

Future Spatial Data Scientist's Group Project

Declaration of Authorship

We, Future Spatial Data Scientist, confirm that the work presented in this assessment is our own. Where information has been derived from other sources, we confirm that this has been indicated in the work. Where a Large Language Model such as ChatGPT has been used we confirm that we have made its contribution to the final submission clear.

Date: 2023/12/18

Student Numbers:

23060146

23083058

23058858

23125714

23064414

Response to Questions

1. Who collected the data?

The data was collected by Wentao Lei from Airbnb(<http://insideairbnb.com/>) and ONS website.

2. Why did they collect it?

The data was collected for a study titled ‘Opportunities and Risks arising from Covid-19,’ aimed at understanding how Covid-19 has affected London’s rental market. To do this, it’s important to look at Covid-19 case and death numbers because they show how the pandemic has impacted people’s decisions on renting. We’re checking how Airbnb listings have changed over time and if these changes are linked to the pandemic’s spread. We also want to see if apartments leaving Airbnb are being rented out long-term instead. Including health data helps explain why the rental market moves in certain ways during the pandemic. The research also involves examining the entry to and exit from the Airbnb marketplace by comparing snapshots of London’s rental data at different time points. The goal is to assess how the pandemic has influenced these markets and to make inferences about the movement of properties between Airbnb and the long-term rental sector. This requires making reasonable assumptions, such as whether all flats withdrawn from Airbnb are re-entered into the long-term rental market, and these assumptions need to be documented and justified within the study.

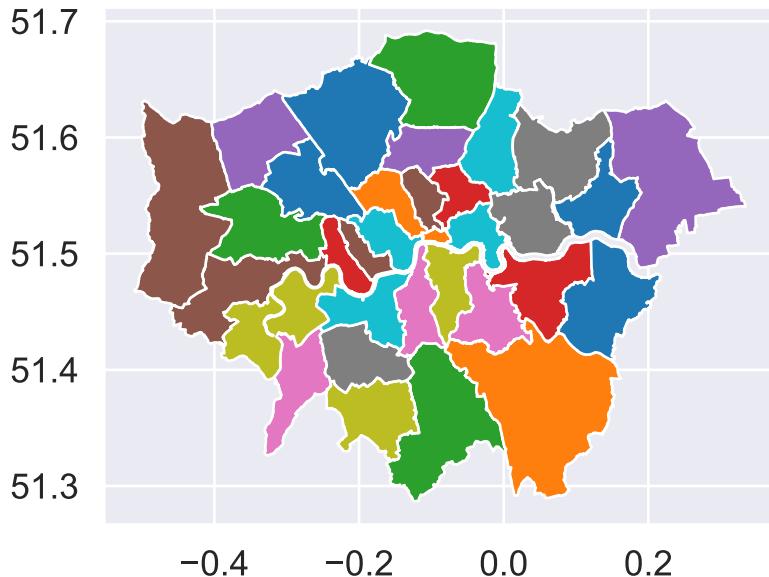
3. How was the data collected?

Although the exact methods of data collection are not specified in the information provided, it is inferred that Airbnb data may have been harvested from public listings on the Airbnb platform or through an API made available by Airbnb (<http://insideairbnb.com/>). For the COVID-19 infection and death data, the likely source is the Office for National Statistics’ (ONS) website. These datasets have undergone a rigorous data cleaning process, an essential step to ensure accuracy and reliability. The data cleaning involved removing duplicates, handling missing values, correcting inconsistencies, and filtering irrelevant information to hone in on pertinent data points. Moreover, to enable a more granular analysis, the Airbnb dataset was further refined by selecting listings from February, April, June, August, October, and December of 2020 as a detailed time-scale for subsequent examination.

The data analysis techniques are illustrated in the `Covid-19.py` script, which include reading CSV files for structured data storage, merging datasets to correlate diverse sources of information, employing geospatial data, and generating visualizations for an enhanced interpretive experience.

Python, a staple in data science, is utilized for these tasks, with libraries such as Pandas for data manipulation and analysis; Matplotlib and Seaborn for visualization; and GeoPandas for geospatial data handling; Functions like `pd.read_csv()` read data from CSV files, `gpd.read_file()` is used for geospatial data, and the `plot` method creates maps and other graphical data representations.

London Map



After cleaning, the data could be structured for analysis, statistical tests were performed, and the findings were visualized through graphs and maps to reveal the geographical distribution and trends of COVID-19's impact on rental markets.

4. How does the method of collection impact the completeness and/or accuracy of its representation of the process it seeks to study, and what wider issues does this raise?

Using “last comment data” to estimate the rental date is an approximation. This may introduce inaccuracies as the rental date may not exactly match the date of the last review. The exact date of the rental may be earlier or later than the last review date, meaning the findings may not accurately reflect the actual rental market.

This error may affect the long-term rental market differently than the short-term rental market. In long-term rental markets, where rentals last for longer periods of time, errors in rental dates may be more significant, affecting the completeness and accuracy of the study. In the short-term rental market, the rental duration is shorter, so this error may be relatively small and will not have a big impact on the study.

Taken together, conclusions based on approximate data may not apply to all rental situations. For example, errors may be more significant in certain areas or specific rental types, so caution is needed when generalizing findings. The accuracy of research is limited by data collection methods, and policymakers and market participants need to be aware of these limitations so that they can be considered when making decisions.

5. What ethical considerations does the use of this data raise?

Privacy and Data Protection: The use of personal data, such as names, locations, and other identifying information, raises significant privacy concerns. Ensuring that the data is anonymized and does not violate individuals' privacy is crucial. This is in line with the principles discussed in Zook *et al.* (2017), who emphasize the importance of anonymization in big data.

Impact on Stakeholders: Releasing detailed Airbnb data can have various impacts:

Hosts (Merchants): They might face privacy breaches or unwanted exposure. Additionally, competitors or local authorities could use this data against them, as indicated by Ert, Fleischer and Magen (2016) in their study on the competitive dynamics in the hospitality industry.

Tourists: If their travel patterns or stays are revealed, it could lead to privacy violations or security risks, which Ferreri and Sanyal (2018) explore in their work on the privacy concerns in urban analytics.

Government/Authorities: The data might reveal regulatory non-compliances or tax evasion, leading to legal actions or policy changes, as discussed by Edelman and Geradin (2016) in the context of digital markets regulation.

Accuracy and Misinterpretation: Ensuring the data's accuracy is vital, as incorrect data can lead to false conclusions and potentially harmful decisions. The significance of data accuracy is underscored by Boyd and Crawford (2012), in their critical examination of big data's impact on decision-making.

Legal and Ethical Compliance: The data must be used in compliance with laws like GDPR in Europe or other local data protection laws. Ethical use also involves considering the potential negative effects of data release on various communities and individuals, as highlighted by Newell and Marabelli (2015).

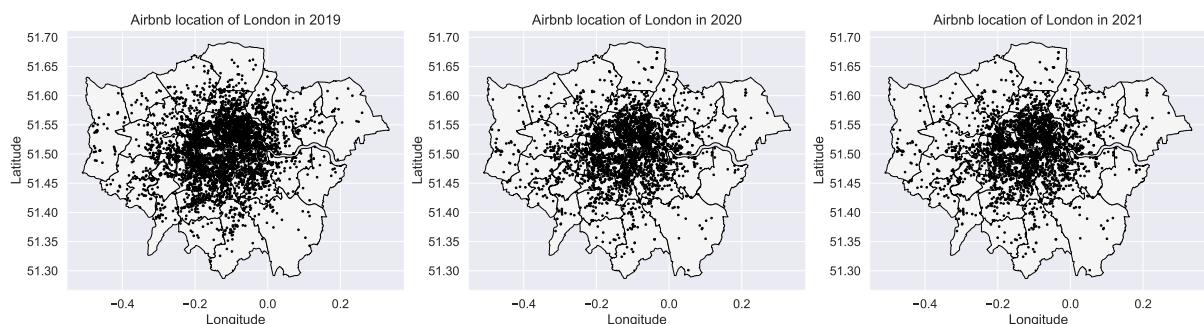
Economic Impact: Revealing certain data about Airbnb's operations might negatively impact local real estate markets, rental prices, and the tourism industry. This economic influence is detailed in the research by Guttentag (2015).

