London Airbnb Rental Market Report — Based on the Analysis of STL

The Stapleton Five

Declaration of Authorship

We, [The Stapleton Five], pledge our honour 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:16/12/2024

Student Numbers: 24107824, 24057654, 24054682, 24070831, 24153654

Brief Group Reflection

What Went Well	What Was Challenging
A Literature collection and review were smooth. We gained a good understanding of Insitebnb from the literature and identified appropriate references to support our arguments.	B Initially, dynamic maps failed to render successfully.
C Successfully applied statistical and spatial analysis methods.	D It was challenging to determine a clear theme amidst a large volume of figures and maps.

Priorities for Feedback

1. Whether our written explanations are clear and the structure is appropriate, and suggestions on areas for improvement.
2. Whether each analytical method used in the code has been correctly applied.

1. Who collected the InsideAirbnb data?

Murray Cox founded InsideAirbnb in 2015 and began collecting Airbnb data (PuckLo, 2018). The website now has collaborators, new members, and an advisory board to support project development and data collection.

2. Why did they collect the InsideAirbnb data?

They collected the InsideAirbnb data to address social concerns about Airbnb's impact on affordable housing and gentrification (Katz, 2017). In 2015, they discovered that Airbnb data could be misleading, hiding the true scale of its effects on communities. The purpose of collecting Inside Airbnb data is to provide transparency, advocate for community housing, and protect the rights of tenants and small landlords.

3. How was the InsideAirbnb data collected?

The data is collected through python scripts, which are sourced from Github and other online resource (Alsudais, 2021), each listing page on the website.

The data collection can be divided into the following stages: Firstly, finding all public listings on the Airbnb website as much as possible. Aditionally, accessing each page through the script to collect information such as the ID of the listing, the type of home, the time it was published, the number of reviews, and the location. Finally, aggregating the data and verifying it with the number of data published by Airbnb. They use the number of reviews as the number of visits to the listing for they lack internal data (Cox and Slee, 2016).

They state that no private information is used in the data collection process and that the names of the listings are compiled by comparing the geographical coordinates of the listings with the city-to-community definitions. Furthermore, in the later development, 50% of the comment rate was converted to an estimated booking and an average length of stay was assigned to each city (Airbnb, n.d.c).

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

It is obvious that the main data source of IA (Inside Airbnb) lacks accuracy, as it is obtained by directly scraping data collected by Airbnb itself from its website (Cox and Slee, 2016). In addition, other data sources for the IA dataset are also unreliable, which are related to the data collection methods of Python scripts. Specifically, Python scripts are likely to generate IA datasets by copying scripts that lack responsible maintenance on any websites (Airbnb, n.d.b), with GitHub being an example (Alsudais, 2021). Moreover, the automated generation of code for data collection services for IA datasets may encounter issues such as ignoring list types and making incorrect links (Alsudais, 2021), which can lead to problems include duplicating data collection, data position errors, and data omissions. For example, automated code may simplify or truncate some property review fields.

Overall, the IA dataset covers a wide range of data, such as various housing types and different landlord characteristics. This is beneficial for the process of data research. However, unreliable data sources and automatically generated codes affect the research results of IA. In addition, due to the setting of monthly updates (Alsudais, 2021), the dataset of IA is not updated in real time, which brings the problem that the existing dataset cannot represent the real data. As a result, the IA dataset can only representbthe process it wishes to study to some extent. Specifically, only when the research scope, problem definition, time constraints, and data processing methods are clearly defined, can the IA dataset effectively reflect analytical research.

These limitations have resulted in serious consequences. For reseachers who use inaccurate IA datasets for research, the reproducibility of their research results will be doubted. Additionally, as the errors in the IA dataset increase, the public's trust in online information will decrease accordingly. Furthermore, as an important reference, if the IA dataset lacks accuracy, relevant political and economic decisions may be misled.

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

We analyze its ethical issues through four stages: purpose of data use data collection, data storage, use and impact of analysis results.

