

PANA's Group Project

Declaration of Authorship

We, **PANA**, 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:

24130341

24213583

24112862

24085798

24089558

Brief Group Reflection

What Went Well	What Was Challenging
Our group worked efficiently by starting early, breaking tasks into smaller parts, and designating responsibilities, which ensured smooth communication and time management.	Addressing weaknesses in the data, such as using the “last reviewed” column for yearly estimates, and managing word count distribution.
Decision-making was effective through voting, and collaboration was enhanced by using shared tools like Google Sheets.	Differences in coding styles and logic among team members required careful integration, and earlier data processing choices often had a cascading impact on later tasks.
These approaches left ample time for final adjustments, resulting in a well-prepared submission.	Rendering the Quarto markdown file to PDF was hindered by unresolved font issues.

Priorities for Feedback

Are there any areas on which you would appreciate more detailed feedback if we're able to offer it?

1. Word Count Allocation: The team faced challenges in distributing word counts across open-ended questions, as some required more detailed discussion than initially anticipated.
2. Task Focus: Identifying which sections needed more attention in hindsight could improve the depth of analysis in future projects.
3. Time Management: Feedback on managing time more effectively would be valuable for better coordination.
4. Task Prioritization: Insights on prioritizing tasks would help optimize resource allocation and streamline workflows in future projects.

Response to Questions

1. Who collected the InsideAirbnb data?

The data is primarily collected by founders Murray Cox and Tom Slee. Murray and Tom, who are respectively a community artist/ activist and a researcher. Key collaborators that may have also contributed to collecting data include Taylor Higgins, Alice Corona and Michael “Ziggy” Mintz (InsideAirbnb, 2016).

2. Why did they collect the InsideAirbnb data?

Cox was motivated to start collecting the data in 2014 while studying gentrification in Brooklyn, the project expanded to analyse Airbnb’s hidden effects in American cities. Cox and Slee, aligned in opposing Airbnb’s portrayal of public data, gathered InsideAirbnb data to expose illegal renting and its negative housing market impacts to reveal the “lies” in Airbnb’s claims, with data providing evidence of scofflaws (Katz, 2017). The data was collected to counter biased narratives, offering independent, open-access insights beyond Airbnb’s statistics or commercial providers.

3. How did they collect it?

Automated web scraping was used to collect the data where scripts extract information from Airbnb’s website by parsing its HTML content. Collaborator “Ziggy” automated the scraping process via python scripts to collect snapshots of public data through the Airbnb website periodically, replacing older data for each location (Alsudais, 2021). Cox compiled, analysed, and built the platform where the first stage browsed the Airbnb website to look for all listings within a city and then in the second stage, the program would visit the page of each listing to collect its detailed information (Cox and Slee, 2016).

4. How does the method of collection (Q3) impact the completeness and/or accuracy of the InsideAirbnb data? How well does it represent the process it seeks to study, and what wider issues does this raise?

An automated web scraping, affects both the completeness and accuracy of its dataset. By capturing a snapshot of publicly available information at a specific moment, it overlooks temporal factors such as seasonal trends. Modifications to listings after data collection, the inability to differentiate between booked and unavailable nights, and Airbnb’s anonymisation policies—which can shift locations by up to 150 metres—skew key metrics like availability and occupancy rates.

Reliance on web scraping introduces further limitations, including assumptions about review-to-booking ratios or average stay durations, which may not reflect actual behaviours. Mislinked data also highlights the risks of inaccuracies, for example, a LA property erroneously connected to a Tokyo experience (Alsudais, 2021).

InsideAirbnb’s reliance on Airbnb’s platform introduces vulnerabilities, including periodic anonymisation of location data, which compromises reproducibility. Reliance on Airbnb’s website necessitates ongoing monitoring to maintain data reliability. The absence of Airbnb’s raw data exacerbates these limitations, restricting access to detailed booking records, guest demographics, and host behaviour over time.

Without this proprietary data, researchers face challenges in conducting comprehensive analyses of pricing strategies, socio-economic impacts, and market trends and the lack of granularity can result in incomplete or misleading findings when applied to housing policy or urban planning (Alsudais, 2021).