Social Consequences: The release of this data might lead to a backlash against Airbnb hosts or guests in certain communities, affecting social harmony.

6. With reference to the data (*i.e.* using numbers, figures, maps, and descriptive statistics), what does an analysis of Hosts and Listing types suggest about the nature of Airbnb lets in London?

1 Add Airbnb location to London Map

Create a large graph with three horizontally aligned subgraphs, each showing the distribution of Airbnb listings in London in different years (2019, 2020, and 2021).



This graph shows the distribution of Airbnb locations in London in 2019, 2020 and 2021.

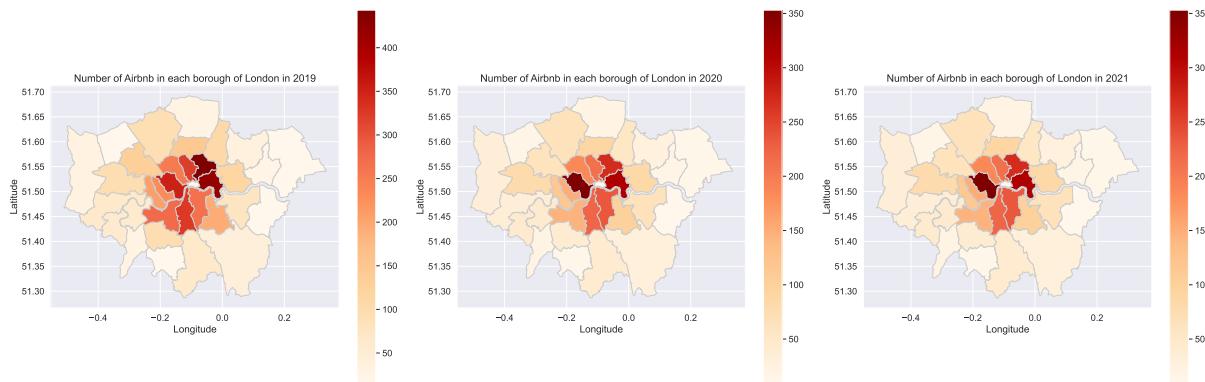
Density of Airbnb locations: In central areas, especially in the black part of the figure, the density of Airbnb is high, which usually indicates a high demand for travel or rentals in the area.

2020 is likely to be affected by the COVID-19 pandemic, leading to changes in patterns of travel and accommodation. So we continue to study

The number of airbin location had declined.

2 Number of Airbnb location for each neighborhood

- This code is used to create and display a map visualization of the number of Airbnb listings in different neighborhoods (or boroughs) of London for the years 2019, 2020, and 2021.
- The code begins by using the value_counts() function on the ‘neighbourhood_cleansed’ column of the listings_2019, listings_2020, and listings_2021 dataframes.
- The results are stored in the Number2019, Number2020, and Number2021 dataframes.
- These dataframes have two columns: ‘Neighborhood’ (neighborhood name) and ‘Number of Airbnb Listings’.



This graph shows the number of AirbnBs in London boroughs in 2019, 2020 and 2021. This is a heat map made by combining geographic information with Airbnb listing data, using different colors to indicate the number of AirbnBs in each borough. The depth of the scale represents the concentration of AirbnBs, and the darker the color, the more AirbnBs there are in the area.

From 2019 to 2021, the number of AirbnBs in central London has decreased. This could be due to policy changes, increased market saturation, or the impact of COVID-19.

Spatial distribution: The largest number of AirbnBs are concentrated in the central area and gradually decrease to the peripheral area.

Time trend: Over the three years, the distribution trend of Airbnb has remained consistent, with the central area always being the densest place.

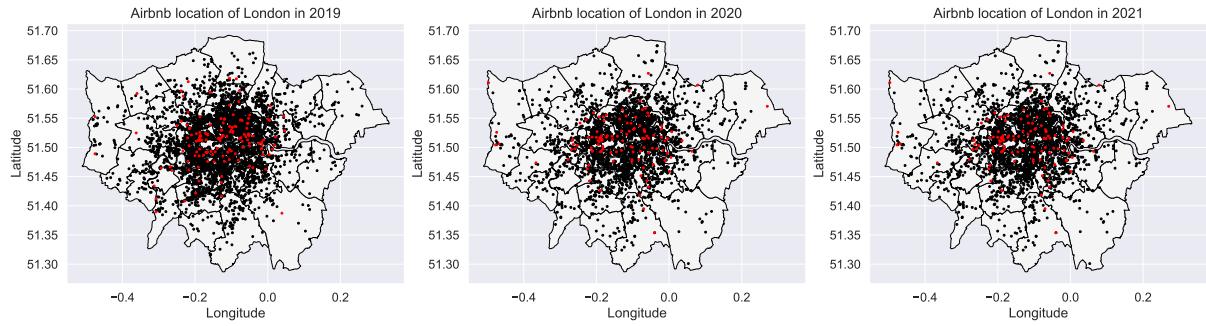
3 Minimum nights from 2019-2021

Add Airbnb location–minimal nights to London Map

Create and display maps of Airbnb property locations in London for the years 2019, 2020, and 2021, differentiating between short-term and long-term rental properties.

According to the Airbin website, the difference between a long lease and a short lease is whether it is greater than 30 days, if it is greater than 30 days, it is a long lease, if it is less than 30 days, it is a short lease.

Through this picture about three years, We can clearly see the distribution of airbin long and short rent in each borough of London.



This chart shows the distribution of Airbnb locations in London for 2019, 2020 and 2021, while distinguishing between short-term rentals (minimum rental terms of less than 30 nights, indicated by a black dot) and long-term rentals (minimum rental terms of 30 nights or more, indicated by a red dot).

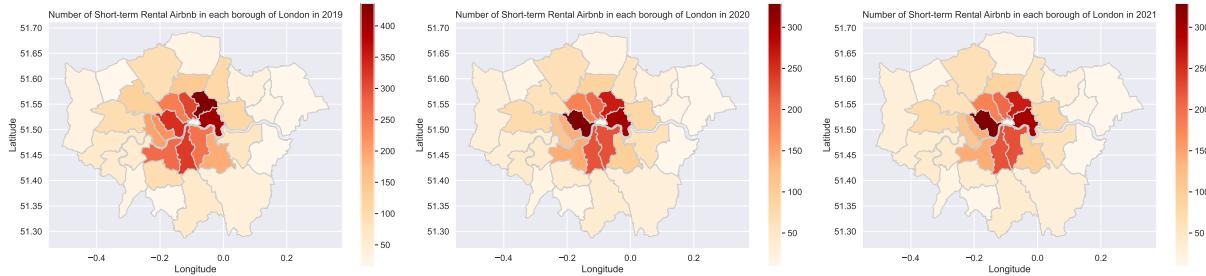
Concentration of short-term rentals: Over three years, short-term rentals (black dots) were highly concentrated in central London.

Rarity of long-term rentals: Long-term rentals (red dots) are relatively rare in London and are concentrated in central areas.

4 Number of Short-term and Long-term Rental Airbnb in each borough of London 2019-2021

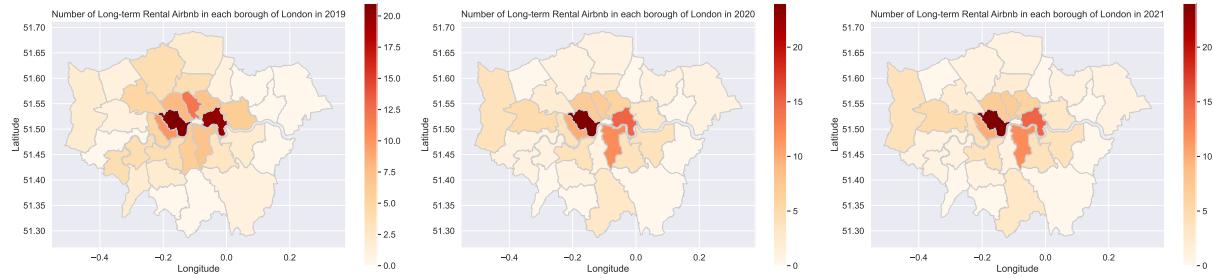
Visualize the number of short-term and long-term rental Airbnb listings in different boroughs of London for the years 2019, 2020, and 2021.