5.1 Purpose of Data Use

- The transparency of InsideAirbnb itself: InsideAirbnb's mission statement is to protect the city from short-term rentals (Airbnb, n.d.a). InsideAirbnb has a bias when analyzing data, which may reduce the objectivity of the results and lead to inaccuracy.

5.2 Data Collection

- Data source: InsideAirbnb says it's not endorsed by Airbnb (Airbnb, n.d.c). Airbnb expressly states that it prohibits using automated means to access or collect data from the Airbnb Platform. But whether InsideAirbnb collects its data through Python scripts may involve legal and ethical controversies that should be considered.
- Data privacy and security: Being able to access or collect data does not mean that it is ethical to use that data (Boyd, Levy and Marwick, 2014). Although InsideAirbnb claims in its disclaimer that the data is safe and full-protected, it is risky that it involves sensitive content such as names, photos, locations and reviews, which may indirectly expose individuals' privacy, and trigger ethical judging of data collection and privacy security.

5.3 Data storage

- Fairness of data access: InsideAirbnb provides the most recent 12 months of data for free, but access to archived data is subject to review or even payment (Airbnb, n.d.c), which may be a hindrance to researchers or organizations with limited resources.

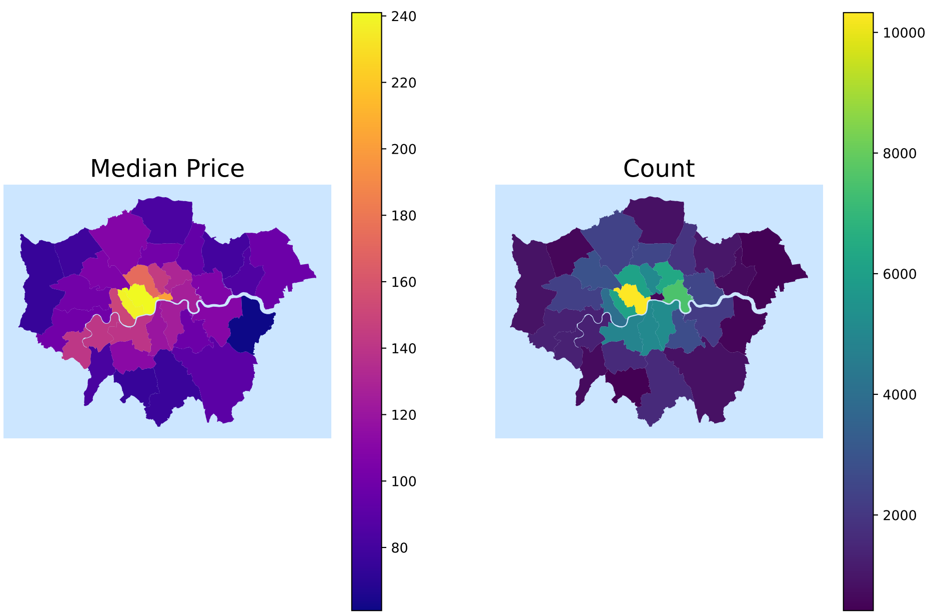
5.4 Usage and Impacts of Analysis Results

- A risk of misuse: The analysis results derived from public data may be influenced by the researcher's position to support unfair policies or biased conclusions.
- Unfair business competition: conclusions and analyses based on InsideAirbnb data could be exploited by competitors in the same industry to harm Airbnb or landlords engaged in lawful operations.
- Challenges of public distrust: deepening public distrust and doubt about the sharing economy, inducing people to overlook the economic benefits that platforms like Airbnb may bring to certain landlords and tenants.