Despite its widespread use, the dataset has undergone limited independent validation, raising concerns about its reliability and broader impact. Ethical issues regarding data accessibility, privacy, and validity further complicate its application in urban development and policy-making, emphasising the need for robust validation and transparency in data collection.

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

The ethical implications of using InsideAirbnb data stem from concerns about data quality, social impacts, and legal considerations.

Reliance on outdated Python scripts for web scraping undermines reliability. Sourced from GitHub, these scripts face accuracy issues due to Airbnb's layout changes (Slee, 2017). The dynamic nature of Airbnb's platform compounds these issues, as deleted listings create incomplete datasets, compromising insights for urban planning and housing policy. Fields like last_review may reflect inactive listings, skewing analyses. Furthermore, limited transparency in extraction processes hinders verification, raising concerns about validity and reproducibility. The periodic datasets ensure currency but complicates the studies replication when earlier versions are unavailable or unspecified (Alsudais, 2021).

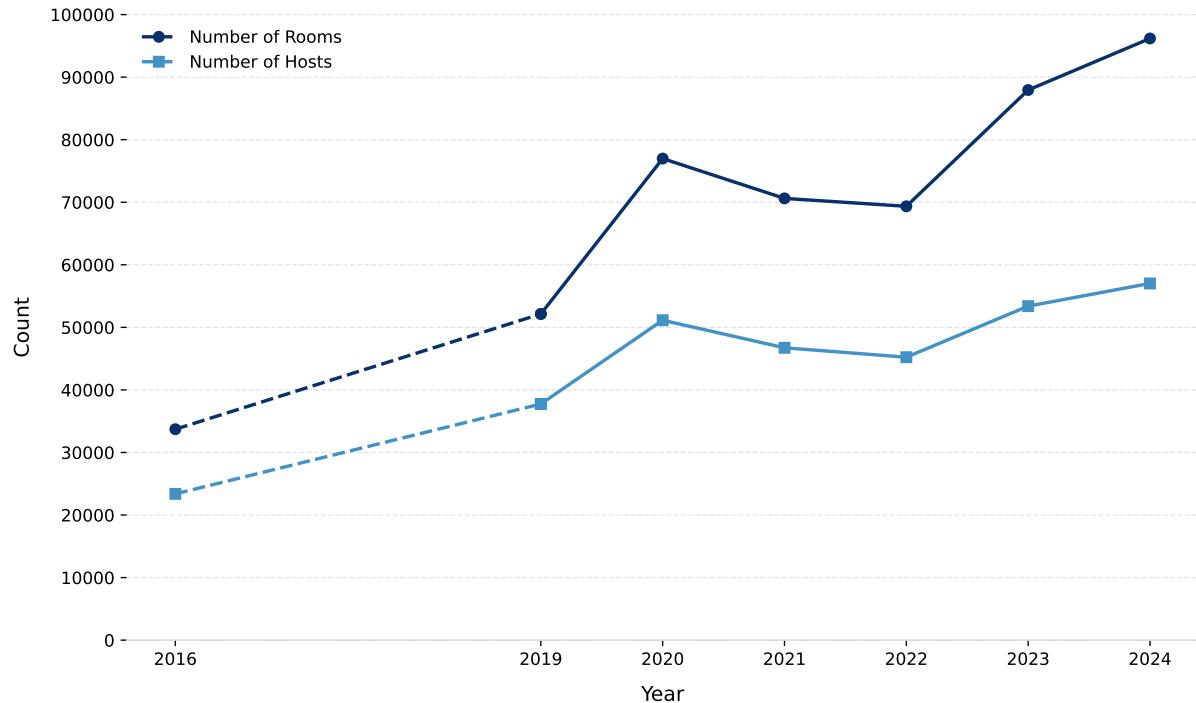
Legal and privacy concerns also arise. The legality of scraping varies by jurisdiction, with some viewing it as a breach of terms of service or copyright laws. Although anonymisation efforts shift geographic coordinates by up to 150 metres, privacy risks persist. The absence of informed consent from individuals whose data is used further complicates ethical compliance (Schaefer, Labude and Nasir, 2018).