4.1 Number for short-term from 2019-2021



This picture shows that between 2019 and 2021, the number of Airbnb short-term listings in the central areas (the areas with the deepest color blocks) decreased significantly. In 2019, the central region had more than 400 listings, the highest concentration of listings. In 2020, the number of listings in the central area decreases, peaking at around 350. Through 2021, the number of listings in the central District continues to decrease, peaking below 175. Overall, the number of short-term Airbnb listings fell in most areas of London from 2019 to 2021

4.2 Number for long-term from 2019-2021



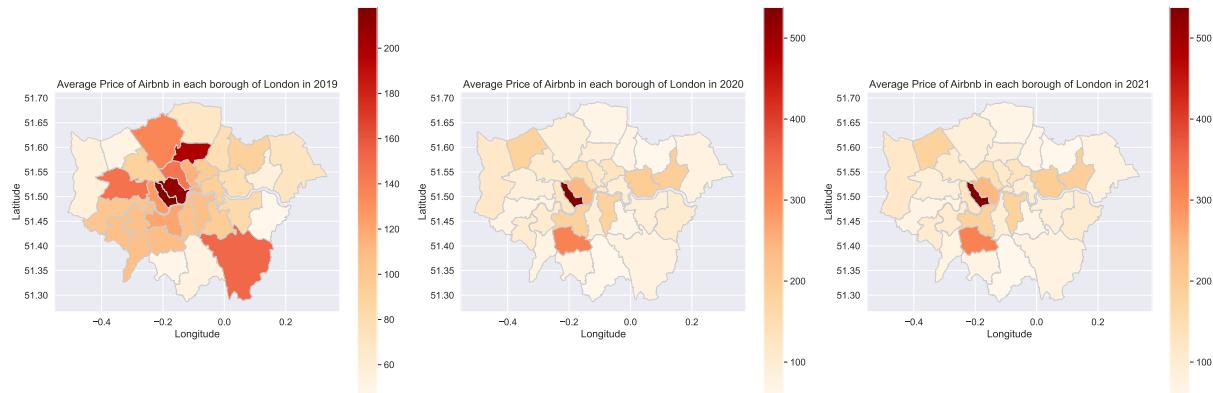
Central area concentration: Long-term rental Airbnb is mainly concentrated in central London and has remained so throughout the three years.

Diminishing numbers: From 2019 to 2021, the number of long-term Airbnb rentals in central London has decreased. The maximum number in 2019 was more than 20, while by 2021 it was reduced to less than 15.

Consistency of distribution: Although the number varied, the distribution pattern of long-term rental Airbnb remained consistent over the three years, concentrating in the central area and gradually decreasing to the periphery.

5 Calculate the average price for each neighborhood from 2019-2021

Calculates the average price of Airbnb listings in different neighborhoods for the years 2019, 2020, and 2021 using the groupby and mean functions. And merges these average price DataFrames with a GeoDataFrame called London based on the neighborhood information, creating new DataFrames named Neighborhood_price2019, Neighborhood_price2020, and Neighborhood_price2021.



In 2019, the average Airbnb price in some areas, such as the central District, was very high, reaching more than £200. By 2020, the average price in these high-value areas had fallen, with the highest price range around £180. In 2021, the high value area shrinks further and the price range drops further, with the maximum price not exceeding £500, but it is noted that the maximum range of the scale chart increases, indicating a greater difference in the price distribution.

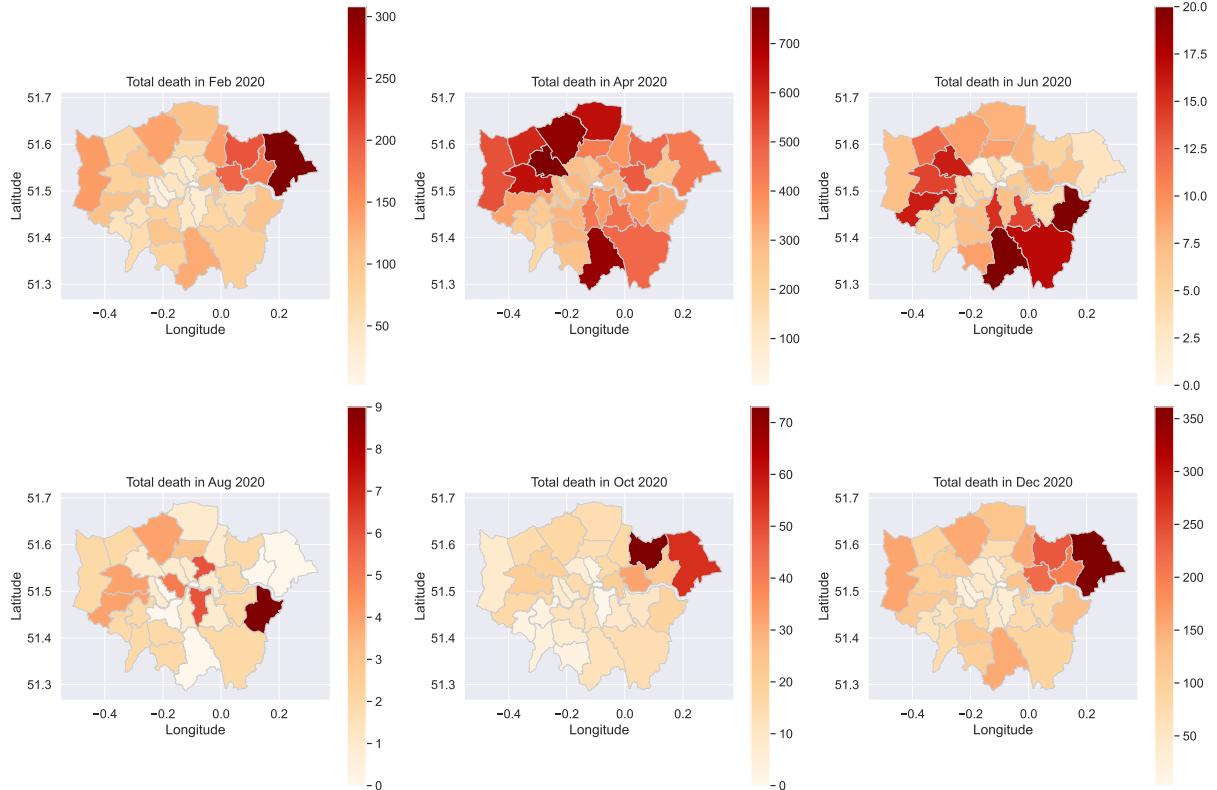
6 Total death in 2020

Covid-19 Death data in 2020

Covid-19 Infection data in 2020

Firstly, combine the death population in 2020 to a datafram

Visualize Covid-19 deaths in different boroughs or areas of London in different months of 2020.



In 2019, the average Airbnb price in some areas, such as the central District, was very high, reaching more than £200. By 2020, the average price in these high-value areas had fallen, with the highest price range around £180. In 2021, the high value area shrinks further and the price range drops further, with the maximum price not exceeding £500, but it is noted that the maximum range of the scale chart increases, indicating a greater difference in the price distribution. February and April: February 2020 saw a relatively low number of deaths in London, but by April there was a significant increase in deaths in certain areas, particularly the central and northern areas, showing a deep hue.

