The impact of different venue types on the prevalence of COVID-19 cases in London

Felix Mantz

3 January 2021

1. **Introduction**
   a. **Background**
   The global COVID-19 pandemic has dominated much of 2020, with countries enacting lock-downs and other socio-economic measures to contain the spread of the virus. Nevertheless, the virus has continued to spread throughout the world. The virus is primarily spread through droplets (e.g. of saliva) from a contagious person to another non-infected person. Venues with close proximity between customers and/or staff may therefore be contributing to the spread of COVID-19.

   b. **Problem**
   This project aims to identify if specific venue types are more prevalent in London boroughs which have seen faster increases in COVID-19 cases than other boroughs.

   c. **Interest**
   The results of this analysis may be of interest to policymakers considering if specific venues should be closed to help curb the spread of COVID-19. It may also be of interest to people aiming to avoid infection with COVID-19 who may chose to avoid certain London boroughs or neighbourhoods.

2. **Data**
   a. **Data sources**
      i. Daily number of new and current total COVID-19 cases for London boroughs from 11 February 2020 to 27 December 2020 is provided by Public Health England and can be accessed here
      ii. The population of each London borough can be found here. This is not presented in a format that is readily downloadable or scrapeable for free, so has to be transcribed manually
      iii. To analyse this data spatially, a GeoJSON file with outlines of London boroughs can found here
      iv. Lastly, information on venue types data can accessed using the Foursquare API, by requesting venues 'near' each of the London boroughs found above
   b. **Data cleaning**
   Initial cleaning of the COVID case data included formatting the dataframe (unstacking, index setting), and expressing the absolute number of new cases in each borough per 100k inhabitants of said borough.

   COVID case data was combined for two boroughs, Hackney and the City of London. The City of London borough only has c10k residents, so my assumption is the COVID data was grouped with the larger neighbouring Hackney borough as it would otherwise be insignificant. This has knock on implications:
      i. The GeoJSON file to create a map of cases needed to have the geometries of these two boroughs combined. This was done by individually extracting the

geometries for these boroughs, merging them using the shapely library, and then updating the GeoJSON file.

ii. As the combined "Hackney and the City of London" does not exist, the Foursquare API was not able to resolve it as a viable geocode and did not deliver any venues. As a proportionate fix for this, for the purpose of querying Foursquare I simply looked for venues in Hackney.

c. **Feature selection**

I opted to focus on the number of new cases each day rather than total cases, as I wanted to get investigate where the virus spread most quickly. The change in total cases day to day is further muddied as it includes recoveries/deaths, and so is not a good indicator of new cases.

I did not trust the Foursquare API to be a comprehensive list of all venues in a borough. So, instead of working with absolute venue type frequencies, I worked out the proportion of each type for each borough. This gives an indication of the composition of venues in the borough, assuming those provided by Foursquare are a representative sample.

3. **Methodology**

a. The first step was to analyse new covid cases per 100k residents of each borough. An initial plot (Fig 1) of new cases in each London borough over time shows that these tend to move together overall, but some boroughs seem to be consistently at the top of each spike in new cases.

Fig 1:

Number of COVID-19 cases per London Borough per 100k residents, 7-day simple moving average
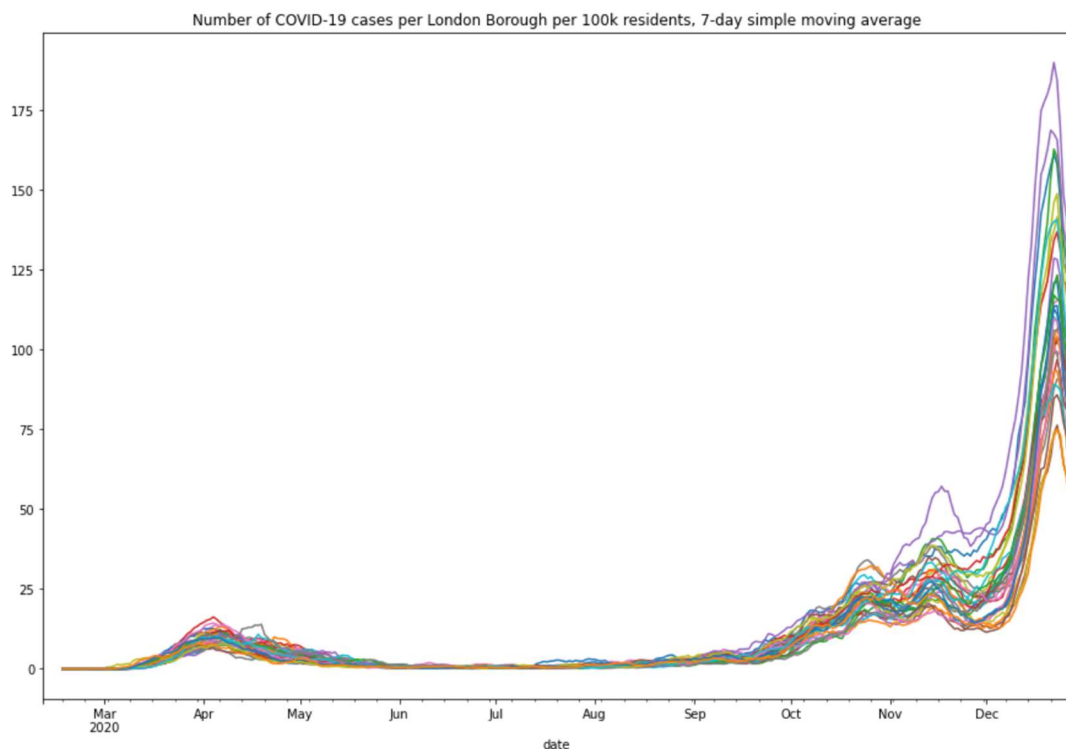
b. Looking at the location of boroughs with high cases in December 2020, it is apparent that boroughs with high cases are predominantly in the east. This choropleth map was chosen to show the COVID spread through neighbouring boroughs.
Fig 2:
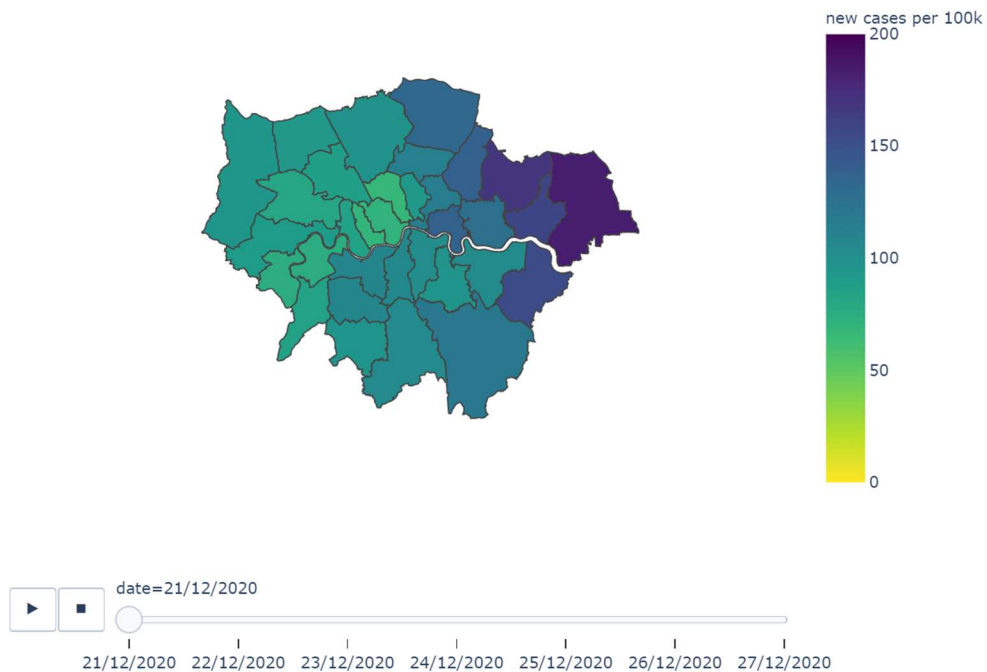
c. Next, looking at the types of venues most common in each borough (per Foursquare API), we look at how similar boroughs are from each other. Fig 3 shows that that for most boroughs, pubs tend to be the most common venue – but it is unclear just how common. I have therefore worked out proportion of venues for each borough, to get a more nuanced view than the simple ranking shown in Fig 3.
Fig 3

|	Borough	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
0	Barking and Dagenham	Supermarket	Coffee Shop	Park	Clothing Store	Pub	Furniture / Home Store	Department Store	Grocery Store
1	Barnet	Coffee Shop	Pub	Fast Food Restaurant	Pharmacy	Bookstore	Pizza Place	Golf Course	Grocery Store
2	Bexley	Pub	Grocery Store	Coffee Shop	Supermarket	Park	Pharmacy	Clothing Store	Gym / Fitness Center
3	Brent	Coffee Shop	Indian Restaurant	Hotel	Clothing Store	Park	Gym / Fitness Center	Pub	Café
4	Bromley	Pub	Grocery Store	Clothing Store	Coffee Shop	Pizza Place	Gym / Fitness Center	Indian Restaurant	Bar
5	Camden	Hotel	Park	Bakery	Pizza Place	Coffee Shop	Indian Restaurant	Garden	Juice Bar
6	Croydon	Pub	Coffee Shop	Hotel	Clothing Store	Park	Grocery Store	Mediterranean Restaurant	Bookstore S
7	Ealing	Pub	Park	Coffee Shop	Hotel	Italian Restaurant	Café	Pizza Place	Burger Joint

d. This information is then combined with the 10 boroughs that tended to have more new cases than London overall on any given day, and we look for any patterns visually