6. With reference to the InsideAirbnb 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?

6.1 Spatial agglomeration

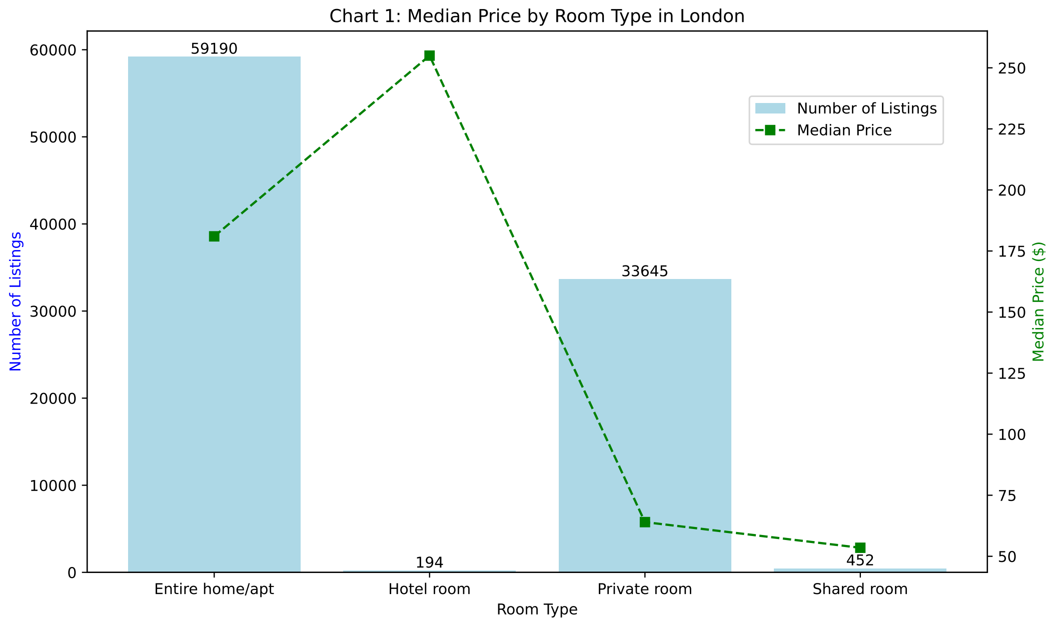
Figure 1: Distribution of Listings and its Price by Boroughs



As shown in the figure 1, the distribution of the price and count of Airbnb listings in London shows a significant spatial clustering, with the highest values in the city center (along the River Thames), while lower values in the outskirts of the city, showing a decreasing spatial distribution pattern from the center.

6.2 Market-Oriented

Landlord market-oriented: The influence of professional hosts is significant. Among all kinds of landlords, although 81% of single-unit landlords have an absolute advantage in terms of quantity, they only control 49% of the units. By comparison, only 19% of professional hosts, who are landlords with multiple properties (Li, Moreno and Zhang, 2016), control 51% of properties.



Property market-oriented: chart 1 shows that the proportion of properties of essential rental property types is the highest, accounting for 63.32%. When the housing is rented out in the form of rooms, its main function is still residential, which does not affect the supply of the housing rental market (Chang, 2020). However, as a large number of landlords place their entire properties in the Short-term Lets market, their properties shift towards commercial attributes.

6.3 Primarily for Short-Term Lets

According to the Deregulation Act 2015, Short-term Lets (STL) is defined as "The total number of nights of use does not exceed 90 nights in a calendar year" (UK Parliament, 2015). Based on InsideAirbnb data, 97.3% of listings in London are classified as STL, indicating that the Airbnb market is largely made up of STL. Therefore, the analysis focuses on the spatial distribution characteristics and types of STL (InsideAirbnb, 2024).

