

Covid-19 Cases and Neighbourhood Venues Analysis for Manchester, UK

Applied Data Science Capstone by IBM/Coursera

Iseul Song

1. Introduction/Problem

Covid-19 is an ongoing global pandemic of coronavirus. Since the outbreak of the virus was first identified in December 2019 in Wuhan, China, more than 23.6 million cases of Covid-19 have been reported in more than 188 countries and territories, leading to more than 814,000 death as from 25 August 2020.

Manchester is one of the hotspots of the Covid-19 cases compared to the rest of the UK. The UK government has recently tightened social distancing measures in the area to stop the spread of the virus.

In this context, this project undertook a case study of Manchester to provide an understanding of the potential relationship between Covid-19 infection rates and neighbourhoods in Manchester. In particular, it focused on how the infection rates differed by the neighbourhoods' most common venue types to investigate if a neighbourhood which had certain types of venues the most was likely to have higher infection rates. This study targeted policymakers and healthcare-sector workers who might be interested in the effectiveness of measures to stop the spread of the virus by restraining people from particular activities.

2. Data

To solve the defined problem above, Middle-Layer Super Output Areas (MSOAs) were used as a unit of neighbourhood data. MSOAs are statistical geography created for the 2011 Census of England and Wales. A typical population is 7,000–10,000 people per MSOA.

The following list of data was used for a MSOA-level analysis.

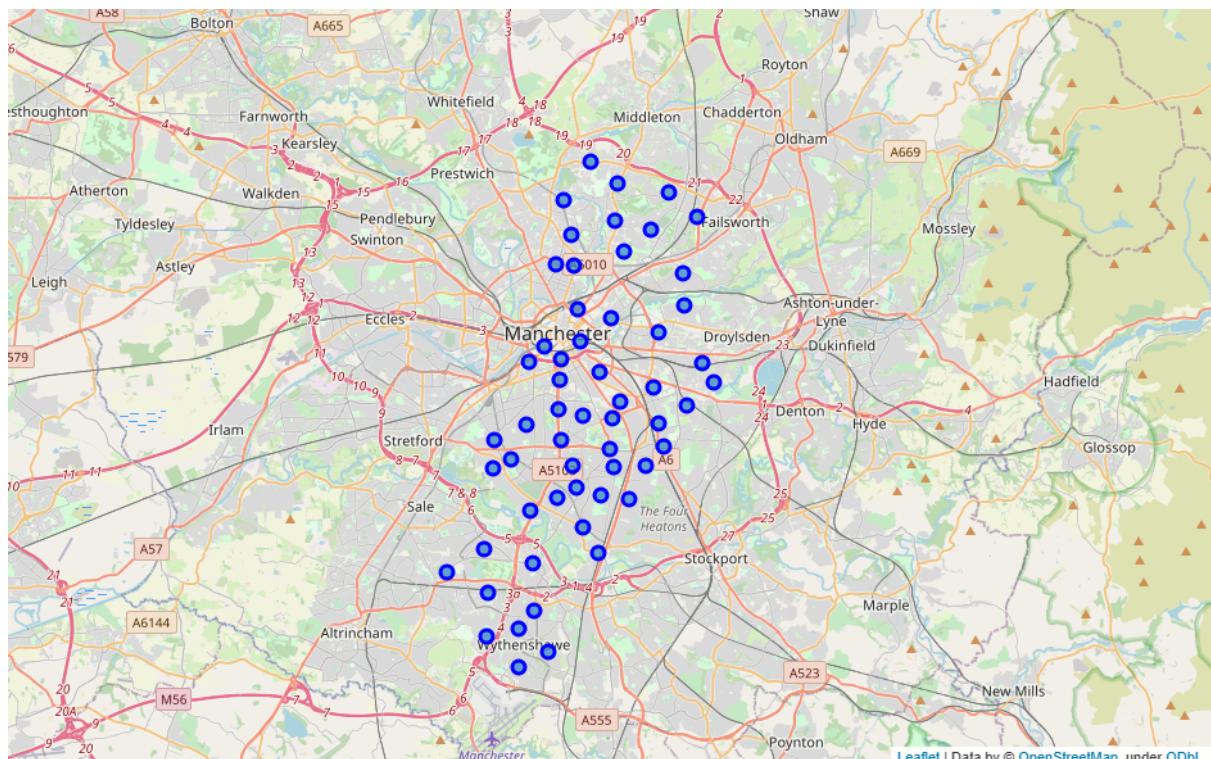
- 2011 Population weighted centroid location by MSOA from the ONS Geography Open Data for coordinates of MSOAs (<https://geoportal.statistics.gov.uk/datasets/middle-layer-super-output-areas-december-2011-population-weighted-centroids?geometry=-2.466%2C53.438%2C-2.032%2C53.510>);
- Neighbourhood names by MSOA from the House of Commons Library MSOA Names (<https://visual.parliament.uk/msoanames>);