4. Results

    a. The top 10 boroughs which tended to have more daily new cases than the rest of London are (noting that this is consistent with the eastern focus shown in Fig 2):

        i. Havering
        ii. Redbridge
        iii. Barking and Dagenham
        iv. Bexley
        v. Tower Hamlets
        vi. Newham
        vii. Waltham Forest
        viii. Enfield
   ix. Ealing
        x. Harrow

b. The venues in these boroughs differed from the London average in a number of places. Fig 4 below shows where each of the top 10 boroughs had substantially more venues of a certain type (positive) or less (negative) than the London average.

Fig 4
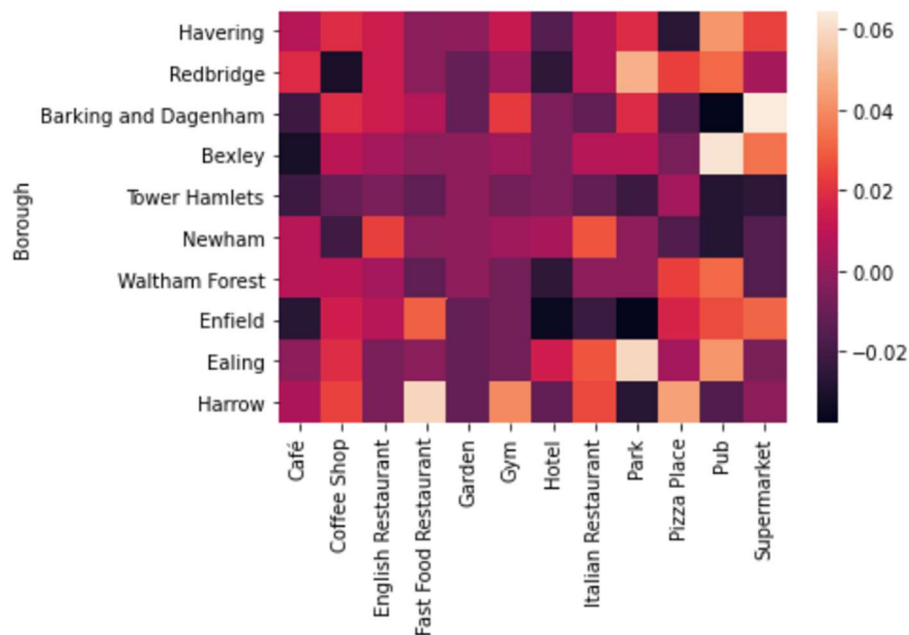
c. Fig 4 does not show a clear trend, but does suggest that some venues tend to be more common among the top 10 COVID boroughs than for London overall. Considering the average for all top 10 COVID boroughs in Fig 5, we see that these tended to have more pubs than the London average.

Fig 5
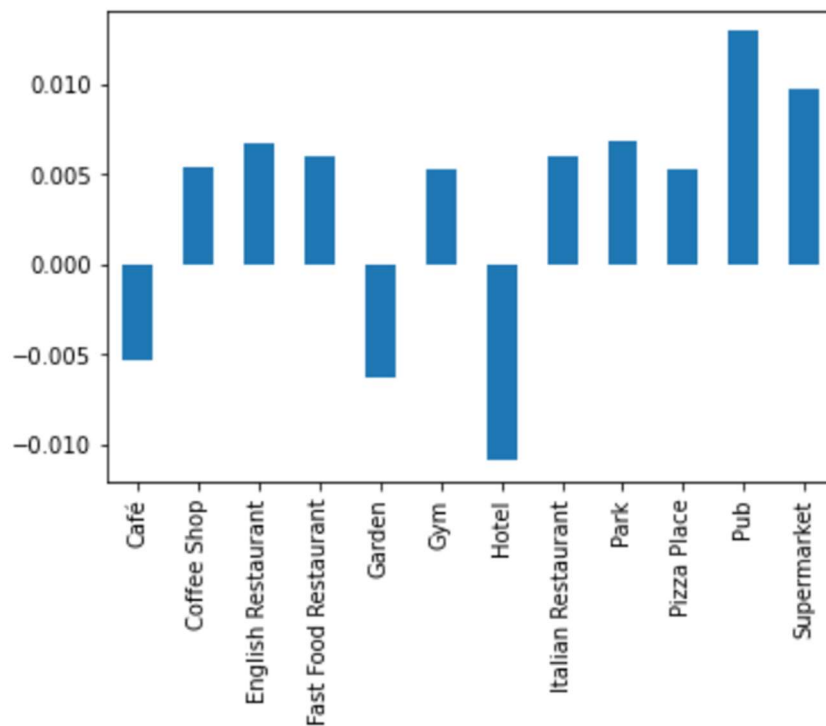
d. Investigating the theme of pubs further, we look a regression plot in Fig 6 showing that over the entire data set (all 32 boroughs considered) there is a general trend that those with more pubs have seen more rapid increases in COVID cases. Given the complexity of the situation, more advanced machine learning tools are unlikely to give a better fit, which is why a simple and widely understood linear regression was chosen.