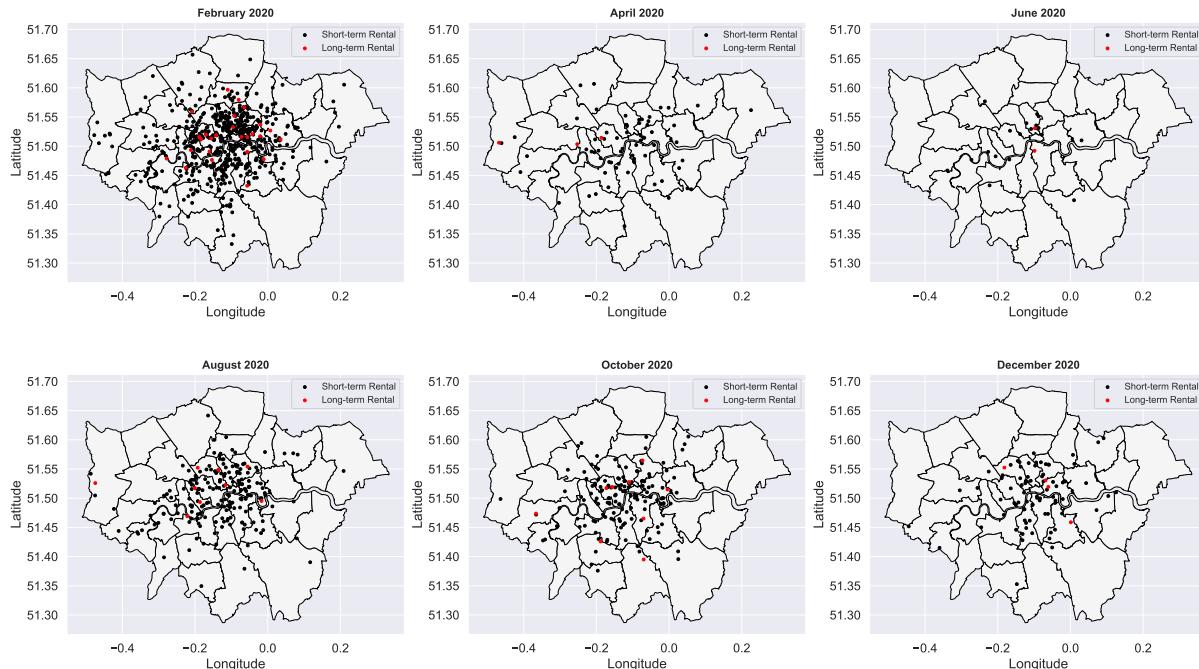
June and August: In June, the number of deaths peaked in several regions, especially the central region. By August, however, the number of deaths dropped significantly and the overall tone lightened.

October and December: The number of deaths increased in October, but not as high as in June. In December, the number of deaths increased again in several regions, but still not as high as in June.

7 Short-term and Long-term Rental Airbnb location of London in 2020 monthly

It first creates Point objects for Airbnb rental properties' longitude and latitude coordinates for each specific month. Then creates a GeoDataFrame (geo_df_2020_feb, geo_df_2020_apr, etc.) for each month's data. For each month, it plots a map of London using the London GeoDataFrame as the base map and overlays Airbnb rental properties. Short-term Rentals (minimum_nights < 30): Represented by black circles. Long-term Rentals (minimum_nights >= 30): Represented by red circles.

Short-term and Long-term Rental Airbnb locations in London for different months in 2020

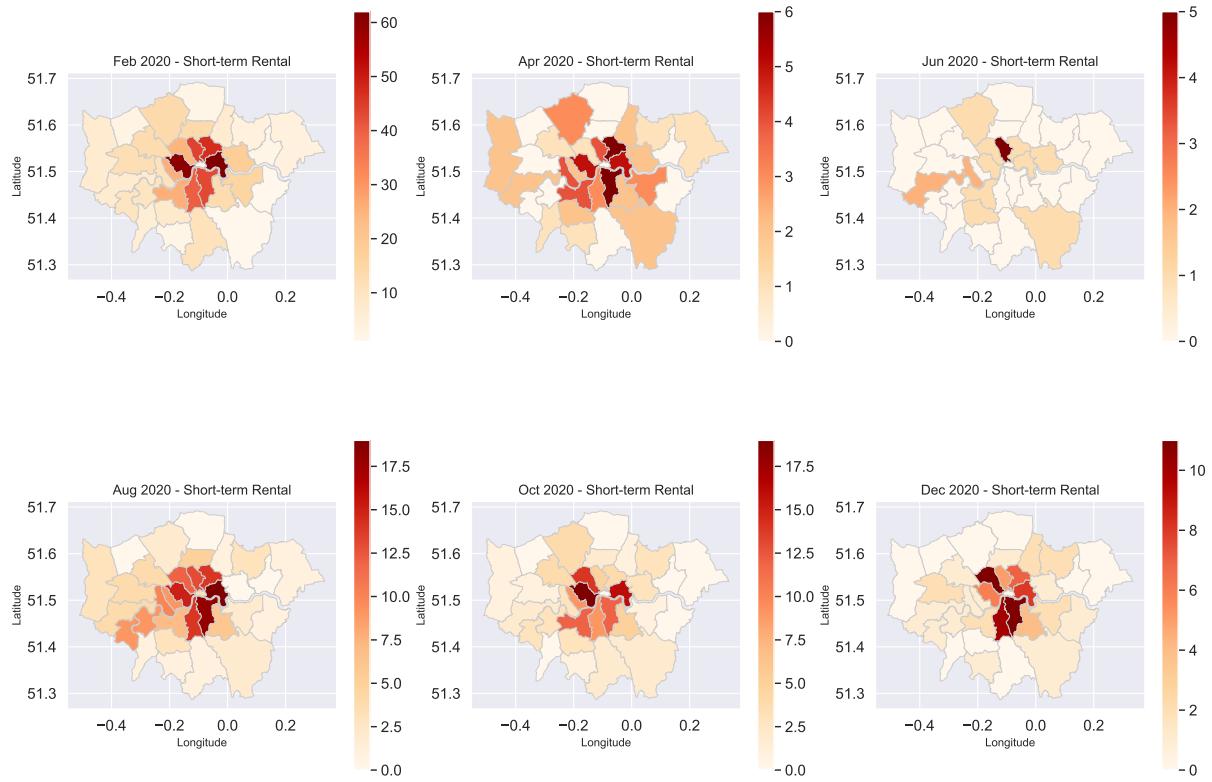


In February, the number and distribution of short-term rentals are relatively high, especially in the downtown areas. By April, the number of short-term rentals decreased significantly, which may be related to the impact of the COVID-19 pandemic, as many countries began to implement travel restrictions and lockdowns at this point in time. The June and August charts show that the number of short-term rentals has recovered somewhat, but remains below February levels. By October and December, the distribution of short-term rentals was thinning again, likely due to the ongoing impact of the pandemic and seasonal changes in travel patterns.

8 Number of Short-term and Long-term Rental Airbnb in each borough of London in 2020 monthly

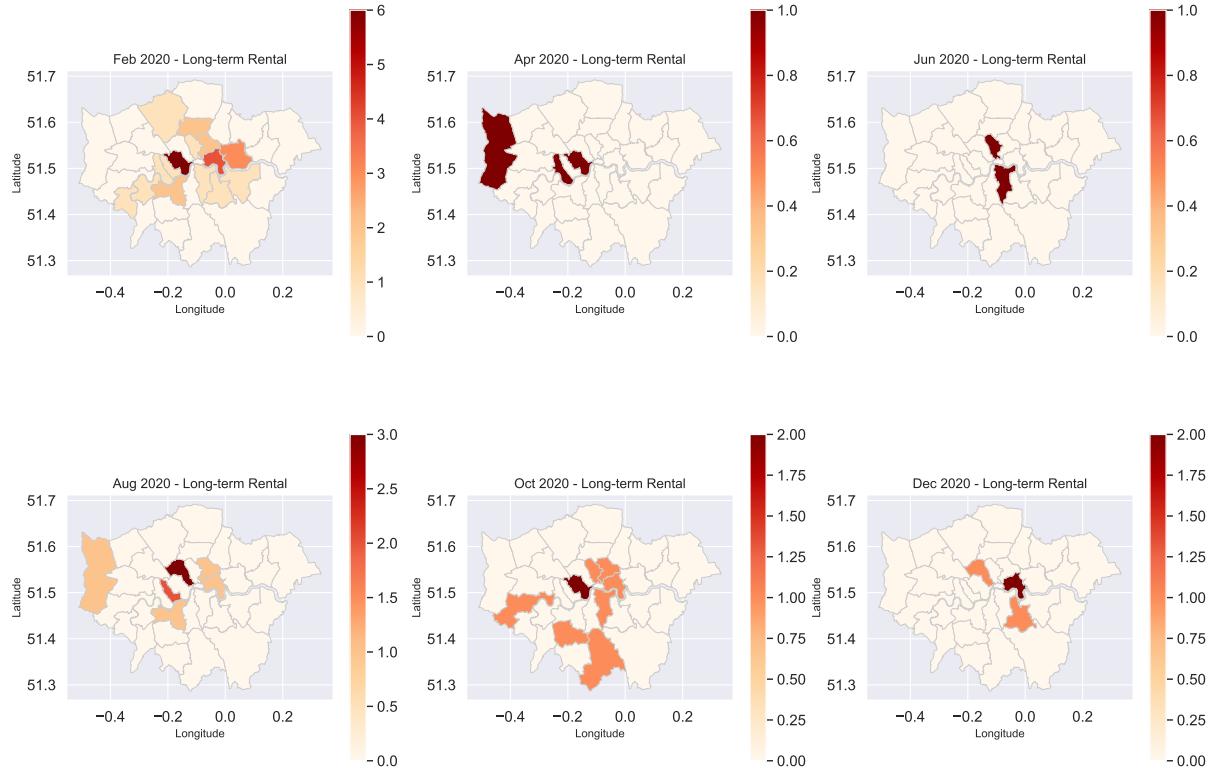
Rental data for different months are processed and combined to produce a dataset that includes geographic information and the number of short-term and long-term rentals

