

Final Project: Battle of Neighborhoods

Presentation for the
Coursera-IBM Data Science
certificate

Introduction

The following study, although modest, pretends to explore the spread of COVID19 on different communities, hoping to find patterns or clues regarding its behaviour. While the most natural assumption would be that the majority of the infections occur on densely populated areas, we want to assess the nature of the neighborhoods instead, trying to conclude whether the dominant presence of social hubs, commercial districts or open areas conveys a significant difference in numbers

Methodology

We will study data from two cities, New York in USA and Toronto in Canada. We will extend the analysis of Manhattan from the course to the rest of the city and compare the results with the ones from Toronto.

More explicitly, after obtaining the neighborhood list of both cities from an external database, we access access Foursquare to obtain information of the venues of each, so we can later find similar zones by clustering algorithms. At this point we will discuss differences and similarities between both cities. Later, using open access data from their respective official city portals, we will try to relate the incidence of COVID with the neighborhood clusters. Finally, we will compare the results on both financial capitals.

Data

For both cities we require geographical and COVID data.

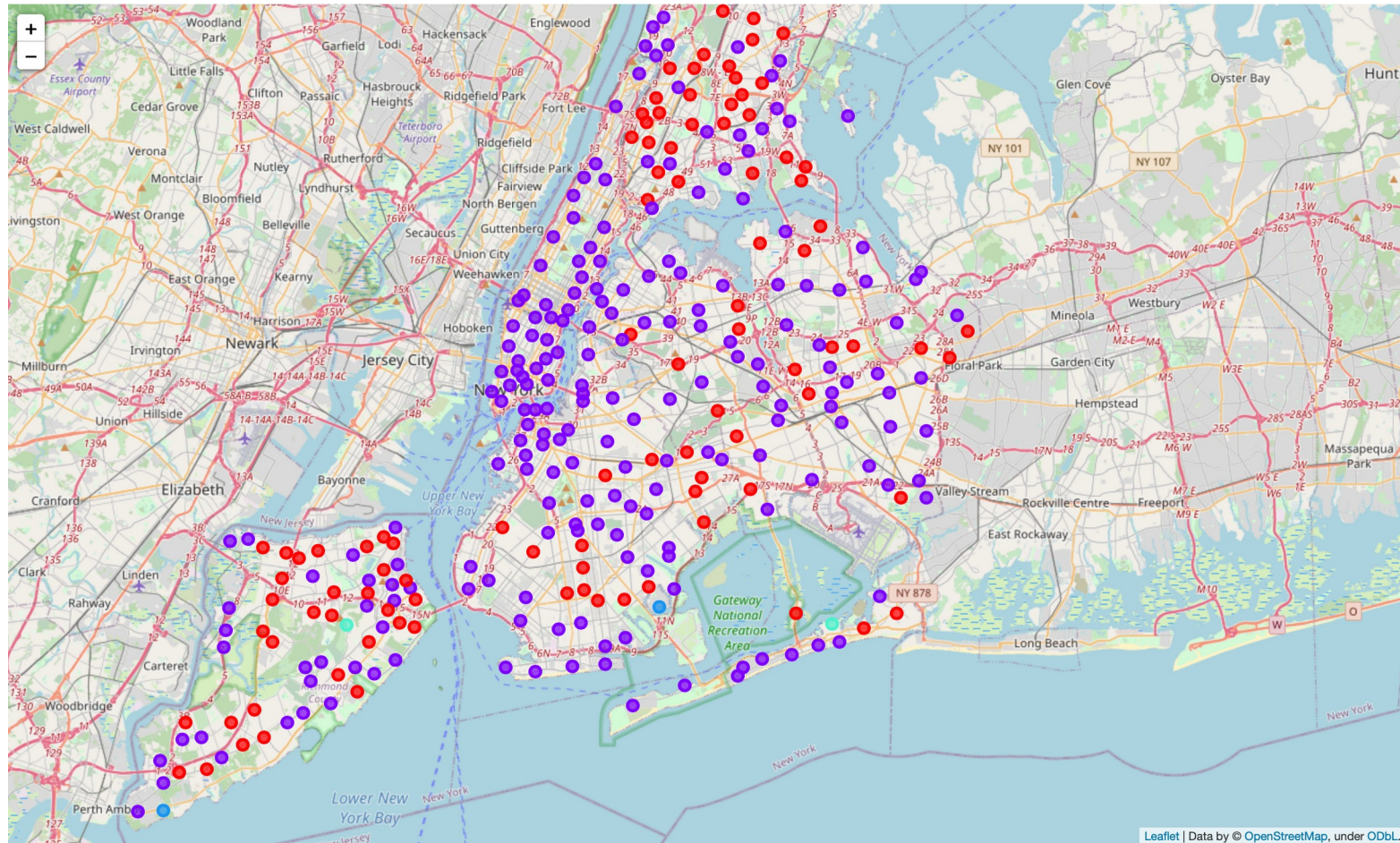
- For NY: the geographical data comes from the json file provided on the course, the venue information from Foursquare. Covid data comes from the official city website:
<https://www1.nyc.gov/site/doh/covid/covid-19-data-boroughs.page>
- For Toronto: the geographical data comes from an html file hosted on wikipedia and a list provided from the course, the venue information from Foursquare. Covid data comes from the city portal:
<https://www.toronto.ca/home/covid-19/covid-19-latest-city-of-toronto-news/covid-19-status-of-cases-in-toronto/>

