NA_Group1′s Group Project

Declaration of Authorship

We, [NA_GROUP1], 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:24048055,24143995,24038681,24130339,24050260

Brief Group Reflection

What Went Well	What Was Challenging
teammates	output to pdf
data clean	analysis data to prove our opinions
useful references	explain our discovery in limit word

Priorities for Feedback

During our group project, we found it challenging to effectively connect the data we analyzed with the points we raised while reading the papers. Although we learned how to perform data analysis and coding exercises in class, and how to read academic papers and form our own insights, combining these two aspects has proven to be difficult. Another issue we faced is that the existing data has many gaps compared to the data we need for our research, such as timeliness issues. We are unsure how to address or substitute for these gaps at the moment. Additionally, answering questions concisely requires practice, which was an area we lacked during the semester.

On the positive side, we learned how to communicate our ideas to the public using simpler, non-academic language.

Response to Questions

1. Who collected the InsideAirbnb data?

The data is sourced from publicly available information on the Airbnb website and is analyzed, cleaned, and aggregated for public discussion (Inside Airbnb, n.d.a). Key contributors include the founder and collaborators who developed tools to enhance data transparency, such as automating search functionality and stabilizing the platform's code (Alsudais, 2021; Inside Airbnb, n.d.a).

2. Why did they collect the InsideAirbnb data?

Inside Airbnb collects data to enhance transparency by addressing incomplete and biased reports(Alsudais, 2021; Inside Airbnb, n.d.c). Studies show that short-term rentals disrupt communities, drive gentrification, and exacerbate housing inequities in cities like New York, London, and Nanjing (Wachsmuth and Weisler, 2018; Jiao and Bai, 2020; Sun, Zhang and Wang, 2021). To promote housing equity, Inside Airbnb focuses on: Increasing Transparency: Highlight the effects of short-term rentals on housing availability and affordability (Garcia-Ayllon, 2018; Airbtics, n.d.). Supporting Policy Development: Provide actionable data to regulate short-term rentals and tackle urban challenges (Jiao and Bai, 2020; UK Government, 2023).

3. How was the InsideAirbnb data collected?

Inside Airbnb relies on publicly accessible data to analyze the platform's impact on housing and communities. Using web scraping, it extracts and aggregates information such as listings, prices, calendars, reviews, and host details from Airbnb's website, which is then cleaned and prepared for public discussions and policymaking(Inside Airbnb, n.d.a). Meanwhile, Airbnb processes proprietary user interaction data through its User Signals Platform (USP), employing real-time analytics to support applications like personalization and market segmentation(Jiao and Bai, 2020).

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?

Impact on Data Completeness and Accuracy

Inside Airbnb data is gathered through web scraping. It may exclude some listings due to technical or legal barriers, such as anti-scraping technologies deployed by Airbnb(Airbnb, 2024c). In addition, data collection is done at intervals, which means dynamic changes such as new or deleted lists can be missed(Gurran and Phibbs, 2017).And the data collection method may underrate the number of listings. This factor contribute to data gaps, potentially overlooking numerous active listings and limiting the accuracy of analyses(Adamiak *et al.*, 2019).

Limitations in Timeliness and Geographic Representation in Reflecting Airbnb Data

InsideAirbnb's data collection method relies on periodic snapshots, with updates occurring every few months. This frequency means it may miss real-time changes, such as new or removed listings, limiting its ability to capture the dynamic nature of Airbnb's platform(Gurran and Phibbs, 2017). Additionally, although InsideAirbnb gathers data from cities in dozens of countries, it does not cover all Airbnb regions, which restricts its ability to fully represent the broader market. This affects the accuracy of its representation of Airbnb's operations across different geographical areas(Inside Airbnb, n.d.c).

Wider issues

On one hand, the possibility that research using this dataset could unintentionally reinforce biases in the representation of the Airbnb market, leading to skewed conclusions about the platform's impact(Adamiak *et al.*, 2019). Additionally, such research might focus on easily accessible data, like listing distribution and pricing, while overlooking more complex phenomena, such as user behavior or platform strategies(Bivens, Sawatzky and Wachsmuth, 2021).On the other hand, scraping data without explicit consent from hosts or Airbnb itself could raise ethical concerns, especially when dealing with sensitive information like earnings or availability(Floridi and Taddeo, 2016).

