537	Electronics Store
5	Guildwood]], Morningside, West Hill	43.763573	-79.188711	Big Bite Burrito	43.766299	-79.190720	Mexican Restaurant
6	Guildwood]], Morningside, West Hill	43.763573	-79.188711	Enterprise Rent-A-Car	43.764042	-79.193371	Rental Car Location
7	Guildwood]], Morningside, West Hill	43.763573	-79.188711	Woburn Medical Centre	43.766631	-79.192286	Medical Center
8	Guildwood]], Morningside, West Hill	43.763573	-79.188711	Eggsmart	43.767800	-79.190466	Breakfast Spot
9	Woburn	43.770992	-79.216917	Starbucks	43.770037	-79.221156	Coffee Shop

2.2 Conclusion:

We will segment these neighbourhoods to understand the similarities between them and analyse further. With all these features, we will further discover useful features and come up with the list of high priority areas for the Toronto Police Department. For example, the area with high population and more public places will be more prone to the virus spread since more people are found in groups.

3. Methodology:

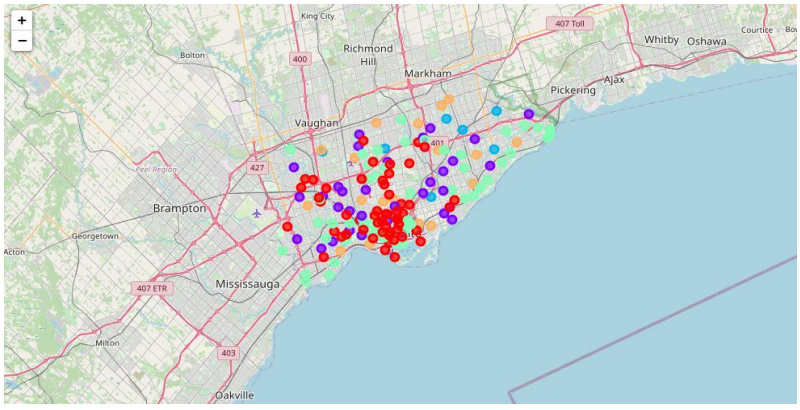


Data scrapping from the Wikipedia page that contains the up-to-date population statistics of Toronto neighborhoods has been used. This is critical to understand the population of each Toronto neighborhood which is one of the key elements in the neighborhood of choice in this project.

Furthermore, we need to know the coordinates and locations of this neighborhoods, and therefore the geocoder API has been used for achieving this objective. This is important so that we can input this information into the location information provider such as Foursquare.com to obtain venue information in these neighborhoods, and this is precisely what we have done for it in this project. We will also use machine learnings techniques such as the K-Means Clustering to segment and cluster these neighborhoods so that we can group them together to understand their similarities. This is critical as we need to help the Toronto Police Department identify the regions where the population is higher and where the crowd movement would possibly be high.

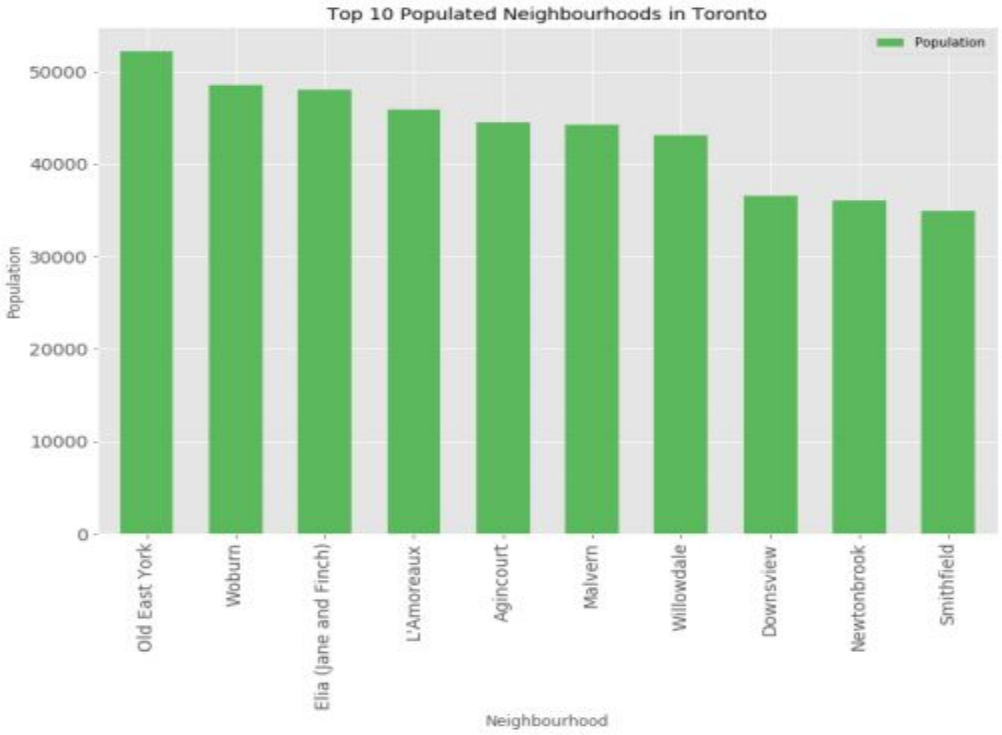
Finally, with all these methodologies, we will then be able to come up with a best analysis to the Toronto Police to their problem which is where the virus outspread is most likely to be higher based on neighborhoods similarities and high population.