InsideAirbnb's data raises broader concerns about its role in addressing negative social impacts, including rising property prices, income inequality and social segregation. The dataset's selective scope limits representativeness, while municipal governments face data deficits due to Airbnb's privatisation of information. These issues hinder regulation of short-term rentals and efforts to mitigate community disruption and other social challenges.

Despite limitations, InsideAirbnb's open-access model reduces research inequality but perpetuates disparities due to limited coverage and the inaccessibility of commercial datasets (Alsudais, 2021). Addressing these concerns requires rigorous validation processes, transparency in data collection, and adherence to ethical standards to ensure fairness, privacy, and accuracy in research and policymaking.

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

Figure 1: Airbnb Trends, Number of Rooms and Hosts (2016, 2019-2024)



Using data from InsideAirbnb via Orca, we analysed Airbnb trends in London from 2016 to 2024. There was growth in both properties and hosts, albeit irregular.

In Figure 1, the property-to-host ratio rose sharply between 2016 and 2019, increasing from approximately 1:1 to 1:2 (to one significant figure). This shift suggests that hosting on Airbnb transitioned from a secondary income stream to a full-time career for many, with more hosts owning multiple properties. However, this trend was interrupted in 2020 by the Health Protection Regulations, which required the closure of non-essential businesses, including short-term lets.

Figure 1 shows the property-to-host ratio fell slightly but remained relatively stable. Host numbers fell more sharply between 2020 and 2021, likely due to the Second National Lockdown in November 2020. Despite the pandemic, professional hosts maintained a strong market presence, leveraging their resources and expertise to navigate the crisis more effectively than smaller, non-professional hosts (Kourtit *et al.*, 2022).

By 2024, property listings had grown by over 10% compared to 2019, while host numbers remained nearly unchanged. This divergence highlights Airbnb's growing profitability in London, favouring larger operators.

Figure 2: Airbnb Listings and Market Share Trends by Host Type and Room Type (2019, 2021, 2023)

	2019	2021		2023		
Host Type		Market Share		Market Share	Market Share	
Individual Hosts	30230	57.98%	39112	55.39%	43381	49.33%
Professional Hosts	7522	42.02%	7623	44.61%	10013	50.67%
Room Type		Share of Room		Share of Room		Share of Room
Entire home/apt	29620	56.81%	39115	55.39%	54575	62.06%
Private room	21430	41.1%	30527	43.23%	32711	37.19%
Hotel room	659	1.26%	412	0.58%	219	0.25%
Shared room	434	0.83%	563	0.8%	441	0.5%

Between 2019 and 2023, Airbnb saw a clear shift toward professionalization in its host base and evolving consumer preferences in property types.

Even though the count of Individual Host rose by 30% from 2019 to 2023 where professional hosts only rose by 25% the market share of Individual Host during this period actually decreased by more than 8% whereas the professional hosts share rose by the same amount. This shift indicates that although individual hosts remain far more numerous, professional hosts have leveraged operational efficiency and standardized management practices to dominate the market share, reflecting a growing inequality between host types that exacerbates socio-economic disparities in the short-term letting market.