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

Firstly, the Inside Airbnb is supposed to protect the privacy of the hosts. While Inside Airbnb asserts that it avoids using personal information and processes data carefully (Inside Airbnb, n.d.b), the raw data scraped from Airbnb's website often includes host names, housing locations, and other sensitive information. Even when locations are obfuscated, the inclusion of identifiable data challenges the hosts' right to privacy.

Compared with privacy rights, the right to know how the hosts' information is being used is well protected by Airbnb and Inside Airbnb. As the privacy policies of Airbnb (Airbnb, 2024a)maintained, the types of personal information they collected are clearly shown on the website. The process and targets of using these data are also informed and legally guaranteed. Once these policies are changed greatly, they will connect the hosts. Hosts also enter into contracts with Airbnb, consenting to the use of their information. However, a key concern is whether hosts fully comprehend these contractual terms(Airbnb, 2024b).

Finally, the legality of the use of Inside Airbnb data is doubtful. Inside Airbnb made use of the skill of web scraping to get the data from Airbnb instead of getting an API from the platform, which is explicitly forbidden by the terms of service from Airbnb(Airbnb, 2024b). Moreover, this data acquisition process broke the laws of many regions around the world such as General Data Protection Regulation (GDPA) of Europe and the Privacy Act of Australia (Australia Government, 1988; Intersoft Consulting, 2018). Although Airbnb has got permissions from the hosts to deal with the sensitive data, Inside Airbnb did not carry out this procedure.

With regard to the indirect ethical influence of using data from Inside Airbnb, the problems of discrimination and inequality can be caused.For instance, according to Wachsmuth and Weisler (2018), certain communities may be over-labeled after

the analysis through Inside Airbnb data, especially those exist gentrification phenomenon. At the meantime, as Horn and Merante (2017) mentioned, Inside Airbnb has a high coverage of popular cities or areas. However, there are insufficient listings for those remote regions and markets that are lack of popularity.

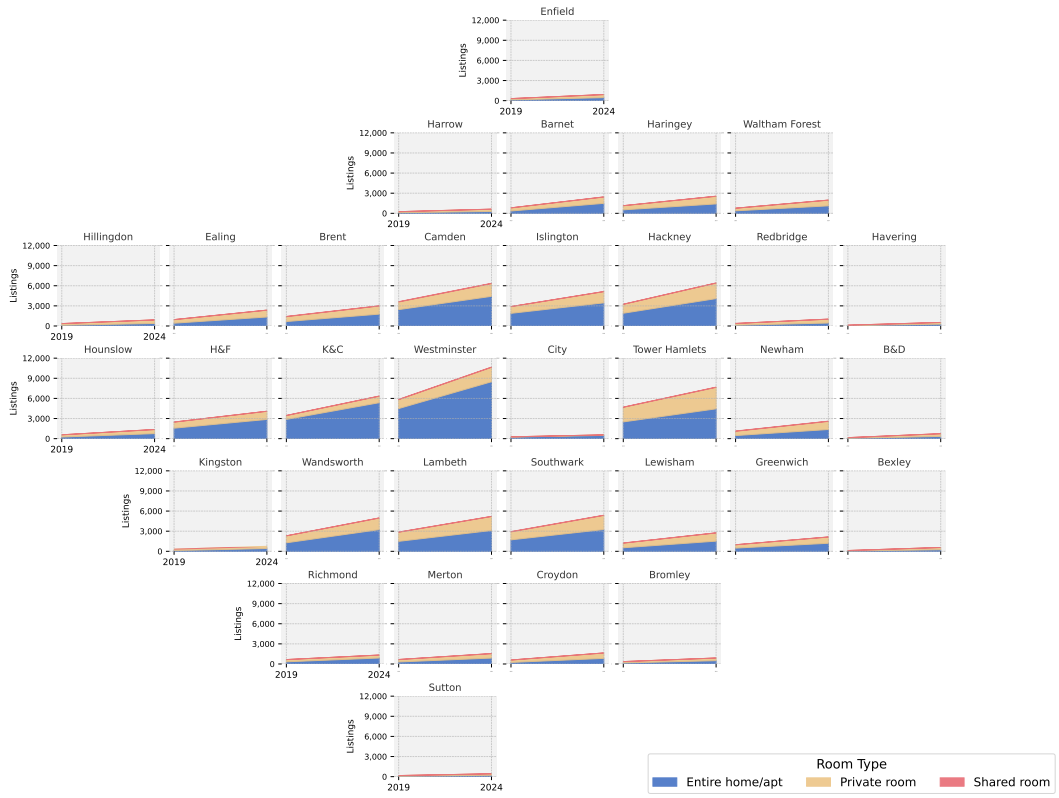
6. With reference to the InsideAirbnb data (*i.e.* using numbers, figures, maps, and descriptive statistics), what does an analysis of Hosts and Listing types suggest about the nature of Airbnb lets in London?