- Number of Covid-19 confirmed cases by Middle Super Output Area (MSOA) from the Government's Coronavirus website (<https://coronavirus.data.gov.uk/cases> retrieved on 23rd August 2020);
- 2018 Population estimates by MSOA from the National Statistics to calculate infection rates (<https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/datasets/middlesuperoutputareamidyearpopulationestimatesnationalstatistics>);
- Foursquare API to get the most common venues of the neighbourhoods in Manchester.

3. Methodology

Using the listed data above, this study clustered 57 neighbourhoods in Manchester, which are presented in Figure 1, based on their similarity of the most common venue types and compared infection rates between the neighbourhood clusters. It was assumed that venues are most likely to be used by usual residents who live in walking distance from the venues in the neighbourhoods rather than people from other neighbourhoods. This was a reasonable assumption under the social distancing rules and the government's advice that people should work from home if they can.

Figure 1 Location of neighbourhoods in Manchester



The first step was an exploration of the frequency of venue types across the neighbourhoods in Manchester, using Foursquare API. The radius for counting venues was set as 800 metres, which is widely considered as a ten-minute walking distance. The limit was set as 200 venues in this analysis. Figure 2 shows 10 rows from the created data frame of venues by

neighbourhood and 1,379 venues were returned by Foursquare in total. By running code to count unique values of the venue category, it was found that there are 197 venue types. The number of venues by venue type was calculated in each neighbourhood to analyse the most common venue types from 1st to 10th in order as shown in Figure 3 (only 5 rows of the neighbourhoods were displayed).

Figure 2 Venues by neighbourhood

	Neighbourhood	Neighbourhood Latitude	Neighbourhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Clayton Vale	53.488394	-2.177365	Clayton Vale	53.484735	-2.177158	Park
1	Clayton Vale	53.488394	-2.177365	Brewers Fayre	53.486428	-2.173513	Pub
2	Clayton Vale	53.488394	-2.177365	Clayton Hall Metrolink Station	53.485007	-2.177443	Tram Station
3	Clayton Vale	53.488394	-2.177365	Hewlet Johnson Playing Fields	53.483562	-2.168788	Park
4	New Islington & Miles Platting	53.484197	-2.216825	Pollen Bakery	53.483487	-2.224372	Bakery
5	New Islington & Miles Platting	53.484197	-2.216825	Cloudwater Brew Co.	53.477914	-2.221918	Brewery
6	New Islington & Miles Platting	53.484197	-2.216825	Ancoats Coffee Co	53.483536	-2.227407	Coffee Shop
7	New Islington & Miles Platting	53.484197	-2.216825	Unit 9 Cloudwater Taproom	53.478033	-2.221925	Beer Bar
8	New Islington & Miles Platting	53.484197	-2.216825	The Hip Hop Chip Shop	53.484978	-2.227510	Fish & Chips Shop
9	New Islington & Miles Platting	53.484197	-2.216825	Elencot	53.484649	-2.228448	Italian Restaurant

Figure 3 Most common venues by neighbourhood

	Neighbourhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Abbey Hey	Gym / Fitness Center	General College & University	Park	Hotel	Fast Food Restaurant	Train Station	Market	Bakery	Fish & Chips Shop	Farmers Market
1	Ardwick	Music Venue	Chinese Restaurant	Concert Hall	Bar	Museum	Café	Beer Bar	Korean Restaurant	Theater	Fish & Chips Shop
2	Baguley East & Wythenshawe Park	Fast Food Restaurant	Tram Station	Gym / Fitness Center	Bus Stop	Women's Store	Electronics Store	Fish Market	Fish & Chips Shop	Farmers Market	Falafel Restaurant
3	Baguley West & Brooklands	Supermarket	Tram Station	Coffee Shop	Grocery Store	Furniture / Home Store	Donut Shop	Pet Store	Clothing Store	Hardware Store	Arts & Crafts Store
4	Beech Road & Chorlton Meadows	Pub	Bar	Pizza Place	Café	Coffee Shop	Grocery Store	Fast Food Restaurant	Park	Turkish Restaurant	Mediterranean Restaurant