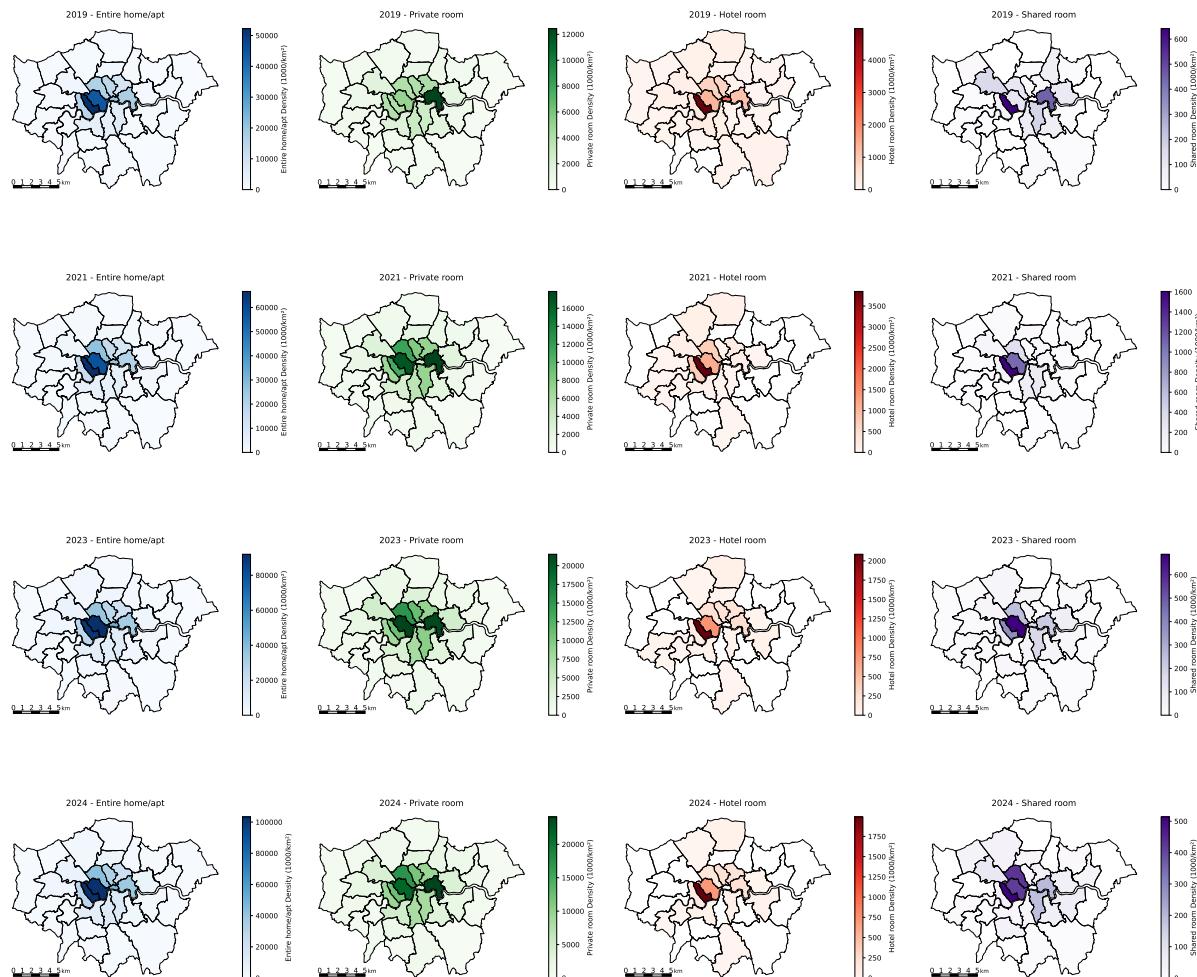
Entire home/apartment listings experienced the most significant recovery after the pandemic, with market share rose by more than 6%, highlighting the increasing preference for private, independent accommodations. Meanwhile, private rooms, despite an increase in total listings, lost market share, and hotel rooms saw a sharp decline, likely due to the pandemic.

Airbnb's trajectory reveals a strategic shift from its sharing economy roots to a more commercialized and professionalized model. While this meets consumer demands, it raises concerns about market concentration and socio-economic disparities, necessitating thoughtful regulation to mitigate housing market disruptions and community challenges.

7. Drawing on your previous answers, and supporting your response with evidence (e.g. figures, maps, EDA/ESDA, and simple statistical analysis/models drawing on experience from, e.g., CASA0007), how could the InsideAirbnb data set be used to inform the regulation of Short-Term Lets (STL) in London?

In the wake of the Second National Lockdown and strict travel restrictions, many smaller-scale hosts exited the market, a trend that persisted even as property listings grew. Figure 1 shows this shift in the host market, with increasing listings and declining hosts pointing to the consolidation of the Short-Term Lets (STL) market, as professional operators took a more dominant role, pushing out smaller hosts. Figure 2 shows entire home and apartment listings experienced the strongest recovery, with a nearly 30% increase in listings between 2021 and 2023, surpassing pre-pandemic levels by 46%. Early reopening in April 2021 allowed these properties to rebound quicker than shared room listings, which were restricted until February 2022. Entire home listings have solidified their role as the STL market's cornerstone, with professional hosts managing an average of five properties each. This professionalisation highlights growing inequality in the STL market, necessitating regulatory measures to address the socio-economic consequences of such concentration.