Table 1: Number of hosts with one or more active Airbnb listings, distribution of host percentages, and listing types managed by hosts

No. listings linked to host ID	Number of Hosts	% of Hosts	% Entire home/apt	% Private room	% Shared room
1	29616	48.9	62.34	37.14	0.5
2	13236	21.85	56.51	42.91	0.57
3	6177	10.2	55.76	43.53	0.65
4 to 10	8820	14.56	58.13	41.0	0.62
11 to 50	2298	3.79	68.89	29.98	0.91
51 to 100	264	0.44	74.62	22.35	0.76
101 to 200	88	0.15	73.86	25.0	1.14
200 or more	66	0.11	69.7	24.24	3.03
TOTAL	**60565**	**100.0**	/	/	/

The data reveals that 48.9% of hosts are non-professional, typically individual homeowners renting out spare rooms or properties part-time. In contrast, professional hosts, who manage multiple listings, control a significant share, highlighting Airbnb's important for commercial property management(Li, Moreno and Zhang, 2016; Kwok and Xie, 2019). Entire home/apartment listings dominate the market, especially among professional hosts, where their share continues to grow. Meanwhile, shared rooms account for a minimal portion, reflecting the market's clear preference for private and independent spaces.

Figure 1: Number of Airbnb Listings in London Boroughs (2019 vs 2024)



In figure1, professional hosts' listings are concentrated in central London and tourist areas(Westminster, Kensington). In contrast, non-professional hosts(aim to supplement income), tend to have their listings on the outskirts of London. The Airbnb market in London has the polarization between the sharing-economy and commercialization.

7.Drawing on your previous answers, and supporting your response with evidence, how could the InsideAirbnb data set be used to inform the regulation of Short-Term Lets (STL) in London?

According to our research in question 6, entire home/apt and non-professional hosts (only have one Airbnb listing) are two significant factors of London Airbnb. Our following studies will make use of these two points.

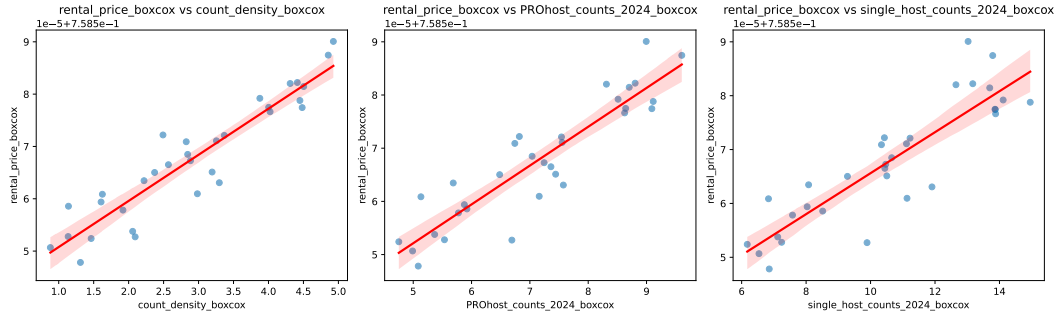
Table 2: Comparison of Average Nightly Revenue from Airbnb vs Open Market Rent in London