Secondly, neighbourhoods were clustered using unsupervised learning K-means algorithm, which is one of the most widely used cluster methods. To determine the optimal k, the elbow method was initially tested considering that it is the most common method. However, as shown in Figure 4, the elbow was not very clear for this case. Therefore, several k values were tested manually and 5 was decided to be the k value for the analysis. It was based on the total number of samples for clustering and the number of samples in each cluster. Figure 5 shows the location of the neighbourhoods, colour-coded by the cluster.

Figure 4 Elbow method for optimal k

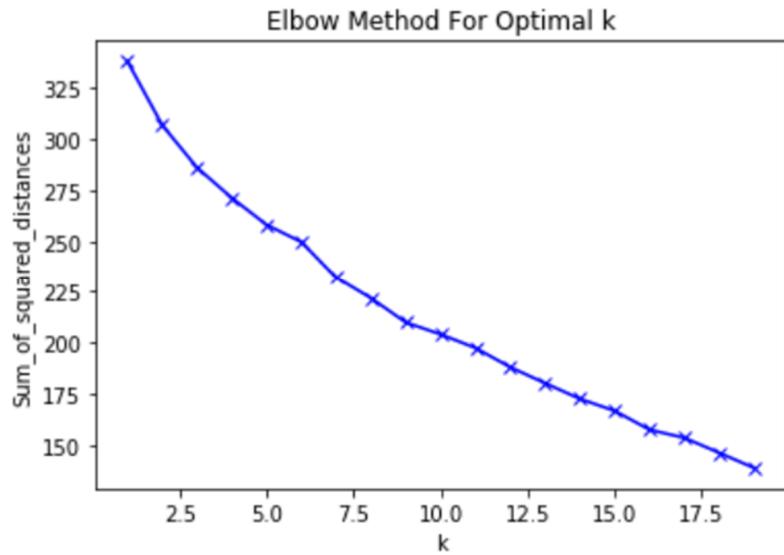
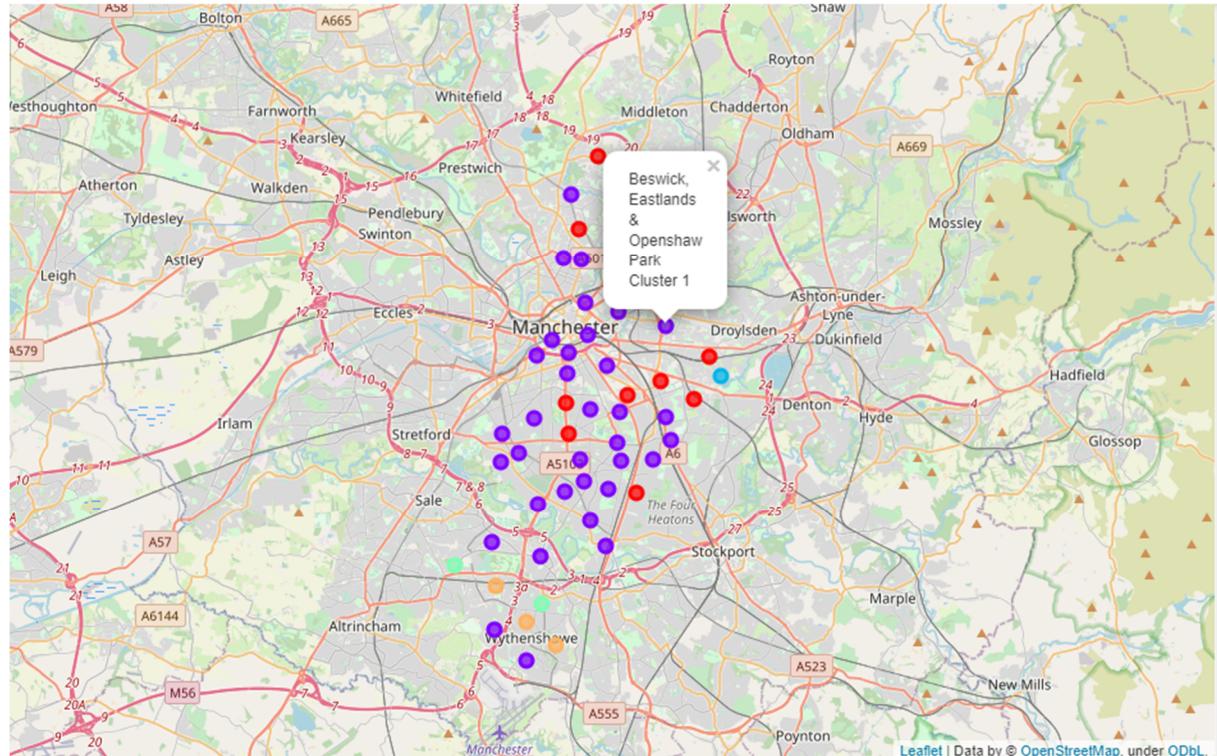


