

TORONTO COVID CASES BY NEIGHBORHOOD

Julio Flores

August 2, 2021

1. Introduction

1.1 Background

During the COVID pandemic, all governments implemented several rules and restrictions in order to mitigate the outbreaks of the disease, these restrictions were usually defined based on the type of business and industry of the venues.

1.2 Problem

Data can contribute to rate the efficacy of the restrictions put in place by the government in the city of Toronto, by analyzing the data of the total COVID cases and the type of venues, all segmented by Postal Code and neighborhood.

1.3 Interest

The target audience would be all the involved government dependencies (Commerce, Health, City Planning) which could determine if the imposed restrictions did work and if any of these should be kept in place. Also, to define a workbook for future pandemics.

2. Data acquisition and cleaning

2.1 Data Sources

Toronto neighborhood data

The neighborhood data was extracted from the Wikipedia page [https://en.wikipedia.org/wiki/List of postal codes of Canada: M](https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M)

COVID-19 Cases in Toronto

Dataset provided by the Toronto city government in their Open Data portal: <https://open.toronto.ca/dataset/covid-19-cases-in-toronto/>

Canada demographic data by Postal Code (2016 Census)

Demographics data from 2016 census (latest data available) from the Canada government statistics page: <https://www12.statcan.gc.ca/census-recensement/2016/dp-pd/hlt-fst/pd-pl/Tables/CompFile.cfm?Lang=Eng&T=1201&OFT=FULLCSV>

Foursquare API

Foursquare API to gather the data of venues by neighborhood and define the most popular categories in each neighborhood <https://api.foursquare.com/v2/venues/explore>

2.1 Data Cleaning

All the collected data was cleaned, grouped, and combined into one final table, with twelve features:

- Postal Code
- Borough
- Neighborhood
- Latitude
- Longitude
- Total COVID cases
- Population
- Cases per 1,000
- Cluster
- 1st most common venue
- 2nd most common venue
- 3rd most common venue