Fig 6
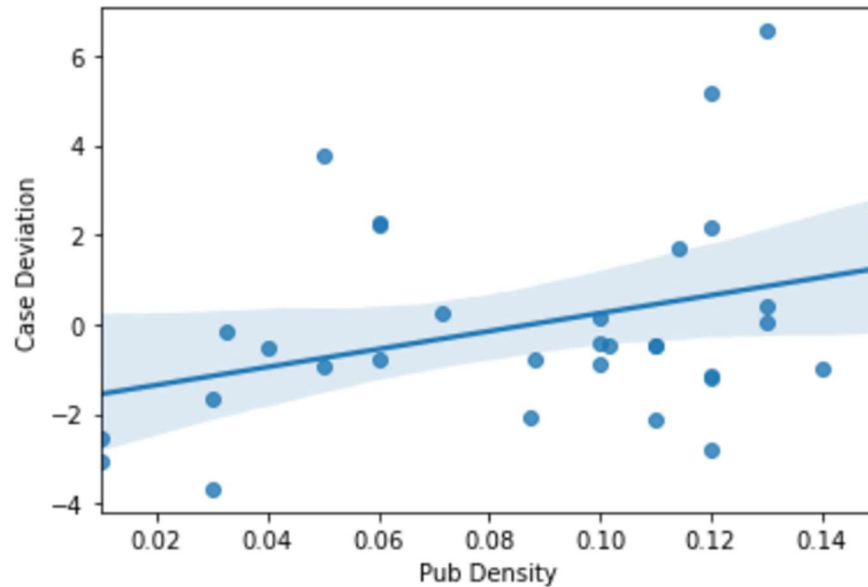
5. Discussion

a. We have seen that there is a positive correlation between the proportion of venues that a borough has that are pubs and the number of new COVID cases it may experience above the London average. However, the trend is not necessarily strong, and outliers clearly exist, suggesting that pub density alone is not sufficient at explaining why some London boroughs have seen more new cases than.

b. While this analysis focuses on venues types in each borough, the Choreoplath map (Fig 2) has clearly shown that neighbouring boroughs tended to have similar rates. In fact, an animation of the map overtime shows a wave of new cases "rolling in" from eastern boroughs in December 2020. This suggests that further analysis should also look at the spread from one borough to a neighbouring one.

c. An additional thought for further analysis is to consider the implications of lockdown, and allowing for when various venues would have been shut. See Expansion A at the end of this report.

6. Conclusion

a. The analysis has shown that borough with more pubs tended to have more new cases than others. However, it has also made clear that pubs alone are not a sufficient explanatory variable, and other factors such as borough neighbours, will impact the spread of COVID.

b. This analysis is therefore tacitly supportive of the Government's measures to close pubs in an attempt to stop the spread of COVID; however, no policy decision should be taken based on this metric in isolation.

Expansion A:

Many venues have been shut at various times over the course of 2020 due to government restrictions in place for London/the UK, so future analysis could focus on time periods where venues were largely open (15 June 2020 to 5 November 2020 – see timeline below).

Timeline of key policy events relevant for the analysis
- 15 June 2020: Non-essential shops re-opened on 15 June following the first wave. Begin of the analysis period
- August 2020: "Eat out to help out" scheme encouraging restaurant use
- 14 September 2020: "Rule of six" capping social meetings at 6 people
- 15 October 2020: London moved to "Tier 2", further restricting meeting
- 5 November 2020: Begin of second UK national lockdown, including closing of all non-essential venues