Figure 5 Neighbourhood clusters



The 1st most common venues were analysed to understand which types appear the most in the neighbourhoods of each cluster, using One-Hot Encoding and visualisation of the data. Figure 6 to 10 correspond the visualisation of the 1st most common venues for the five clusters respectively.

Figure 6 Cluster0 - 1st most common venue

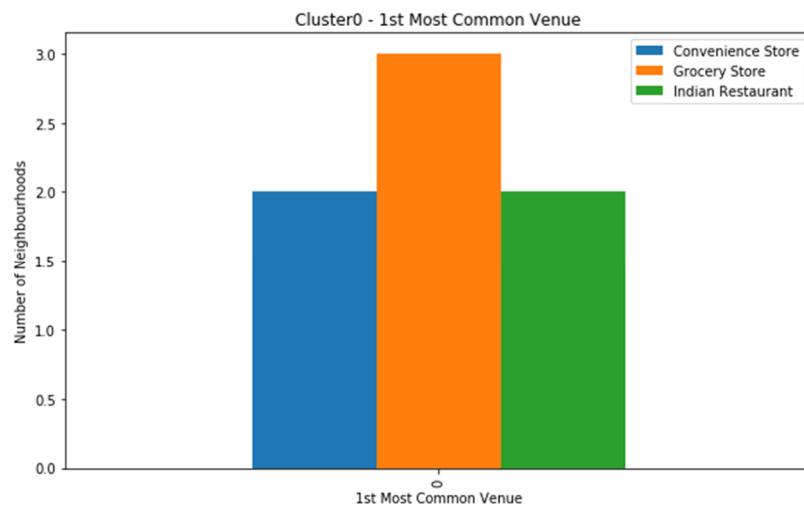


Figure 7 Cluster1 - 1st most common venue

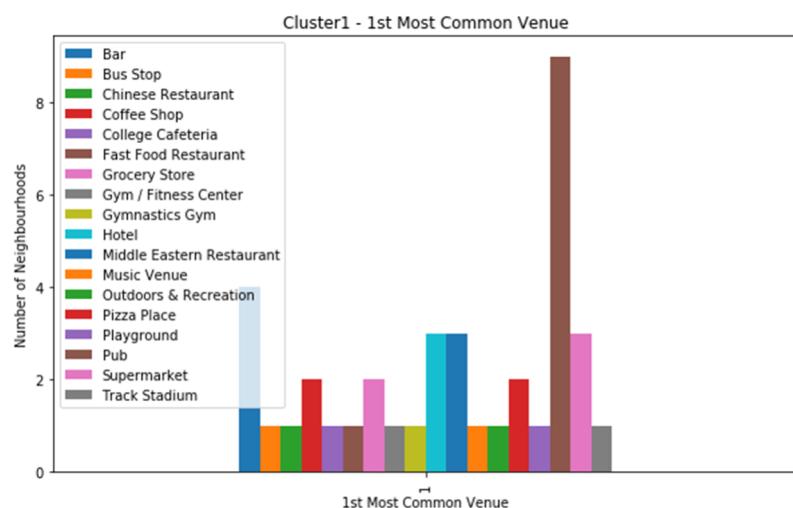


Figure 8 Cluster2 - 1st most common venue

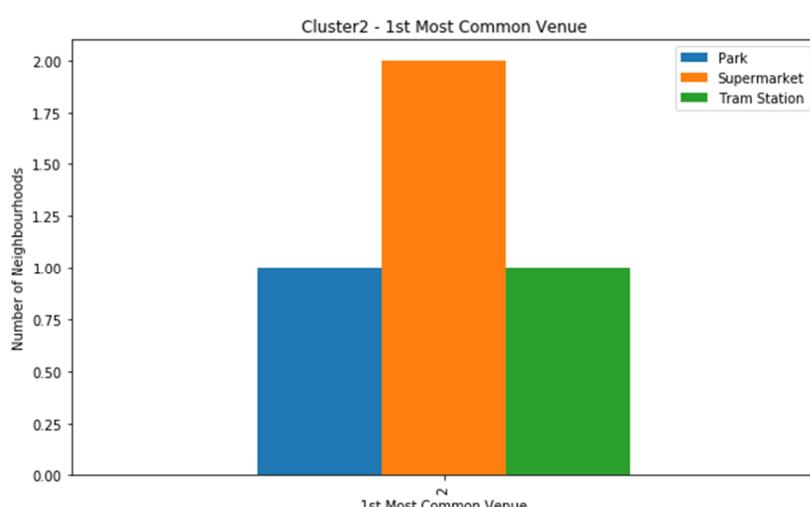


Figure 9 Cluster3 - 1st most common venue