Visualize short-term rental data for different months



February: The number of short-term rental AirbnBs is higher and darker in certain central areas, especially the central areas. April and June: As we move into April and June, the number of AirbnBs in the central area decreases significantly and the overall map becomes lighter in tone, especially in June. August and October: By August and October, the number of AirbnBs in the central area had picked up, but not to the number seen in February. December: The number of AirbnBs in the central region decreased further in December, showing a decreasing trend.

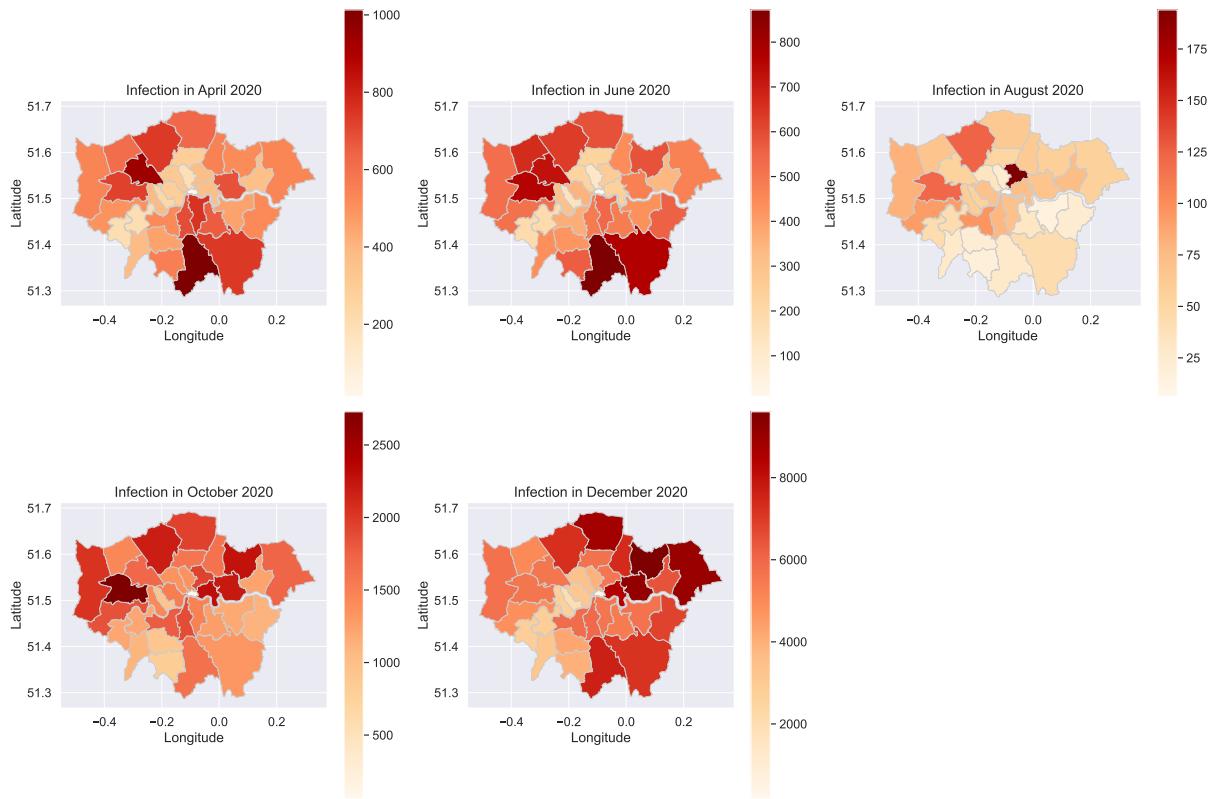
Visualize long-term rental data for different months



In February and April, certain central areas had a higher number of long-term rental Airbnb listings, showing a darker hue. By June, the number of these areas decreases and the overall map becomes lighter in tone. In August, the number of central areas picked up somewhat, but remained below the levels of February and April. In October and December, the number of long-term rental Airbnb listings increased in some areas, but the overall number remained low.

9 Infection in 2020

Combining data on COVID-19 infections in different communities in London with London map data, Geospatial Visualization was then used to show the distribution of COVID-19 infections across London neighborhoods at different time points (April, June, August, October and December 2020).

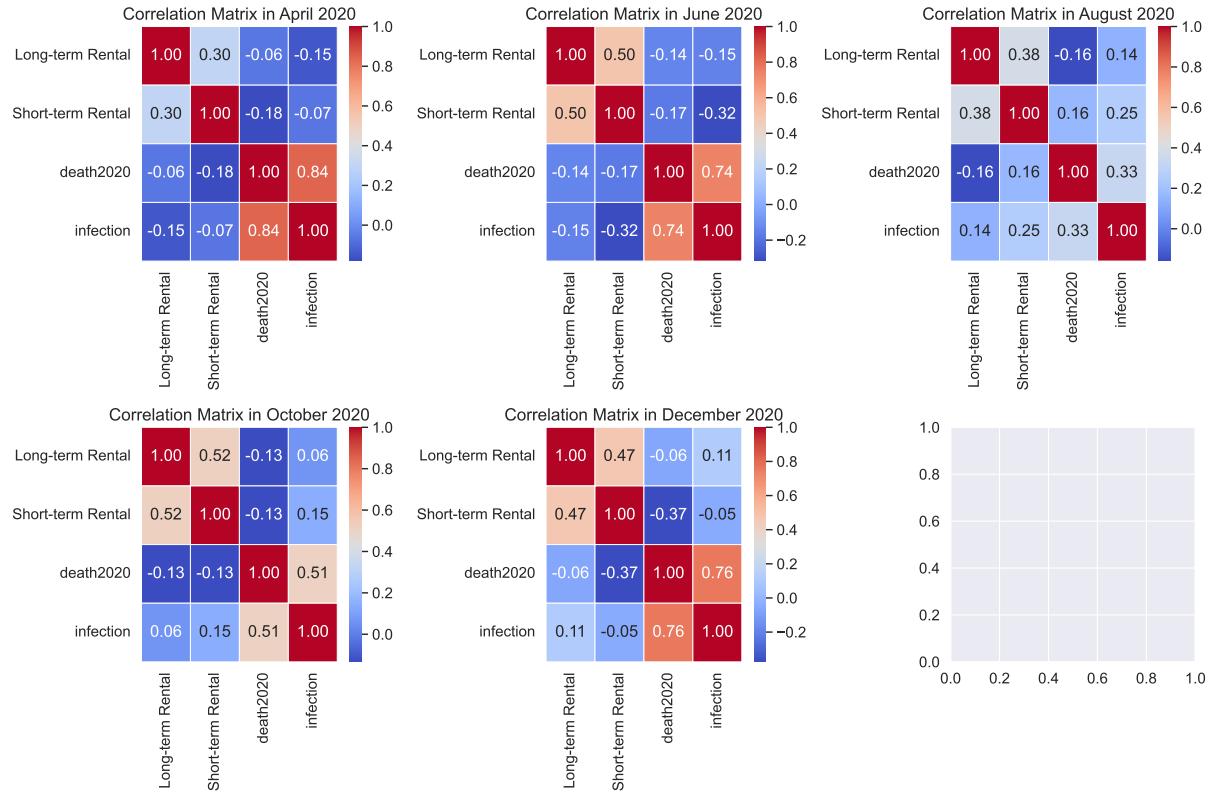


April - The infection is on the order of 200 to 800, with the most severe infection in the central area. June - There has been a decrease in infections, with most areas in the order of 100 to 600 infections. August - Further reduction in infections, with most regions in the order of 50 to 175, with a marked decline in the number of infections in the central region. October - The number of infections has rebounded, returning to between 1,500 and 2,500 infections in several regions. December - The number of infections rose sharply, with almost all regions experiencing infections of an order of magnitude between 2,000 and 8,000, showing a significant increase in infections.