4. Result



With K-Means clustering technique, the top 5 clusters of similar neighborhoods have been apparent in the result, see above. These clusters are group together based on the common nearby venues in each of the neighborhoods. This information is critical so that we can target on the cluster with high total score.

With bar chart visualization technique, we can easily tell what are the top population (i.e. higher number of residences) in the neighborhood cluster. The top 10 neighborhoods with highest number of populations are as follows.



With Foursquare.com API, we are also able to leverage on the data to find out the top common nearby venues and their categories in each of these neighborhoods. This is critical as we want to recommend a neighborhood whereby the venue score is high and then use this score further to get the total score.

	Neighbourhood	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
0	Agincourt	Coffee Shop	Yoga Studio	Fast Food Restaurant	Empanada Restaurant	Ethiopian Restaurant	Event Space	Exhibit	Falafel Restaurant
1	Alderwood	Pizza Place	Gym	Dance Studio	Pub	Coffee Shop	Donut Shop	Bank	Convenience Store
2	Alexandra Park	Bar	Café	Vegetarian / Vegan Restaurant	Restaurant	Coffee Shop	French Restaurant	Dessert Shop	Yoga Studio
3	Allenby	Coffee Shop	Sushi Restaurant	Gym	Italian Restaurant	Café	Fruit & Vegetable Store	Liquor Store	Gastropub
4	Amesbury	Bakery	Fast Food Restaurant	Park	Sandwich Place	Flea Market	Fish Market	Fish & Chips Shop	Filipino Restaurant

5. Discussion

Based on the results above, the top high priority zones are identified with respect to the total score generated. According to the total score, the top 5 high alert zones are: Old East York, Malvern, Woburn, Elia (Jane & Finch) and High Park North. These are the neighbourhoods that need high attention. The third cluster has the highest scores; hence it has to be prioritized.

The top neighbourhoods of each cluster could also be considered as virus hotspots, since they are the ones with high population and also more busy venues. The top neighbourhoods must be strictly under lockdown and the testing of the virus must happen in a large scale in these areas. The zones can be assigned as per the clusters formed: Red zone (cluster 2,3 and 5) ; Orange zone (cluster 1 and 4)

The Toronto Police Department can further put its requirements to the data science team to improve the security and safety of its citizens. The project can be further developed when a feature regarding the existing number of active cases in a neighbourhood is available. That will give a better picture of the scenario.

6. Conclusion

We have concluded that the third cluster is highly critical of experiencing a virus outspread. The top neighbourhoods in this cluster must be given more attention. The top populated neighbourhoods are found out so as to consider them as critical zones.

It is also recommended to re-run this data science program to get the updated result and use the result into consideration so as to perform a better analysis. This is critical not only to make sure that they got the updated result for better decision making, but also to make sure that they can re-validate the findings from this project. Finally, thank you for the opportunity in this project and we wish you the best success in your business.