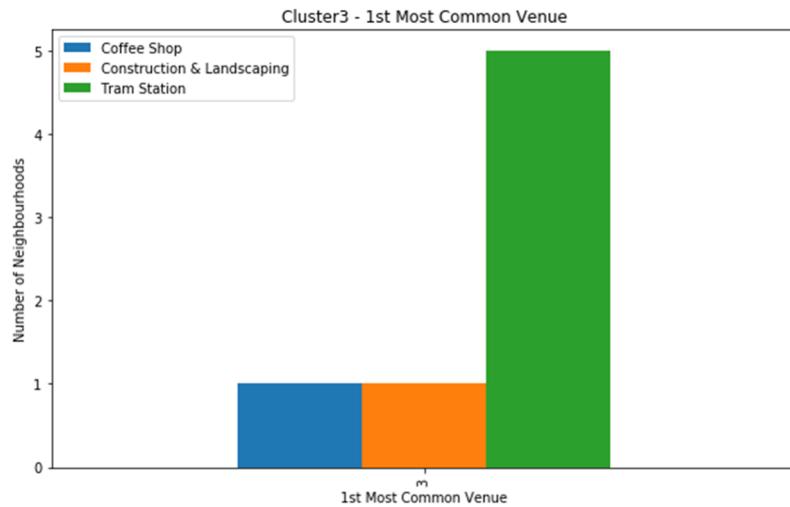
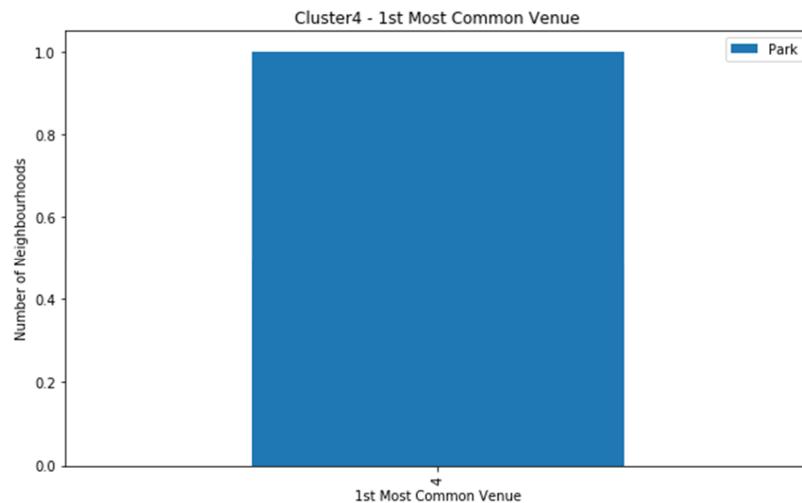


Figure 10 Cluster4 - 1st most common venue



For Cluster 0, grocery stores are the 1st most common venue in most of the neighbourhoods, followed by convenience stores and Indian restaurants. For Cluster 1, pubs and bars are the 1st most common venues in the majority of neighbourhoods. Supermarkets are the 1st most common venue in most of the neighbourhoods in Cluster 2. Tram Stations are the 1st most common venue in the majority of neighbourhoods in Cluster 3. Cluster 4 has only one neighbourhood and parks are the 1st most common venue.