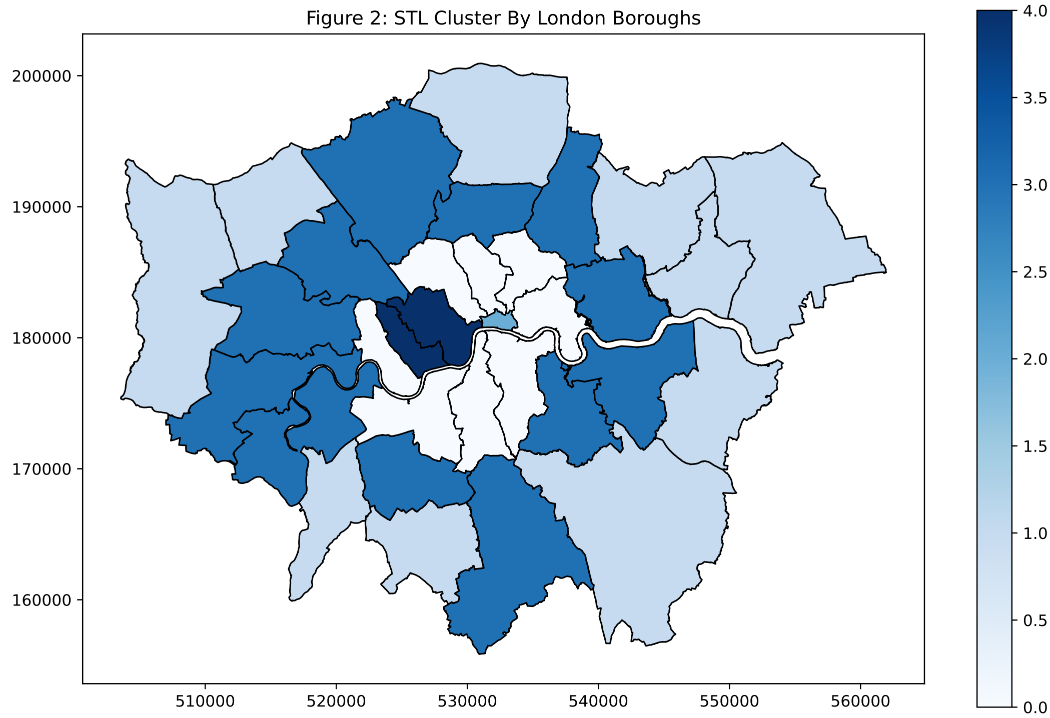
6.4 Summary

The Airbnb rental market in London has clear characteristics of spatial agglomeration, market-oriented, and short-term leasing. This market can meet short-term accommodation needs. However, it has potential negative impacts on the long-term rental market. The government needs to introduce policies to strengthen the regulation of the STL market.

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

Based on the clustering analysis results of the InsideAirbnb dataset, this section explores the characteristics of STL markets in different boroughs of London clusters. Then, several regulatory recommendations were put forward to promote the sustainable and multi-center development of the STL market.

Figure 2: STL Cluster By London Boroughs



Using the standardized variables "Entire home/apt," "Private room," and "median price," London is divided into 5 clusters (Figure 2). The central areas (Cluster 4, Cluster 2, and Cluster 0) are high-density short-term rental areas, dominated by entire home listings with high prices. The outer areas (Cluster 1 and Cluster 3) are low-density short-term rental areas with limited housing supply. According to these characteristics, the following regulatory suggestions can be made.

7.1 Price Regulation

High-density short-term rental areas (central areas): strictly enforce the 90-day rental limit to prevent housing resources from being taken away from long-term lets. Apply tiered taxes based on rental prices to stabilize housing costs.

