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.

The key driver forcing the government to relax the lockdown is economic crisis. 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 have tried 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 used 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 have used 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 date for Boroughs in UK was 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 was enriched.
3. Foursquare API is 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 km^2 according to Wikipedia. *https://en.wikipedia.org/wiki/Districts_of_England*
4. Cluster analysis was run on this borough wise data, with K=10. The K value was arrived at after testing with multiple values to get the most meaningful number of clusters.
5. The cluster labels were combined with the borough wise infection data from the CSV file.
6. Average infection rate (infections per 1000 population) was calculated for each cluster.

Results

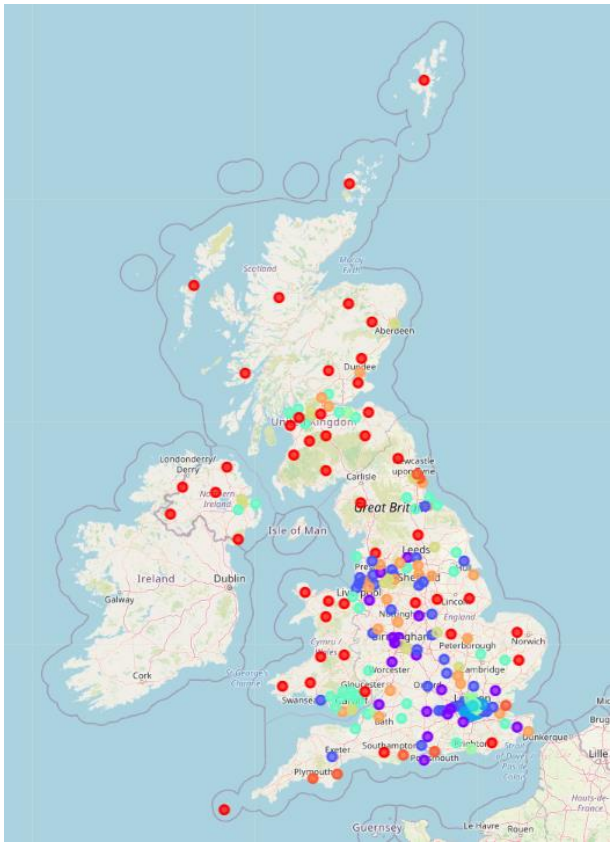
Summary of the average infection rate per cluster, and the number of boroughs per cluster is given in the table below. The table is ordered in ascending order of the infection rate.

Cluster Number	No. of infections per 100k	No. of boroughs
0	220.6	46
9	226.1	7
1	255.2	20
6	258.1	11
4	289.9	9
8	297.1	25
3	304.3	10
2	309.1	34
5	335.2	35
7	341.4	18

As can be seen in the table above and the Jupyter notebook provided,

- There is almost an even distribution of boroughs across clusters.
- About half the boroughs are below the point of 300 infections per 100 K population.
- Starting the relaxation of lockdown in clusters with least number of infections per 100K and then scaling up down the list would be a safe strategy.

The Cluster map is given below:



Discussion:

Based on this analysis, the **46 boroughs in Cluster 0** which have the lowest infection rate of 220.6 could be the ones that can be relaxed from lockdown with least risk of a steep rise in infections.

18 boroughs in cluster 7 with the highest infection rate of 341.4 should be the last ones to come out of lockdown, and should have the requisite medical facilities ramped up to tackle with rise in infections.

Overall, the clusters maybe further classified into 3 segments based on the average infection rate as follows:

Cluster Number	No. of infections per 100k	No. of boroughs	Segment
0	220.6	46	Low Risk
9	226.1	7	Low Risk
1	255.2	20	Low Risk
6	258.1	11	Low Risk
4	289.9	9	Medium Risk
8	297.1	25	Medium Risk
3	304.3	10	Medium Risk
2	309.1	34	Medium Risk
5	335.2	35	High Risk
7	341.4	18	High Risk

This kind of segmentation could be **used to drive policy decisions** around relaxation of lockdown as well as the focus **to ramp up medical facilities** based on the risk segment a borough is in.

Conclusion:

While clustering may be a good preliminary approach to frame strategy around lockdown relaxation for different boroughs, some more data around the below parameters would be necessary to frame a concrete policy decision.

1. Mortality rate in each borough. Different mortality rates might require more medical facilities or lockdown rules
2. While the current data considered business venues in each borough, another significant mode of infection spread is public transportation, especially the underground trains. This would have a higher effect on urban clusters compared to rural clusters and suburbs
3. Office locations and industries in the vicinity is also an important data point. Especially industries that cannot have employees working from home would result in congregation of people in one location as well as higher usage of public transportation