In summary, the neighbourhood clusters were given the following names to reflect their most common venues.

- Cluster 0: Supermarket and Indian Restaurant Cluster
- Cluster 1: Pub and Bar Cluster
- Cluster 2: Supermarket Cluster
- Cluster 3: Tram Cluster
- Cluster 4: Park Cluster

Lastly, to compare Covid-19 infection rates between the neighbourhood clusters, an infection rate was defined as the number of cases per 100,000 people, which is the most commonly used definition in the UK. Using the data of the total number of confirmed cases and the total population by neighbourhood, infection rates were calculated as shown in Figure 11.

Figure 11 Infection rates by neighbourhood

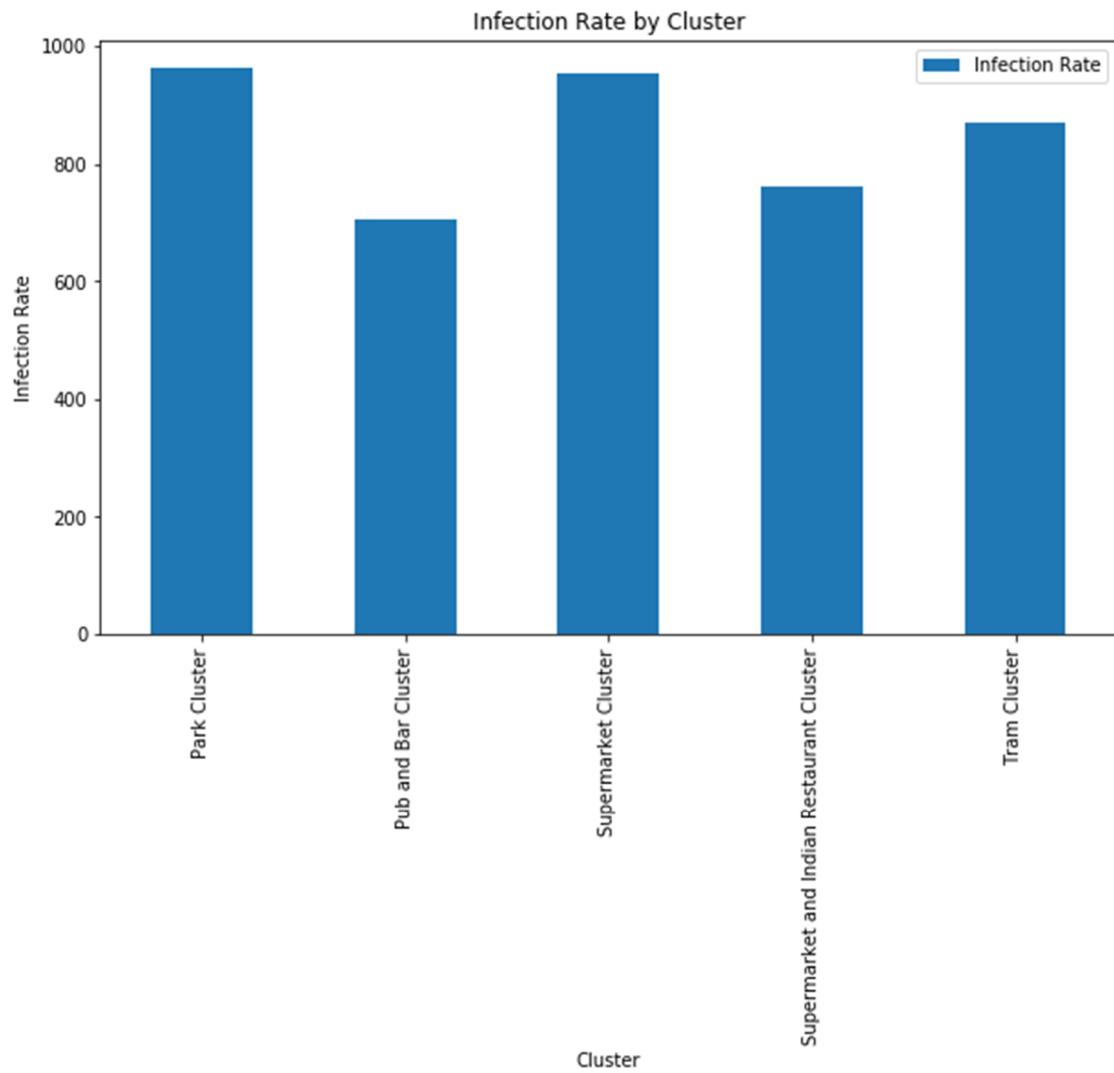
Local authority district name	Msoa code	Neighbourhood	Total case	All Ages	Infection Rate
0	Manchester E02001045	Boothroyden & Higher Blackley	64	8269	773.975088
1	Manchester E02001046	Blackley	98	10377	944.396261
2	Manchester E02001047	Charlestown	98	10002	979.804039
3	Manchester E02001048	Crumpsall North & Heaton Park	93	8841	1051.917204
4	Manchester E02001049	New Moston	75	9764	768.127816