Results

NY - We intentionally chose 4 clusters (see Discussion below). We mainly obtain two very big and two very small groups

- The first group (red) notoriously consists on bars, restaurants, banks and metro stations; i.e. zones with a lot of people and movement. It doesn't seem to have any geographical tendency, except from Manhattan, in which most the group represents the northern part. T
- The second group (purple) contains bars and restaurants as well, but also shows an important amount of stores and health and fitness venues; although crowded, it represent less activity than group 1
- The last two groups consist on a pair of neighborhoods each, however, strikingly uniform and both seemingly more into the 'open air'

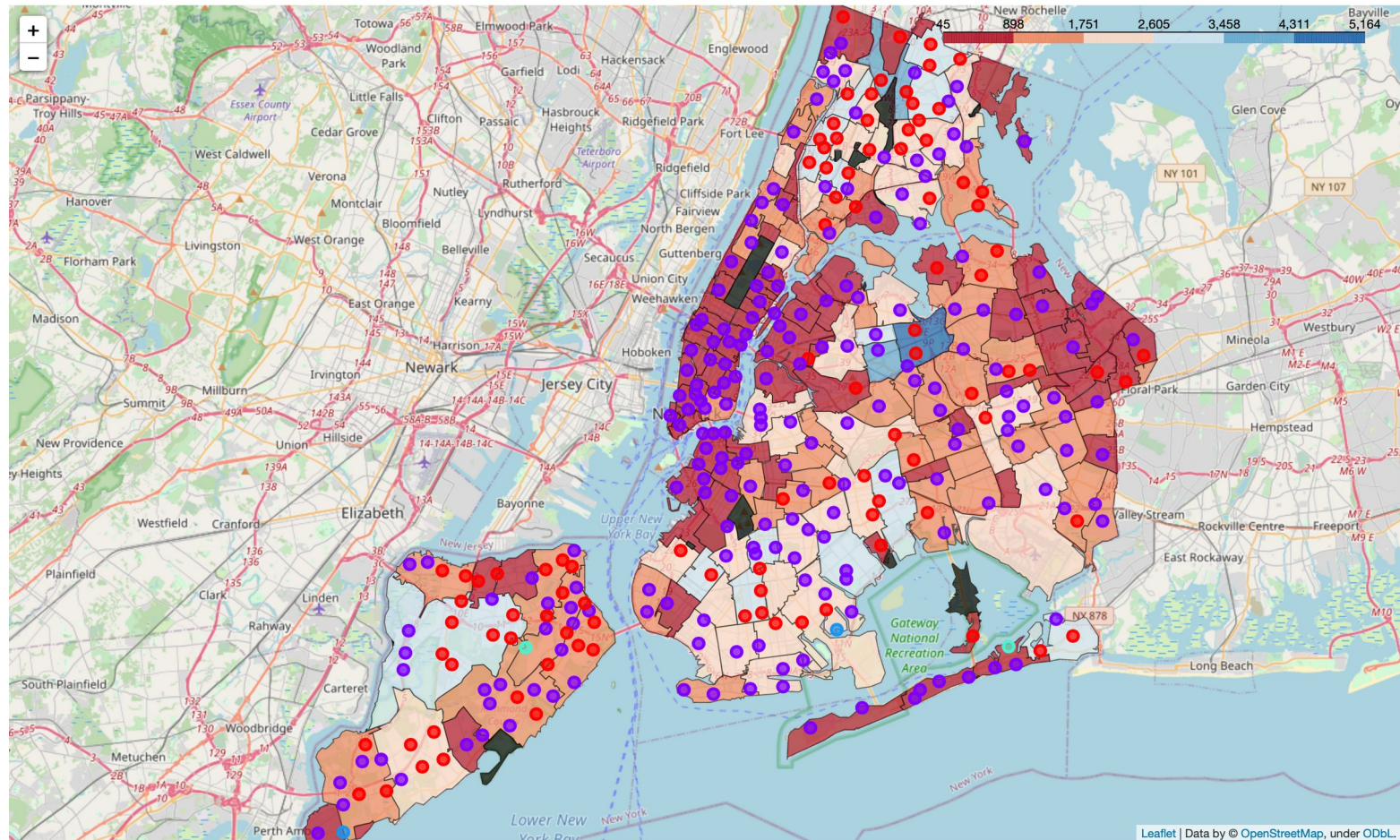
NY Clusters



NY – The heat map portrays the covid data. We ‘draw’ the clustered neighborhoods on top of it

- Interestingly, group 2 (purple) fits well to the zones on bright red, corresponding to the lower case count.
- Following is group 1 (red), mostly appearing on the orange-light red regions standing for low to mid case count.
- It doesn't seem to be a particular pattern for the regions with high case count numbers, nor a specific results for groups 3 and 4.