Price Benchmark	Rental Period (nights/year)	Revenue per Night (£)	Revenue per Annum (£)
VOA (mean)	365	66	24445
STL (mean) 90 nights	90	238	21489
STL (mean) 102 nights	102	238	24354

As shown in Table 2, we use Valuation Office Agency (VOA) data to estimate the average daily income from long-term rentals in London at £67 per day, resulting in an annual income of approximately £24,446. In comparison, data from the Inside Airbnb dataset estimates the average daily income from short-term lets (STL) at £239. If limited to the legally permitted 90 nights per year, STL generates £21,489 annually, slightly below the income from long-term rentals. However, exceeding the 90-night limit (around 102 nights) raises STL earnings to £24,354, surpassing long-term rental income. Despite additional STL costs, such as cleaning and vacancy risks, this higher return suggests the potential for speculative behavior.

Secondly, we use "Airbnb Count Density", "Single Host Counts 2024", and "Professional Host Counts 2024" as independent variables, with "Rental Price" as the dependent variable. While these variables are not initially normally distributed, they achieve normality after applying Box-Cox transformations.

Figure.2 Normality Analysis



The scatterplot and correlation analysis reveal a positive correlation between all three variables and "rental price", with "count density Box-Cox" showing the strongest correlation (0.932). "PROhosts" also exhibit a high correlation (0.912), likely because professional hosts tend to list higher-priced properties. In contrast, non-professional hosts show a weaker and more scattered correlation (0.885), suggesting they have less influence on market prices.However, there may be a very large correlation effect between them.

To prepare for the Ordinary Least Squares (OLS) Regression, we carry out the Variance Inflation Factor (VIF) Test beforehand. The data for all the three independent variables are greater than 10, indicating that they have a high degree of multicollinearity. Therefore, we gradually removed the variables with the highest VIF values and recalculated the VIF. After removing the "PROhosts box-cox", VIF data for both of the other two variables are below 10.

Regression Equation:

*Y=0.7585+[8.355*10^(-6)]X1+[2.22*10^(-7)]X2*

Y: rental price box-cox

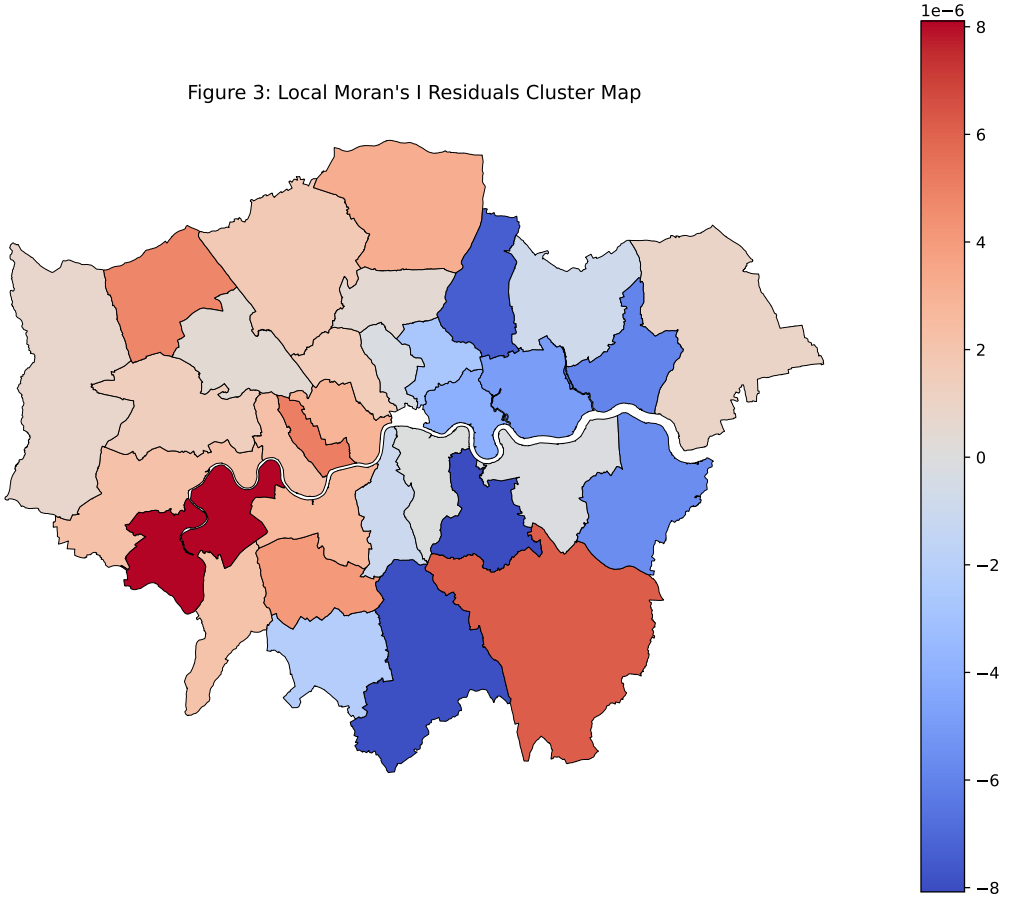
X1: count density box-cox

X2: single host counts 2024 box-cox

According to the equation, "count density box-cox" has a significant positive effect on "rental price box-cox". This suggests that demand for housing in high density areas is usually greater and rental prices are therefore higher, especially in large cities (e.g. London) where there is a strong positive correlation between housing density and rents. However, the effect of "single host counts 2024 box-cox" on the

dependent variable is not significant (p-value = 0.798). This indicates that the market behavior of non-professional landlords has a weak impact on the overall rent level. The rental market may be dominated by professional landlords, who usually offer higher prices and are more concentrated in the core area, while single landlords are mostly individual homeowners who lack the scale effect to influence market prices. The overall model explains "87.1%" of the variance in the dependent variable (R^2 = 0.871), which is a good fit, with residuals basically conforming to a normal distribution.

Figure 3: Local Moran's I Residuals Cluster Map



This map shows the spatial distribution of the residuals. The Red areas indicate positive residuals, where actual values are higher than model predictions, concentrated in the South West and South East regions of London. The Blue areas indicate negative residuals, where actual values are lower than model predictions, concentrated in the East and North East regions. Although there is some clustering of red and blue regions, there is no clear pattern of spatial clustering in the overall distribution. This is consistent with the results of Moran's I and p values. The small value of Moran's I and p-value > 0.05 indicate that the residuals are not significantly spatially clustered and the distribution is essentially random. Therefore, the spatial autocorrelation is not significant.

Recommendations

1. Strictly Enforce the 90-Day Limit: Regulate STL to prevent speculative activities that disrupt the housing market.

2. Neighborhood-specific Controls: Adjust Airbnb listings to align with local housing needs, promoting balance and sustainability.
3. Address Broader Influences: Consider landlord behavior, economic conditions, and structural factors in housing policy.

Limitations

1. Data Gaps: Missing values and incomplete neighborhood coverage affect analysis accuracy.
2. Spatial Variations: While residuals lack overall clustering, local spatial differences still require deeper investigation.
3. Box-Cox Transformation: Ensuring normality may alter some data's original characteristics.
4. Temporal Discrepancies: Airbnb pricing reflects specific points in time, whereas local rents are annual averages.
5. Single-Year Data: The model captures only one year of data, limiting trend analysis.

Conclusion

Our analysis indicates the density of Airbnb listings as a key driver of local rental prices. Regulation of short-term rentals limited to 90 days and neighborhood-specific quantity controls are critical to achieving sustainability in the rental market.

References

Adamiak, C. *et al.* (2019) 'Airbnb offer in spain: Spatial analysis of the pattern and determinants of its distribution', *ISPRS International Journal of Geo-Information*, 8(3). Available at: https://doi.org/10.3390/ijgi8030155.