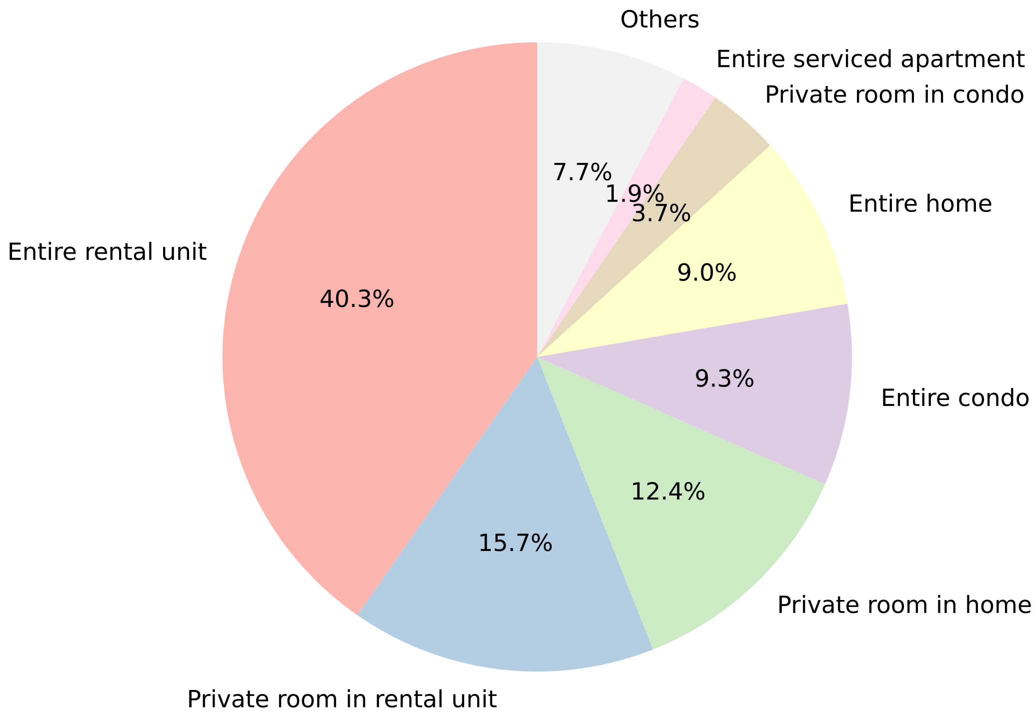
Low-density short-term rental areas (outer areas): introduce incentives to reduce housing pressure in the central areas. For example, offer tax benefits or subsidies to attract more landlords to the outer areas and create a multi-center short-term lets market in London.

7.2 Area-Based Regulation

Restricting entire rentals in high-density areas: As there is a squeeze on the long-term housing market due to STL of entire properties (chart 2), stricter policies on STL of entire properties should be implemented in areas with dense housing stock and high rental pressures (e.g. central London), such as limiting the number of days or properties of STL.

Considering the distribution of regional STL properties, implement differentiated regulatory policies and impose different day limits for areas with different densities.