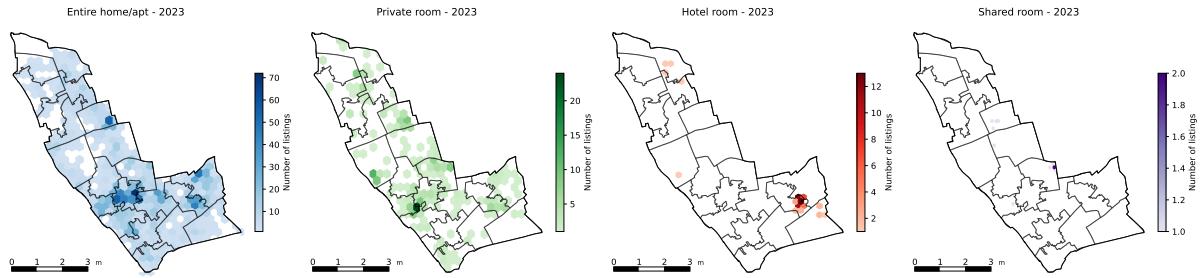
Figure 3: Professional Host Density by Room Type in London (2019, 2021, 2023)



Cox and Murray identify Airbnb as a significant driver of gentrification that worsens economic divides in neighbourhoods. Our report, however, situates Airbnb within a market marked by structural inequalities, amplified by the pandemic. Further analysis of bi-annual trends between

2019, 2021, and 2023 reveals significant shifts in host activity, market recovery, and professionalisation in relation to post-pandemic recovery. Spatial analysis of professional host STL activity across London reveals stark variations in how different boroughs are affected. Figure 3 demonstrates that Kensington and Chelsea is the most heavily impacted borough, with its clustering of professional host activity across three of the four property types analysed. This concentration underlines the need for a borough-specific regulatory approach. Other boroughs, such as Tower Hamlets and Westminster, are also impacted by private room listings, highlighting borough-specific dynamics requiring tailored solutions.

Figure 4: Professional Host Density by Room Type in Kensington and Chelsea (2023)



As shown in Figure 4, in Kensington and Chelsea, the dominance of entire home listings is particularly concerning as the impact spreads borough-wide. Entire homes account for 79.6% of the borough's Airbnb market, while shared rooms and hotel rooms contribute very little. This suggests that many professional hosts in the borough are absentee landlords, detached from the communities where their properties are located. Additionally, the borough has a significantly higher availability rate than the London average, with 49.1% of listings rented for more than 90 nights annually (Rozena and Lees, 2021). This exacerbates the negative impact on community cohesion and neighbourhood stability, as extensive use of entire home listings reduces the stock of permanent housing for local residents. The ripple effects of the Grenfell fire in 2017 further highlighted the deep socio-economic class divisions within Kensington and Chelsea, underscoring the urgency for addressing the wider housing and community stability issues exacerbated by the professionalisation of the STL market (Sanchez, Sanchez and Carlos, 2019).

The proliferation of professionalised STL activity has worsened the housing crisis in Kensington and Chelsea, contributing to an increase in unoccupied homes, which grew from 941 in 2015 to 1,857 in 2017 (Lusher, 2018). This trend, driven by foreign investment and offshore ownership, exacerbates local housing pressures, with professional operators dominating the STL market. To address these challenges, Kensington and Chelsea should be prioritised for piloting regulatory measures to limit the impact of professional STL operators. A two-year targeted intervention strategy is proposed for the borough, including two key measures: a community tax on entire home listings operated by professional hosts, with the revenue funding initiatives to address social deprivation, and stricter letting period limits for professional hosts. Additionally, a community takeover programme would allow local organisations like 240 Project (a not-for-profit Arts and Health Activity Centre for those affected by homelessness and exclusion) to use properties for community-focused projects for two-week periods every three months. This initiative would revitalise areas affected by short-term letting, foster local talent, and restore balance to the housing market. These measures aim to balance Kensington and Chelsea's appeal as an investment hub for STL's with the need to protect local communities and housing, fostering a more equitable and sustainable urban environment.