7. Drawing on your previous answers, and supporting your response with evidence (e.g. figures, maps, and statistical analysis/models), how *could* this data set be used to inform the regulation of Short-Term Lets (STL) in London?

10 Correlation Matrix

Pull together different sets of Data, including Airbnb data, Death Data, and infection data. The absence of infection data from the February data was addressed and the infection data was merged into the combined results of the Airbnb data and the death data. Perform correlation analysis. For each month's data, it calculates correlations between Long-term Rental (long-term Rental), Short-term Rental (short-term rental), death2020 (death data), and infection (infection data), and visualizes the correlation matrix using heat maps. The title of each correlation matrix contains the corresponding month.



April - The infection is on the order of 200 to 800, with the most severe infection in the central area. June - There has been a decrease in infections, with most areas in the order of 100 to 600 infections. August - Further reduction in infections, with most regions in the order of 50 to 175, with a marked decline in the number of infections in the central region. October - The number of infections has rebounded, returning to between 1,500 and 2,500 infections in several regions. December - The number of infections rose sharply, with almost all regions experiencing infections of an order of magnitude between 2,000 and 8,000, showing a significant increase in infections. Relevance of the rental market:

Long-term and short-term rentals remained positively correlated throughout the time period, reaching the highest in October (0.52). This means that these two types of rental markets may be affected by the same market forces, or their performance in the market may be correlated. Relationship between rental market and mortality:

The correlation between long - and short-term rentals and deaths in 2020 showed different patterns over different months, but for the most part, these values were negative or close to zero, suggesting that there is no strong direct relationship between the rental market and mortality. The relationship between rental market and the number of infections:

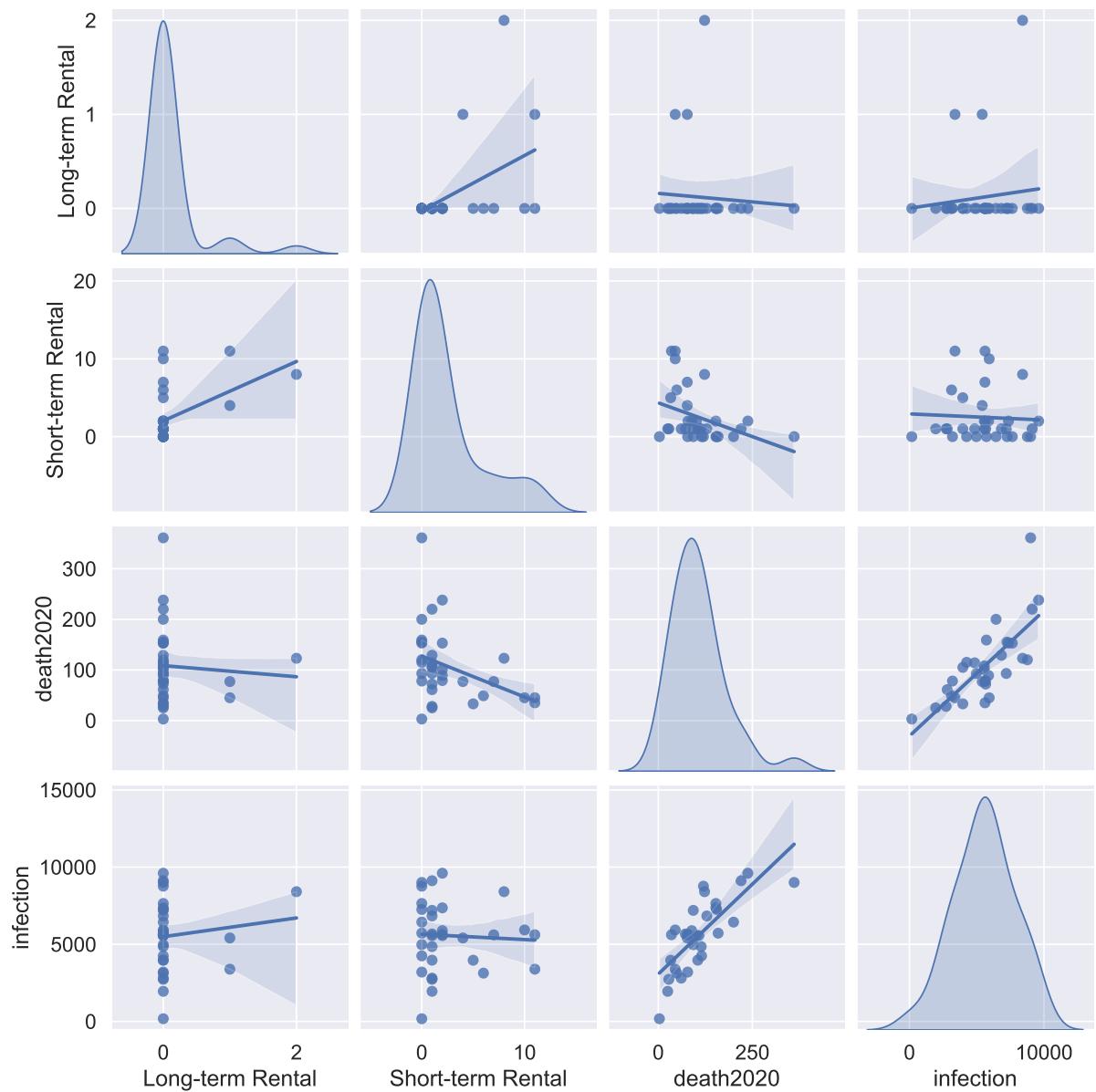
The correlation between long-term rentals and the number of infections varied little from month to month, sometimes positive (0.14 in August and 0.06 in October) and sometimes negative (-0.15 in April and -0.15 in June), which may indicate that there is not a consistent trend or relationship. Short-term rentals showed a strong negative correlation (-0.32) with the number of infections in June, but little correlation in other months. Relation between deaths and infections:

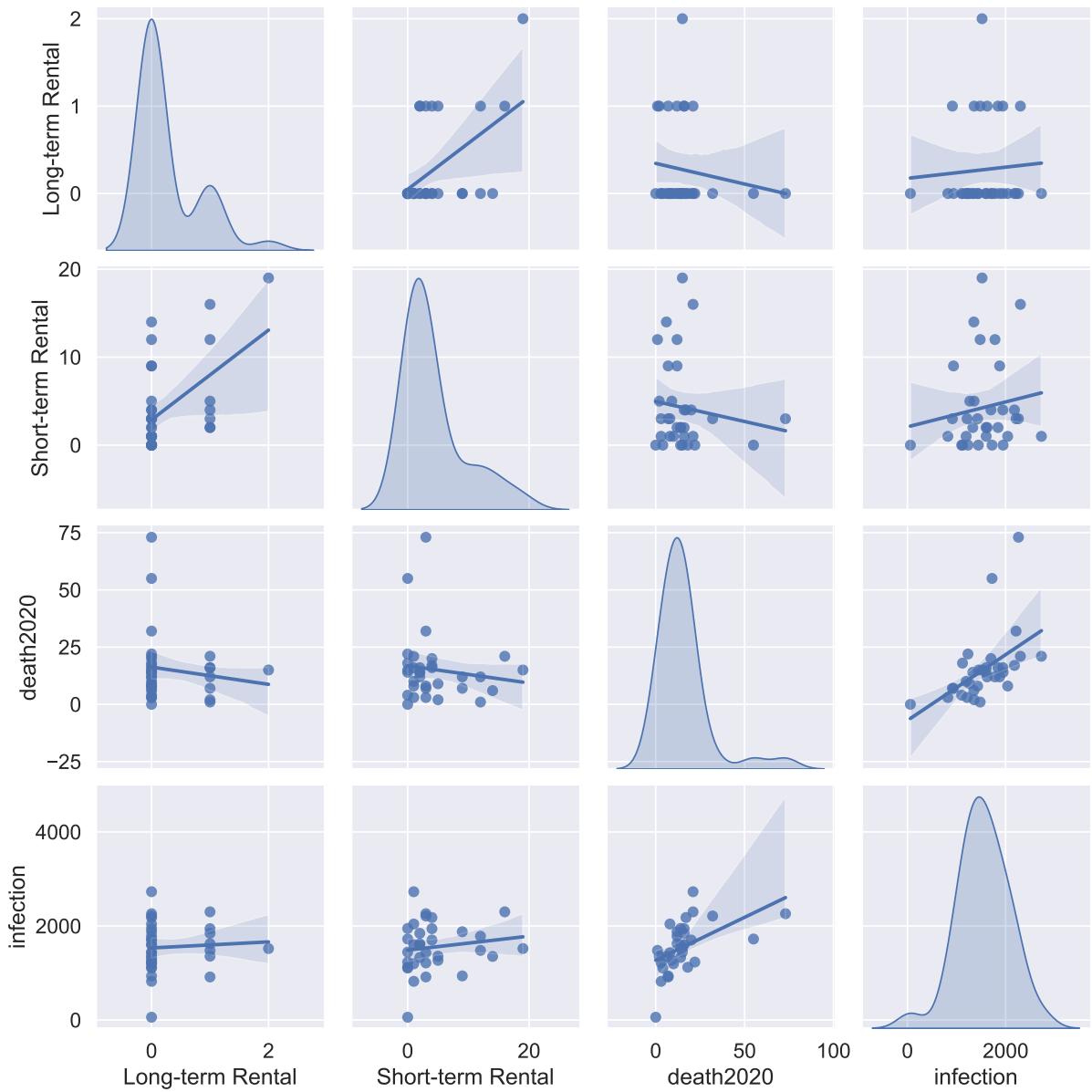
There was a strong positive correlation between deaths and infections in April and June (0.84 and 0.74, respectively), but this correlation dropped to 0.76 in December. Despite the decrease, it still indicates that the increase in the number of infections is in sync with the increase in the number of deaths. Time trend:

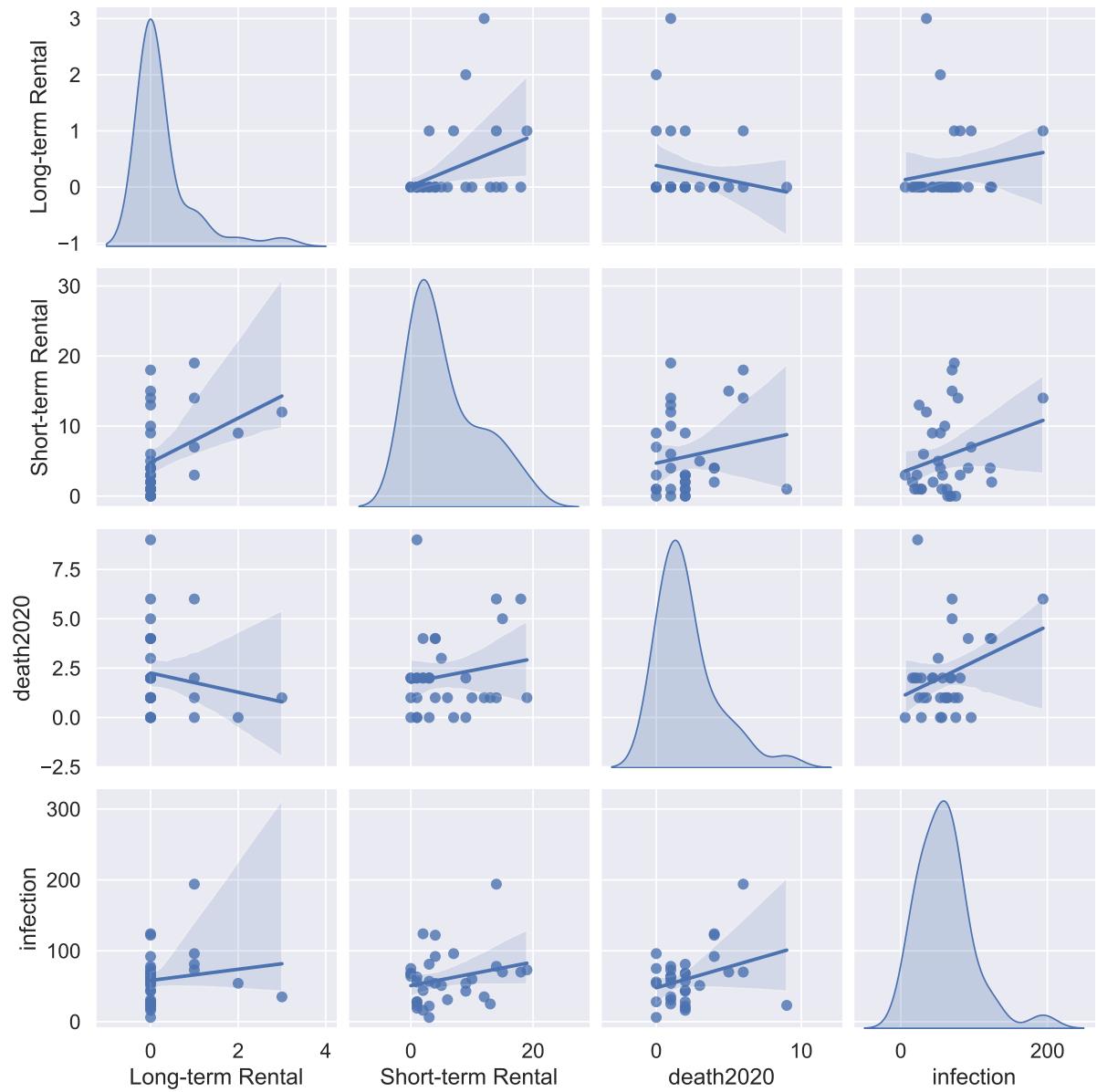
In April and June, the correlations between long-term and short-term rentals and the number of infections were negative or close to zero, while by August, these correlations turned positive, possibly reflecting a possible change in the market's response to the outbreak over time.

