Leveraging Location Data to Model Hotspots for Possible COVID-19 outbreaks in Suburban Neighbourhoods

Shaik Abdillah

June 30, 2021

1 Problem Introduction

The COVID-19 pandemic precipitated in the end of 2019, where a then unknown SARS-like virus started spreading through the global population. The pandemic has caused widespread nation-wide lockdowns beginning in early 2020 and its effects are still in play some 2 years later. Much research has been conducted in an attempt to slow the spread of the virus; these included

1. Wearing Face masks
2. Social Distancing
3. Working from Home

It is thus of interest for public health officials to model possible hotspots based on venues that require unmasking and close physical contact. Thus, this project aims to develop a hotspot map of a neighbourhood based on the existing businesses that provide services which violate the above list of recommended safety measures.

2 Data Collection

The neighbourhood in question is ultimately arbitrary; but the location data used will be from the **FourSquare API**. Businessess will be ranked based on:

1. If the service(s) provided requires unmasking
2. The duration of which customers are present for
3. The possiblity of social distancing

Upon this three risk factors, a data set of types of business will be given points, one for each risk factor that they fall under. For example: a barbershop requires the stylist to be in close proximity to the client for an extended time and the client has to be unmasked for an extended period of time. Thus the barbershop will be given 3 points. On the other hand, a Retail shop requires only for the customers to be in close proximity to each other for a short period of time and thus will be given 1 point.

3 Modelling

The neighbourhood will be clustered using K-means clustering to determine centroids of businesses within a particular neighbourhood. Each centroid will then be ranked based on how many risk points all of the businesses in its vicinity and then plotted on a map to provide possible hostpots for local health officials to continually monitor.

4 Data Analysis

The data for the businesses around the area was obtained from the FourSquare API. The chosen location is **Christchurch** in New Zealand. In total there was more than 80 businesses in the vicinity.

5 K Means Clustering

The K means clustering was done on the latitiude and longitude of data set. The cluster number of 3 was obtained as based on visual inspection, made the most sense in the context of this project. The data was plotted against the map of the City.

6 Caveats of the model

There are apparent caveats of the model, in that it favours locations with localised high density of businesess. It would be different if it were in a different borough where the density of businesses was more spread out. Furthermore, there was not enough data for less populated areas such as the suburbs thus, the data may not reflect reality.

7 Discussion

Based on the results of the K-means clustering; we can see that there are 3 distict clusters where there are areas of higher traffic, thus, we can assume that the spread will be more concentrated in those regions and the regions surrounding it. In relation to the problem situation, we can thus take a more measured

approach when dealing with traffic in these regions. For example, authorities may perhaps consider banning certain high-risk high exposure activities such as exercise, and other activities that are done for prolonged durations an requires unmasking.

8 Conclusion

In conclusion, the data reflected 3 possible clusters where there might be a high occurrence of COVID-19 due to the high risk activities in those locations. The model can be expanded to include a more detailed analysis of the frequency of visitors and the population density of the area. A coupled differential equation of modelling the spread of the may be implemented in tandem with this clustering model in order to provide a greater insight of the spread of the virus. This can allow officials to take a data based apporach in legislating changes in the fight against the pandemic.