While its post-lockdown growth is vital for London's economic recovery, the InsideAirbnb dataset is insufficiently detailed, lacking host demographics and nuanced location data, risking the marginalisation of disadvantaged groups. The dataset's limitations—stemming from selective

and automated web scraping—offer an incomplete snapshot of Airbnb's impact, particularly in areas like Kensington and Chelsea. This undermines informed policymaking and the development of equitable regulations.

To address these gaps, we propose InsideAirbnb and Kensington and Chelsea council collaborate with Airbnb, leveraging commercial datasets to balance investment with measures that mitigate gentrification. Airbnb's 2022 evidence submission to the UK Government (Airbnb, 2022), underscores their commitment to partnerships and inclusivity, stating, “We look forward to working closely with the Government, policymakers... and communities.” They advocate for cost-free registration for non-commercial hosts and balanced approaches to “protect livelihoods... whilst balancing the needs of local communities.” Such openness to collaboration presents opportunities for equitable urban tourism recovery, improved data access, and shared responsibility in addressing neighbourhood concerns sustainably.

This report underscores the transformative professionalisation of London's short-term letting market, shaped by post-pandemic shifts, with significant socio-economic implications for boroughs and host types, highlighting the need for tailored, data-driven policies to address pressures and foster urban equity.

References

- Airbnb (2022) ‘Written evidence submitted by airbnb - committees’, *UK Parliament*. Available at: <https://committees.parliament.uk/writtenevidence/42199/pdf>.
- Alsudais, A. (2021) ‘Incorrect data in the widely used inside airbnb dataset’, *Decision Support Systems*, 141. Available at: <https://doi.org/10.1016/j.dss.2020.113453>.
- Cox, M. and Slee, T. (2016) ‘How airbnb’s data hid the facts in new york city’, p. 14. Available at: <https://insideairbnb.com/reports/how-airbnbs-data-hid-the-facts-in-new-york-city.pdf>.
- InsideAirbnb (2016) ‘Inside airbnb: home’. Inside Airbnb. Available at: <https://insideairbnb.com/get-the-data/>.
- Katz, M. (2017) ‘A lone data whiz is fighting airbnb — and winning’. WIRED. Available at: <https://www.wired.com/2017/02/a-lone-data-whiz-is-fighting-airbnb-and-winning/>.
- Kourtit, K. et al. (2022) ‘Airbnb and COVID-19: SPACE-TIME vulnerability effects in six world-cities’, *Tourism Management*, 93, pp. 104569–104569. Available at: <https://doi.org/10.1016/j.tourman.2022.104569>.
- Lusher, A. (2018) ‘Almost 100 grenfell families without home of their own this christmas, 18 months later’, *The Independent*. Independent Digital News; Media. Available at: <https://www.independent.co.uk/news/uk/home-news/grenfell-tower-fire-families-housing-christmas-kensington-chelsea-council-theresa-may-a8696941.html>.
- Rozena, S. and Lees, L. (2021) ‘The everyday lived experiences of airbnbification in london. Social & cultural geography’, 24(2), pp. 253–273. Available at: <https://doi.org/10.1080/14649365.2021.1939124>.
- Sanchez, B., Sanchez, B. and Carlos, J. (2019) ‘Towering grenfell: Reflections around socioeconomic disadvantage in antidiscrimination law’, *Social Science Research Network*.
- Schaefer, G.O., Labude, M.K. and Nasir, H.U. (2018) ‘Big data: Ethical considerations’, in *The palgrave handbook of philosophy and public policy*. Springer International Publishing, pp. ”593–607”. Available at: https://doi.org/10.1007/978-3-319-93907-0_45.
- Slee, T. (2017) ‘Airbnb data collection: Methodology and accuracy’. Available at: <https://tomslee.net/2017/10/airbnb-data-collection-announcement.html>.