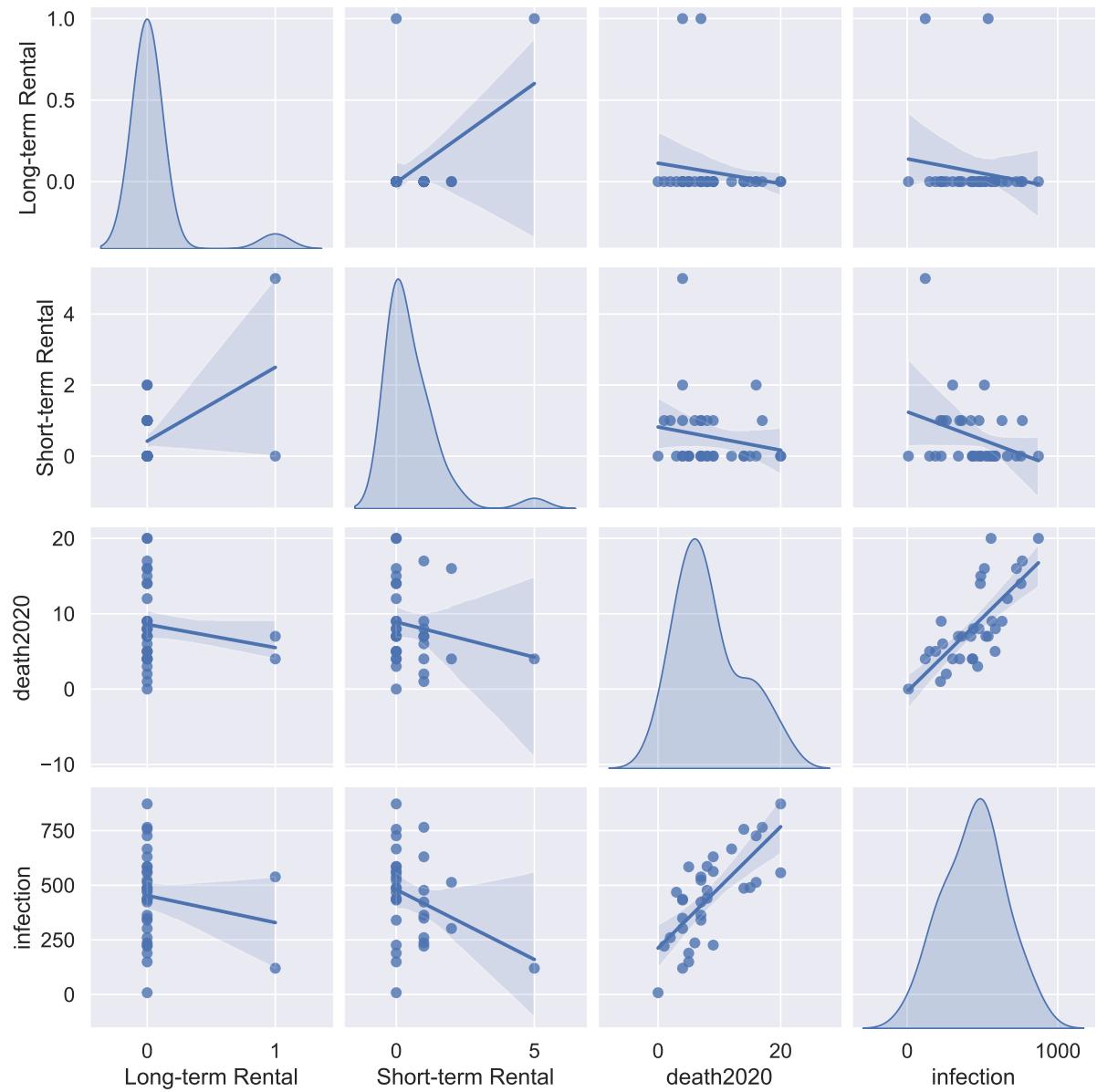
11 Multivariable Joint distribution

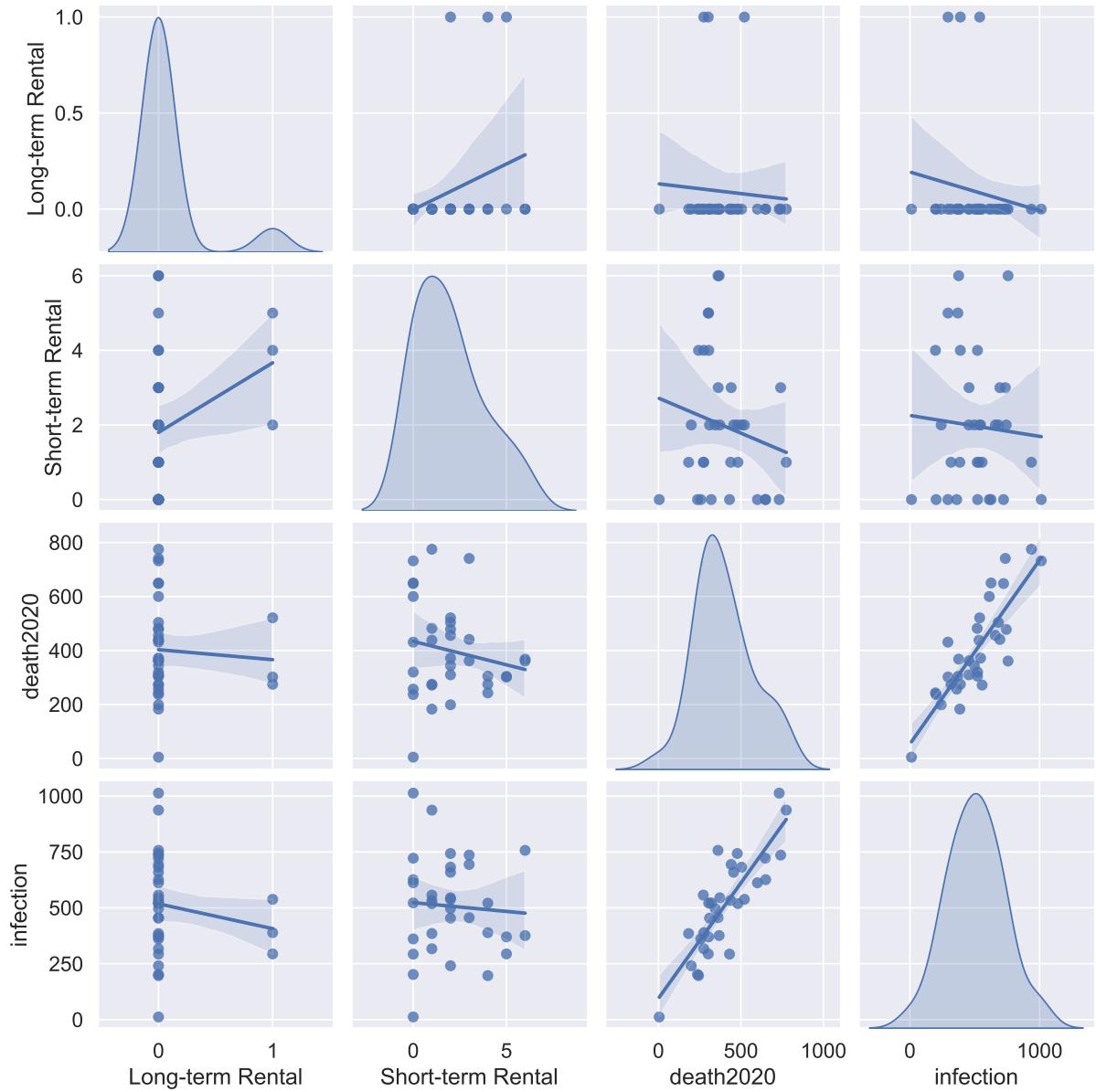
Create a series of scatter plots and Kernel Density Estimates (KDE) to study the relationships and distributions between different variables.











Relationship between long-term and short-term leases: The scatter plot shows that there is no obvious linear relationship between long-term and short-term leases, which means that the two leasing patterns may be relatively independent.

Relationship between long-term leases and deaths and infections: The number of long-term leases does not appear to have a clear linear relationship with the number of deaths and infections. This suggests that the number of long-term leases may not be the main factor influencing the number of COVID-19 deaths and infections.

Relationship between short-term rentals and deaths and infections: The chart shows that there is also no clear linear relationship between the number of short-term rentals and the number of deaths and infections, although the regression line shows a slight positive or negative trend, the distribution of data points suggests that the relationship is not strong.

Relationship between deaths and infections: It is expected that there is a positive correlation between the number of deaths and the number of infections, and the scatter plot and regression line show that as the number of infections increases, so does the number of deaths.

References

- Boyd, D. and Crawford, K. (2012) ‘CRITICAL QUESTIONS FOR BIG DATA’, *Information, Communication & Society*, 15(5), pp. 662–679. doi: [10.1080/1369118X.2012.678878](https://doi.org/10.1080/1369118X.2012.678878).
- Edelman, B. G. and Geradin, D. (2016) ‘Efficiencies and regulatory shortcuts: How should we regulate companies like airbnb and uber?’, *Stanford Technology Law Review*, 19, pp. 293–328.
- Ert, E., Fleischer, A. and Magen, N. (2016) ‘Trust and reputation in the sharing economy: The role of personal photos in airbnb’, *Tourism Management*, 55, pp. 62–73.
- Ferreri, M. and Sanyal, R. (2018) ‘Platform economies and urban planning: Airbnb and regulated deregulation in london’, *Urban Studies*, 55(15), pp. 3353–3368.
- Guttentag, D. (2015) ‘Airbnb: Disruptive innovation and the rise of an informal tourism accommodation sector’, *Current Issues in Tourism*, 18(12), pp. 1192–1217.
- Newell, S. and Marabelli, M. (2015) ‘Strategic opportunities (and challenges) of algorithmic decision-making: A call for action on the long-term societal effects of ’datification”’, *The Journal of Strategic Information Systems*, 24(1), pp. 3–14.
- Zook, M. *et al.* (2017) ‘Ten simple rules for responsible big data research’, *PLOS Computational Biology*, 13(3), p. e1005399.