

Identifying clusters in UK for a phased relaxation of the COVID-19 lockdown

Introduction – Problem and Background

The COVID-19 pandemic has forced the UK government to impose a nationwide lockdown. This has resulted in all businesses and activities coming to a standstill. It is necessary to start reopening businesses and restart the economy while ensuring a safe exit from the lockdown without a sharp rise in infections.

To strike a balance, relaxation should also be in clusters of similar economic activity and density with lower infection rates. Lockdown relaxation in a single borough would not be practical, as a borough is too small an economic unit to survive. Hence the solution considers clusters of boroughs. The substitute for economic activity has been the types and density of business venues in a borough.

I will attempt to segregate the UK boroughs into clusters based on the type of venues in them and then cross-reference that with borough wise data on number of recorded COVID infections. This data would be useful to orchestrate the borough wise relaxation of lockdown based on the infection rates. It would also give a clear direction to health authorities to focus their resources on the high risk boroughs.

Data Description

Key data sources in this analysis are:

1. Borough wise COVID infection data
2. Foursquare API location data for the United Kingdom
3. Venues data for Boroughs in the United Kingdom

I plan to use the venues data for each borough as a symbol of the economic activity in that borough. Venues data is a good parameter to consider as business venues like restaurants, shops, etc. have a higher chance of people crowding in, and for spread of infections like COVID-19.

Methodology

Key steps in the methodology were:

1. CSV file with COVID-19 case data for Boroughs in UK to be uploaded into the Watson project storage and imported into the Jupyter notebook as a data frame
2. Using Geopy, the latitude and longitude values of all the Boroughs to be enriched.
3. Foursquare API to be used to get the top 100 venues for each of the borough within a radius of 10 km, as the maximum size of a borough is 5000 km² according to Wikipedia.
https://en.wikipedia.org/wiki/Districts_of_England
4. Cluster analysis to be run on this borough wise data. The K value to be determined after testing with multiple values to get the most meaningful number of clusters.
5. The cluster labels to be combined with the borough wise infection data from the CSV file. Cluster maps to be plotted.
6. Average infection rate (infections per 1000 population) to be calculated for each cluster. Cluster maps to be plotted for the clusters with highest and lowest infection rates.