NY Covid v/s Clusters

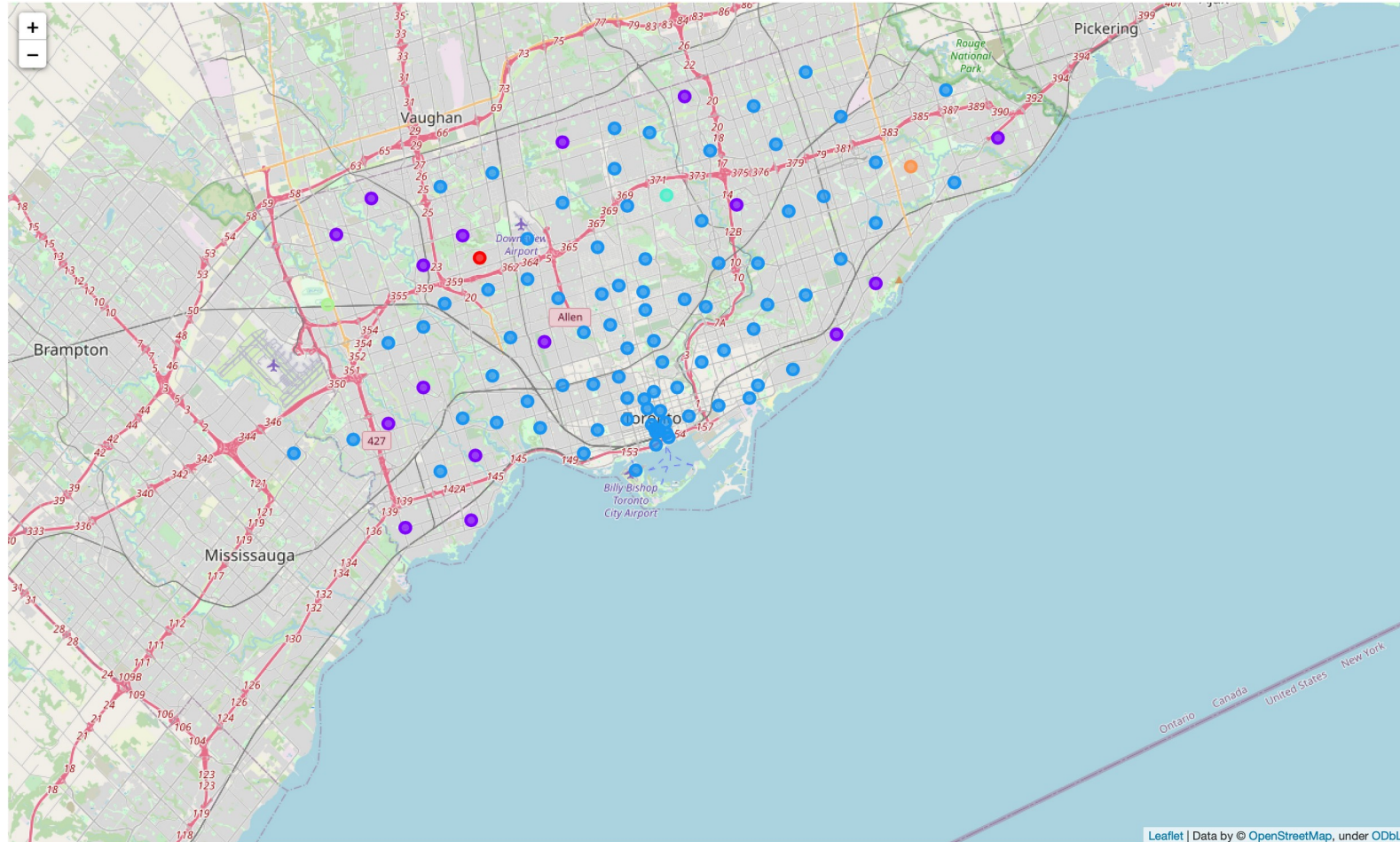


Results

Toronto - Here we try 5 clusters. , a mid sized group spread mostly outside of it and 3 individual clusters also at the outskirts:

- We get a highly dominant group (blue) concentrating towards the heart of the city. It consists mostly, as in NY, on neighborhoods with restaurants, bars and banks.
- The second, mid sized group (purple) is mostly spread outside the center and consists, also as in NY, predominantly on areas with shopping malls, parks and health and fitness venues.
- The second group (purple) contains bars and restaurants as well, but also shows an important amount of stores and health and fitness venues; although crowded, it represent less activity than group 1
- We also have 3 individual, isolated clusters at the outskirts. Noticeable, 2 of them include a Zoo and are close to an airport.