The average infection rates between the neighbourhoods were calculated for each neighbourhood cluster. Figure 12 and 13 show the average infection rates by the cluster.

Figure 12 Average infection rates by neighbourhood cluster

Cluster Labels	Infection Rate
Park Cluster	961.654045
Pub and Bar Cluster	706.562403
Supermarket Cluster	953.633227
Supermarket and Indian Restaurant Cluster	761.859195
Tram Cluster	869.388880

Figure 13 Graph of the average infection rates by neighbourhood cluster



4. Results

From the analysis, the Park Cluster has the highest infection rate among the five clusters with 962 cases per 100,000 people. However, it is worth noting that the Park Cluster has only one neighbourhood. Therefore, it is more likely that there are other factors which contribute to the high infection rate in the particular neighbourhood of the cluster. Following the Park Cluster, the Supermarket Cluster and the Tram Cluster have the second and third highest infection rates respectively.

The Pub and Bar Cluster has the lowest infection rate with 707 cases per 100,000 people, which is approximately 1.4 times lower than the infection rate of the Park Cluster. This cluster contains the neighbourhoods in the Manchester city centre and commercially vibrant neighbourhoods in the south of Manchester. The Supermarket and Indian Restaurant Cluster has the second-lowest infection rate, contrasting with the Supermarket Cluster having the second-highest infection rate.

5. Discussion

Although it is widely believed that there is a high risk of infection in venues of the hospitality industry such as restaurants, pubs and bars, the results of the analysis indicate that neighbourhoods with more of those venues tend to have lower infection rates. However, it is important to consider that hospitality venues were entirely closed during the national lockdown period in the UK and only reopened about two months ago. The most common venues not being operated could contribute to the lower infection rates in the neighbourhoods.

A limitation of this study is that it assumes infection happens in people's home neighbourhoods. In reality, some people might be contracted outside of their home neighbourhoods and that is not considered in this study.

For a better understanding of any relationship between venue types and Covid-19 infection rates, future studies could analyse the number of cases only recorded after all the venues were allowed to open. This will allow us to better capture the role of each venue type under no formal restriction on a particular business. Furthermore, it would be valuable to look at wider areas and more various characteristics of neighbourhoods to see if there is any general pattern of infection rates by neighbourhood characteristics.

6. Conclusion

The study analysed Covid-19 infection rates and the most common venue types across the neighbourhoods in Manchester. Although the results show the neighbourhoods which have hospitality venues most commonly have the lowest infection rate in Manchester, the results should be further investigated with other factors such as the government's restrictions on certain business. Policymakers and healthcare-sector workers who are working hard to stop the spread of the virus might be interested in pursuing further studies on this topic. It will give them a clearer understanding of what type of activities they should restrain people from doing and the effectiveness of the interventions, minimising negative impacts on the national and local economy.