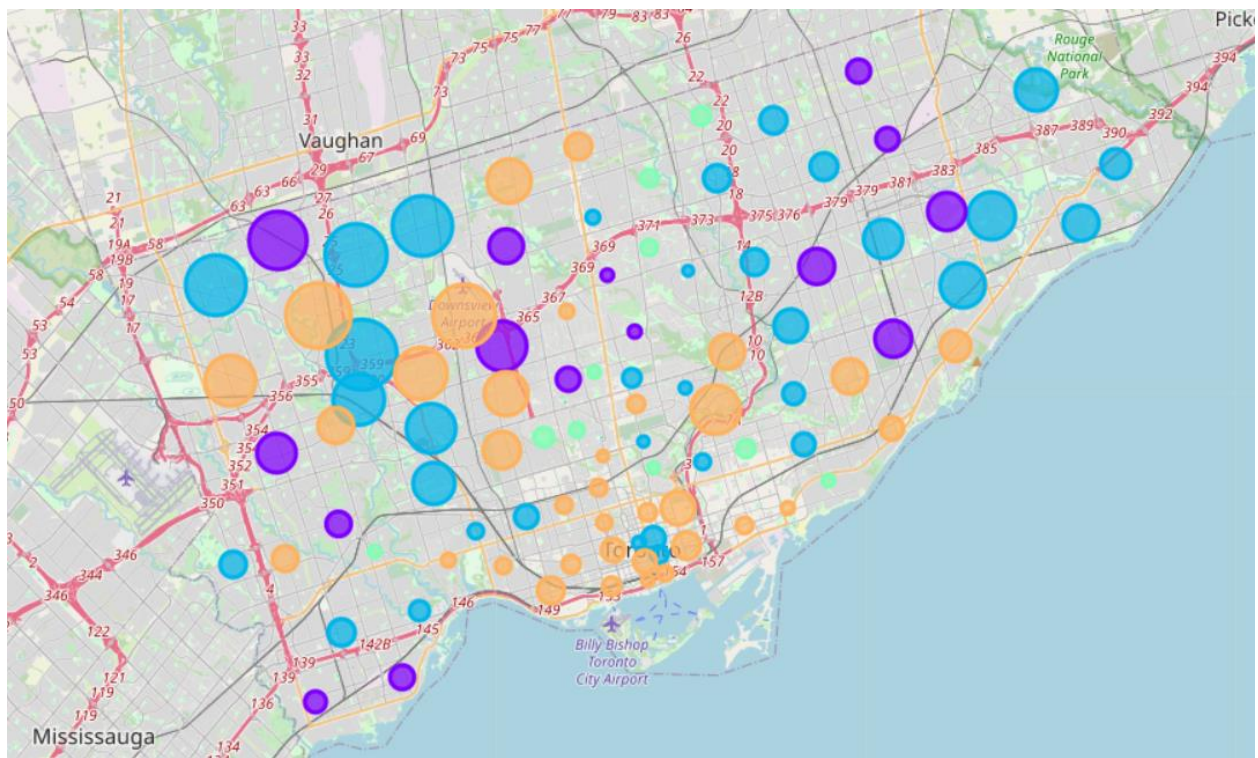
3. Methodology

To determine the effects of the neighborhood venue types in the spread of the COVID-19, we clustered the Toronto's Postal Codes with the k-Means methodology, once clustered, we tried to predict the COVID cases per 100 and confirm if there is a relation with the COVID transmission and the venue type in each Postal Code.

After experimenting with several numbers of clusters (k), we determined that four (4) was the most beneficial number for this study. Once the methodology was applied and the different clusters evaluated, we determine the following definition for each cluster:

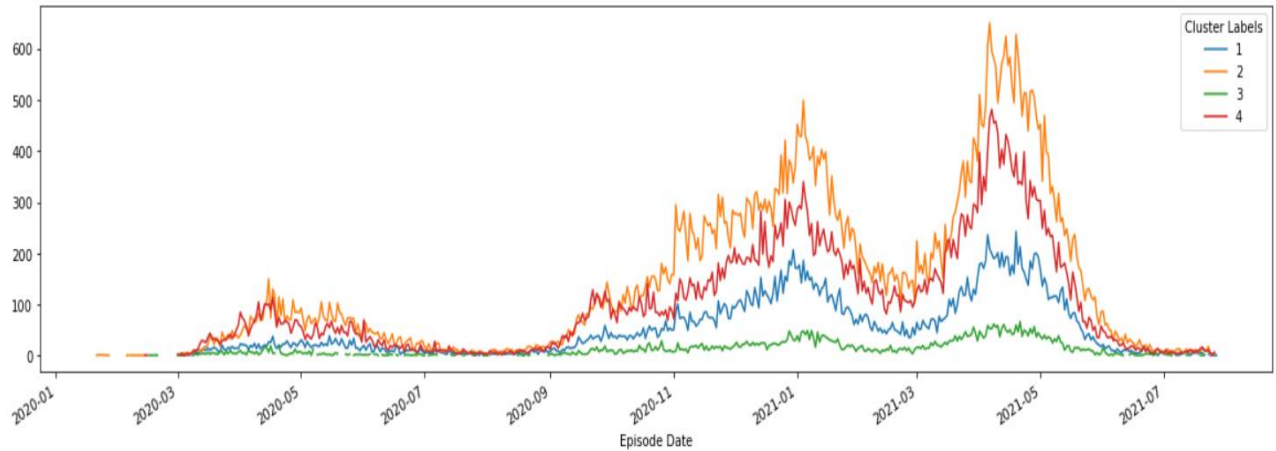
- Cluster 1: Shop & Service venues
- Cluster 2: A combination of Food and Shop & Service venues
- Cluster 3: Outdoors & Recreation venues
- Cluster 4: Food venues

Cases per capita in each neighborhood and cluster:



Cluster 1 – Purple, Cluster 2 – Blue, Cluster 3 – Green, Cluster 4 - Orange

Total cases per Cluster and date



4. Conclusion

By looking at the data, we can infer that the greater number of COVID cases per capita occurred in the neighborhoods which main category was related with Shop & Service. However, the greater number of total cases, occurred in the neighborhoods which Food venues are the first or second category.

As expected, the smaller share per capita occurred in the neighborhoods which main category is Outdoors & Recreation.

Based on this data, it seems that better rules and restrictions should have been applied to the Food