# COVID-19 High alert zones

**An analysis of the neighbourhoods in Toronto to help Toronto Police Department fight against the virus outspread**

# INTRODUCTION

- **Location** : Toronto, Canada
- Toronto Police Department
- Strategic Plan and Overview
- **Problem** :  To identify the neighbourhoods in the city which are highly prone to have a virus outspread.
- Neighbourhood with high population and more busy venues
- **Success criteria**: The classification into priority zones should be advantageous and easy to analyse.
- It should have the capability to enable new information whenever available

# DATA DESCRIPTION

| | Neighbourhood | Population | Land Area | Density | Population % | Commuting | Latitude | Longitude |
|---|---|---|---|---|---|---|---|---|
| 1 | Agincourt | 44577 | 12.45 | 3580 | 4.6 | 11.1 | 43.785353 | -79.278549 |
| 2 | Alderwood | 11656 | 4.94 | 2360 | -4.0 | 8.8 | 43.601717 | -79.545232 |
| 3 | Alexandra Park | 4355 | 0.32 | 13,609 | 0.0 | 13.8 | 43.650758 | -79.404308 |
| 4 | Allenby | 2513 | 0.58 | 4333 | -1.0 | 5.2 | 43.711351 | -79.553424 |
| 5 | Amesbury | 17318 | 3.51 | 4,934 | 1.1 | 16.4 | 43.706162 | -79.483492 |
| 6 | Armour Heights | 4384 | 2.29 | 1914 | 2.0 | 10.8 | 43.743944 | -79.430851 |
| 7 | Banbury | 6641 | 2.72 | 2442 | 5.0 | 6.1 | 43.742796 | -79.369957 |
| 8 | Bathurst Manor | 14945 | 4.69 | 3187 | 12.3 | 13.4 | 43.763893 | -79.456367 |
| 9 | Bay Street Corridor | 4787 | 0.11 | 43,518 | 3.0 | 17.1 | 43.664457 | -79.387189 |
| 10 | Bayview Village | 12280 | 4.14 | 2,966 | 41.6 | 14.4 | 43.769197 | -79.376662 |

## Data sources :
• Wikipedia pages
• Foursquare.com

## Data Requirements:
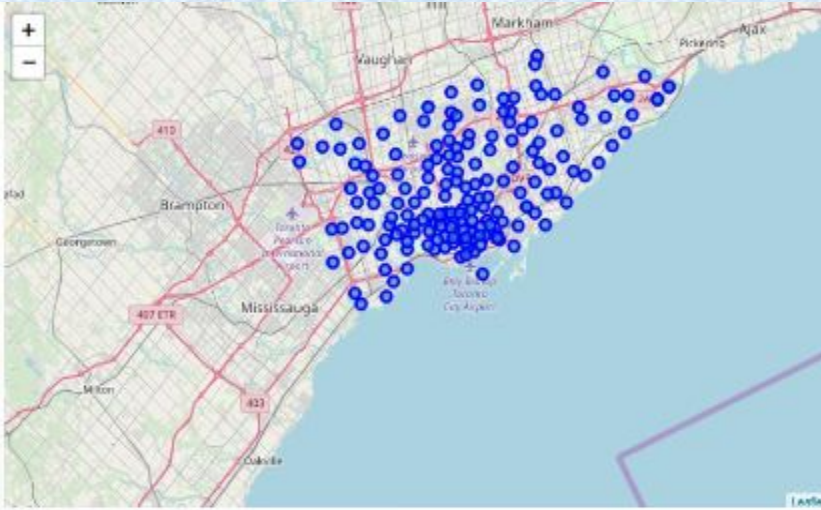•   Neighbourhood information ( Name, coordinates, population)
• Venue information ( name, coordinates)

## Data preprocessing:
• The Data needs to be cleaned
• The data must be structured to the suitable format

# METHODOLOGY



**Data scrapping technique:**
• To scap the neighbourhoods data from the wikipedia page
• Structure the scrapped data

**Foursquare.com**
• Retrieving venue information from the API

**K-means Clustering :**
• Machine learning classification technique
• segment and cluster similar neighbourhoods
• grouping based on most common venues