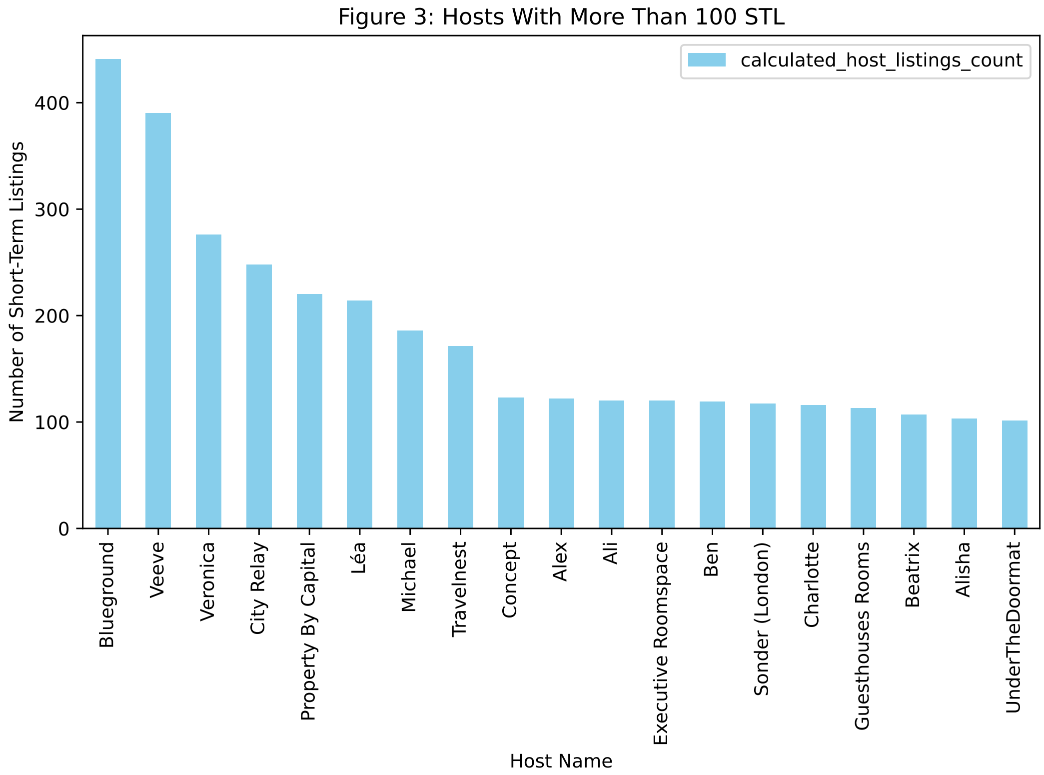
Chart 2: STL Property Type Pie Chart



7.3 Special supervision for professional hosts

Figure 3 shows that a small number of professional hosts have a large number of STL properties and extremely high market share, which may be related to their commercial operations. In addition, professional hosts have lower management marginal costs because of operating multiple properties simultaneously (Xie, Heo and Mao, 2021), which will give them a more advantageous competitive position in the market, ultimately leading to unfair competition.

In response to this phenomenon, it is recommended to implement special supervision for professional hosts:



- Limit on the Number of Listings: Set limitations on the number of STL a host can operate. Implement stricter regulations on hosts with multiple properties to prevent professional hosts from dominating the market.
- Implementation of Permitting Mechanism: Require hosts with multiple properties to apply for a commercial permit to ensure legality and compliance.
- Additional Tax Policies: Implement a progressive tax system based on the number of STL properties owned by a host, where the more properties a host operates, the higher the applicable tax rate. Additionally, impose urban development tax and housing security surcharge on professional hosts, using the revenue to fund Long-Term Lets supply and community welfare programs.

References

Airbnb, I. (n.d.a) 'About inside airbnb'. Available at: http://insideairbnb.com/about.html.

Airbnb, I. (n.d.b) 'Behind the story'. Available at: http://insideairbnb.com/behind.html.

Airbnb, I. (n.d.c) 'Data assumptions'. Available at: https://insideairbnb.com/data-assumptions/.

Alsudais, A. (2021) 'Incorrect data in the widely used inside airbnb dataset', *Decision Support Systems*, 141, p. 113453. doi: 10.1016/j.dss.2020.113453.

Boyd, D., Levy, K. and Marwick, A. E. (2014) 'The networked nature of algorithmic discrimination', *Data and discrimination: Collected essays*, pp. 43–57.

Chang, H. H. (2020) 'Does the room sharing business model disrupt housing markets? Empirical evidence of airbnb in taiwan', *Journal of Housing Economics*, 49, p. 101706.

Cox, M. and Slee, T. (2016) 'How airbnb's data hid the facts in new york city'. Available at: https://insideairbnb.com/research/how-airbnb-hid-the-facts-in-nyc/.

InsideAirbnb (2024) 'London airbnb listings'. https://insideairbnb.com/london/.

Katz, M. (2017) 'A lone data whiz is fighting airbnb — and winning'. Available at: https://www.wired.com/2017/02/a-lone-data-whiz-is-fighting-airbnb-and-winning/.

Li, J., Moreno, A. and Zhang, D. (2016) *Pros vs joes: Agent pricing behavior in the sharing economy.* Working Paper 1298. Ross School of Business.

PuckLo (2018) 'Practitioner profile: Murray cox'. Available at: https://morethancode.cc/2018/05/30/practitioner-profile-murray-cox.html.

UK Parliament (2015) *Deregulation act 2015.* London: Her Majesty's Stationery Office.

Xie, K., Heo, C. Y. and Mao, Z. E. (2021) 'Do professional hosts matter? Evidence from multi-listing and full-time hosts in airbnb', *Journal of Hospitality and Tourism Management*, 47, pp. 413–421.