Airbnb (2024a) 'Privacy policy'. Available at: https://www.airbnb.com/help/article/3175.

Airbnb (2024b) 'Terms of service'. Available at: https://zh.airbnb.com/terms/old.

Airbnb (2024c) 'Understanding automated hosting on airbnb'. Available at: https://www.airbnb.co.uk/help/article/3418?_set_bev_on_new_domain=1734253678_EAMzJhY2QwOTBhNz.

Airbtics (n.d.) 'Inside airbnb data guide'. Available at: https://airbtics.com/inside-airbnb-data/.

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

Australia Government (1988) 'Privacy act 1988'. Available at: https://www.legislation.gov.au/C2004A03712/latest/text.

Bivens, J., Sawatzky, M. and Wachsmuth, D. (2021) 'The impact of new regulations on airbnb in canadian cities: Case analysis'. PDF. Available at: https://upgo.lab.mcgill.ca/publication/impact-of-new-regulations/impact-of-new-regulations.pdf.

Floridi, L. and Taddeo, M. (2016) 'What is data ethics?', *Philosophical Transactions of the Royal Society A*, 374(2083), p. 20160360. Available at: https://doi.org/10.1098/rsta.2016.0360.

Garcia-Ayllon, S. (2018) 'Urban transformations as an indicator of unsustainability in the P2P mass tourism phenomenon: The airbnb case in spain through three case studies', *Sustainability*, 10(8). Available at: https://doi.org/10.3390/su10082933.

Gurran, N. and Phibbs, P. (2017) 'When tourists move in: How should urban planners respond to airbnb?', *Journal of the American Planning Association*, 83(1), pp. 80–92. Available at: https://doi.org/10.1080/01944363.2016.1249011.

Horn, K. and Merante, M. (2017) 'Is home sharing driving up rents? Evidence from airbnb in boston', *Journal of Housing Economics*, 38, pp. 14–24. Available at: https://doi.org/10.1016/j.jhe.2017.08.002.

Inside Airbnb (n.d.a) 'About inside airbnb'. Available at: https://insideairbnb.com/about/.

Inside Airbnb (n.d.b) 'Data assumptions'. Available at: https://insideairbnb.com/data-assumptions/.

Inside Airbnb (n.d.c) 'Platform failures: Airbnb's illegal listings'. Available at: https://insideairbnb.com/report/platform-failures/.

Intersoft Consulting (2018) 'General data protection regulation (GDPR) – legal text'. Available at: https://gdpr-info.eu/.

Jiao, J. and Bai, S. (2020) 'Cities reshaped by airbnb: A case study in new york city, chicago, and los angeles', *Environment and Planning A: Economy and Space*, 52(1), pp. 10–13. Available at: https://doi.org/10.1177/0308518X19853275.

Kwok, L. and Xie, K.L. (2019) 'Pricing strategies on airbnb: Are multi-unit hosts revenue pros?', *International Journal of Hospitality Management*, 82, pp. 252–259. Available at: https://doi.org/10.1016/j.ijhm.2018.09.013.

Li, J., Moreno, A. and Zhang, D. (2016) 'Pros vs joes: Agent pricing behavior in the sharing economy'. Rochester, NY: Social Science Research Network. Available at: https://doi.org/10.2139/ssrn.2708279.

Sun, S., Zhang, S. and Wang, X. (2021) 'Characteristics and influencing factors of airbnb spatial distribution in china's rapid urbanization process: A case study of nanjing', *PLOS ONE*, 16(3), p. e0248647. Available at: https://doi.org/10.1371/journal.pone.0248647.

UK Government (2023) 'Consultation on a registration scheme for short-term lets in england'. Available at: https://www.gov.uk/government/consultations/consultation-on-a-registration-scheme-for-short-term-lets-in-england.

Wachsmuth, D. and Weisler, A. (2018) 'Airbnb and the rent gap: Gentrification through the sharing economy', *Environment and Planning A: Economy and Space*, 50(6), pp. 1147–1170. Available at: https://doi.org/10.1177/0308518X18778038.