**Scoring weightage:**
• 50% each for population and the common venues in the neighbourhood

# RESULT

**Clustering the neighbourhoods:**
• Clustering the neighbourhoods based on the most common venues

**Total score:**
• A score is calculated for each neighbourhood based on the venue score ( based on the top 10 common venues) and the normalised population score.
• Higher the score, higher the virus outspread threat



**High priority cluster:**
• The clusters obtained as the result help in indicating the priorities of the clusters
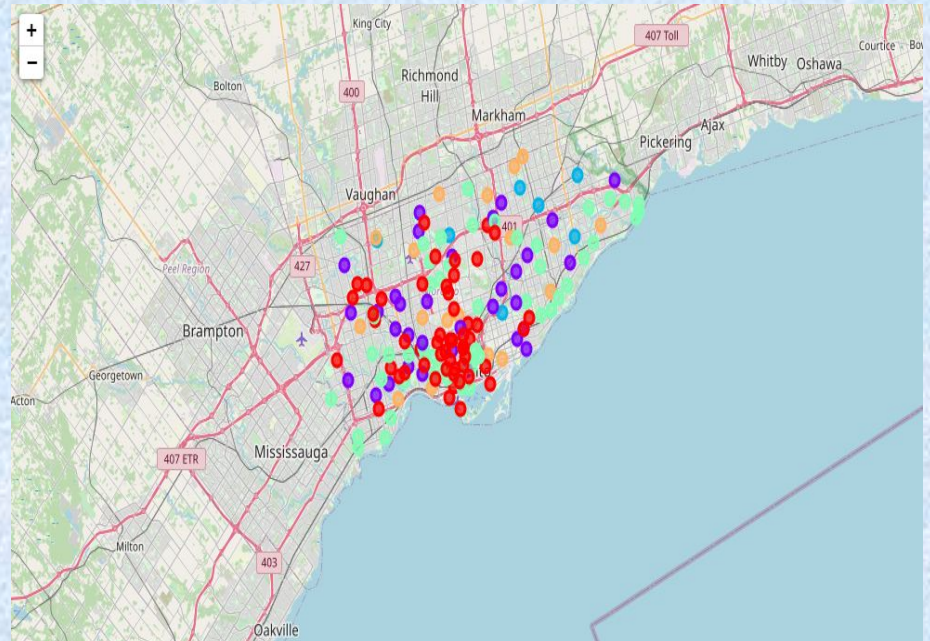
# DISCUSSION

| | Neighbourhood | Population | Land Area | Density | Population % | Commuting | Latitude | Longitude | Population Score | Venue Score | Total Score | Cluster Labels |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | Old East York | 52220 | 7.94 | 6577 | -4.6 | 22.0 | 43.699971 | -79.332520 | 2.161626 | 0.7 | 1.430813 | 2 |
| 5 | Malvern | 44324 | 8.86 | 5003 | 1.5 | 13.6 | 43.809196 | -79.221701 | 1.834774 | 0.7 | 1.267387 | 2 |
| 1 | Woburn | 48507 | 13.34 | 3636 | -1.5 | 13.3 | 43.759824 | -79.225291 | 2.007928 | 0.0 | 1.003964 | 2 |
| 2 | Elia (Jane and Finch) | 48003 | 7.66 | 6267 | -10.0 | 13.0 | 43.757253 | -79.517697 | 1.987065 | 0.0 | 0.993533 | 2 |
| 27 | High Park North | 22746 | 2.18 | 10,434 | -1.6 | 27.5 | 43.657383 | -79.470961 | 0.941562 | 1.0 | 0.970781 | 1 |

**Result review:**
• the top high priority zones are identified with respect to the total score generated. According to the total score, the top 5 high priority zones are: Old East York, Malvern, Woburn, Elia(Jane & Finch) and High Park North.
• The third cluster has the highest scores, hence it has to be prioritised.

**Other observations:**
• 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 zones can be assigned as per the clusters formed: Red zone: Cluster 2,3 and 5 Orange zone: Cluster 1 and 4

# CONCLUSION

**Analysis summary:**

- the top 5 high priority zones are: Old East York, Malvern, Woburn, Elia(Jane & Finch) and High Park North
- Red Zone: Cluster 2,3 and 5
- Orange Zone: Cluster 1 and 4

**Additional note:**

- Rerun of program is encouraged to get the latest information.
- Updated or additional information on the infected people in each neighbourhood will help analyze better.

**Acknowledgement:**

- Appreciation notes from the Data Science team for this project