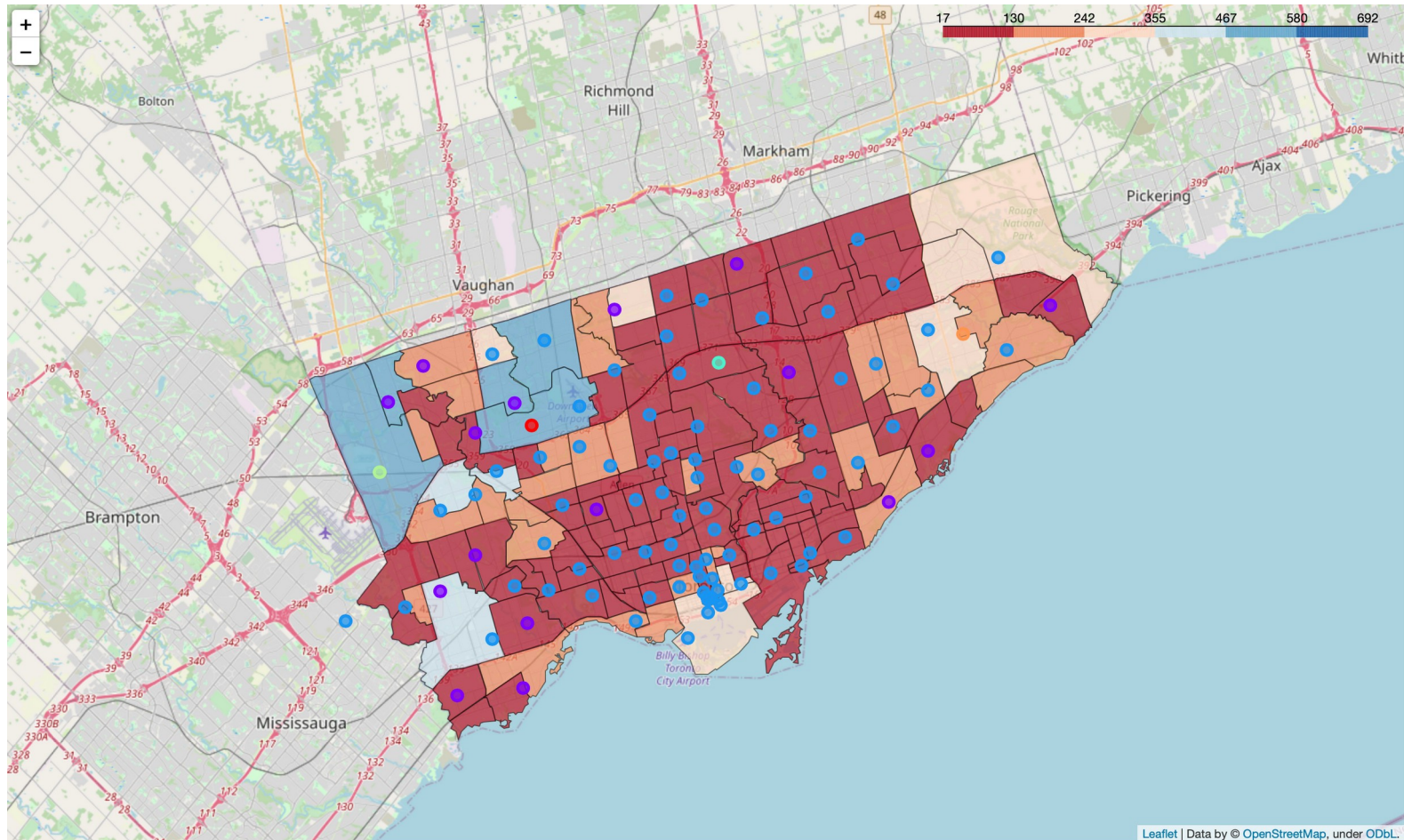
Toronto Clusters



Toronto – The heat map portrays the covid data. We ‘draw’ the clustered neighborhoods on top of it

- Again, the biggest group (blue) fits well to the zones on bright red, corresponding to the lower case count.
- The mid sized group (purple) is present on some bright red zones as well, but they these are particularly surrounded by other number densities.
- Interestingly, two isolated clusters appear on the areas with high case counts, both being close to an airport

Toronto Covid v/s Clusters



Discussion

- Both cities present very similar cluster distribution: two highly dominant groups consisting mostly on areas with bars, restaurants and banks, while the rest of the groups are small, without any particular pattern. From the two dominant groups, the biggest focuses on social activities while the second includes an important component of health and shopping districts. Clearly, this is related to the center and residential areas of each city, although the existence of two well differentiated central groups was unexpected.
- Second, there seems to be a tendency between the type of neighborhood the incidence of covid, with the biggest clusters, also being the ones showing more activity have a low incidence, while the isolated ones have the highest. **However**, most likely this is just a reflection of most of the people not living on the city center. A further study taking into consideration population density and, particularly, the presence of hospitals should follow.

Conclusions

- The main purpose of this modest study was to put into practice the knowledge obtained during the Capstone course. It is important to note that, for the sake of simplification, we make use of many assumptions that are hard to justify. For a proper, in-depth study on the differences between residential areas and the spread of covid more sophisticated tools and analysis should be done.
- However, some conclusions can still be extracted, which could be motivator for further, more serious investigations.