Group Name's Group Project

Declaration of Authorship

We, `505 not found`, 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:

Student Numbers:

Brief Group Reflection

What Went Well	What Was Challenging
A	B
C	D

Priorities for Feedback

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

Response to Questions

See the raw file for examples of how to hide computational output as there is code hidden here.

1. Who collected the InsideAirbnb data?

Inside Airbnb data was collected by Murray Cox, a data activist and the project's founder, John Morris, the website designer and report producer, and Taylor Higgins, a master's student focusing on sustainable tourism at the Università degli Studi di Firenze.

2. Why did they collect the InsideAirbnb data?

The purpose of InsideAirbnb data is to provide data-driven insights into the impact of Airbnb on residential housing markets, thereby contributing to public discourse on the regulation and effects of short-term rental platforms in urban areas.

3. How did they collect it?

The data is collected by utilizing web scraping techniques such as self-made bots, inside Airbnb and AirDNA (Pawlicz and Prentice, 2021) (Prentice and Pawlicz, 2023) to extract publicly available information from Airbnb's website, focusing on various aspects of listings such as location, price, availability, and host details. This approach allows for the assembly of comprehensive datasets, which are then cleansed and organized to facilitate thorough analysis.

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?

The data collection method used by InsideAirbnb raises data quality issues, mainly related to data incompleteness, reliance on website structure, and technical challenges. In terms of accuracy, data is automatically retrieved from the website, which possess the risk of capturing inaccurate or outdated information due to the dynamic nature of web content (Krotov and Johnson, 2023).

In addition, due to privacy measures, the geographic coordinates provided by Airbnb may not reflect the exact location of the listing, which adds a layer of inaccuracy. And as web scraping depends heavily on the structure of the Airbnb website. Changes to the website layout or measures to block scraping activities may disrupt data collection efforts, like Airbnb's anti-scrap measures including CAPTCHA or IP bans, pose additional challenges (Prentice and Pawlicz, 2023). This burdens data analysts by requiring them to constantly develop and maintain scraping scripts.

In the discussion of the structure of InsideAirbnb data, it contains all aspects of the Airbnb market, including the distribution and characteristics of listings, pricing models, and the impact of Airbnb on the local housing market. And it is a relatively complete dataset and can assist with comprehensive analysis study.

Besides the accuracy concerns, the use of InsideAirbnb data raises technical, legal, and ethical issues. Legally, as discussed in Sobel (Sobel, 2021), scraping faces challenges in different jurisdictions, depending on how it intersects with privacy laws and terms of service agreements. This could affect the legality of the Inside Airbnb data collection process, especially if it violates Airbnb's terms of service. Scraping also raises ethical issues, particularly regarding the consent of data subjects (Airbnb hosts and guests) whose information is collected without explicit permission. This raises significant privacy issues, as highlighted in the study by Xie and Karan (Xie and Karan, 2019), where users' awareness and concerns about how their data is used influence their privacy management behaviours.

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

The use of the InsideAirbnb database does raise several ethical considerations.

Firstly, there are issues of legal compliance. Web scraping can conflict with legal standards and ethical norms, particularly when data is collected without explicit consent, potentially leading to legal actions (Krotov and Johnson, 2023).

Secondly, privacy concerns for individuals must be addressed. Although the data might be publicly accessible, individuals typically do not anticipate their rental information being extensively aggregated and analyzed (Brenning and Henn, 2023).

In many instances, data subjects (hosts and guests) are neither directly informed nor asked for consent when their data is scraped and analyzed. This presents a significant ethical dilemma: using their information without explicit permission, especially when such data might be utilized to draw conclusions or influence policies that could directly impact them.

Moreover, there is the issue of how policymaking might be influenced by the data. Since the scraped data can contain errors, issues with accuracy and potential misrepresentation may lead to misleading conclusions that could negatively affect Airbnb hosts, guests, and policy decisions.

Additionally, the misuse of data poses a significant ethical concern. When analyzing Inside Airbnb data, it is crucial to ensure that the data is not used for purposes unintended by the original data providers, such as market manipulation, unfair competition, or research that adversely impacts hosts and guests.

Lastly, transparency and accountability are crucial. Ethical research involving data scraping should clearly disclose its methodologies, the specific data collected, and how this data is utilized. Such transparency is especially important for accountability, particularly if the research has the potential to influence public opinion or policy (Brenning and Henn, 2023).

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?

6.1 Analysis of Hosts

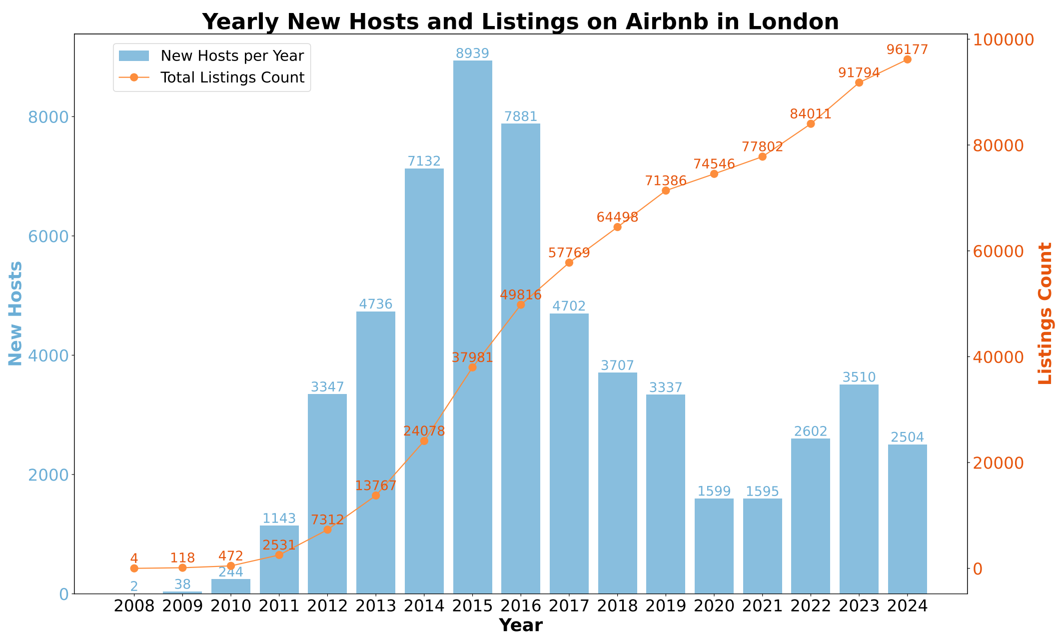
6.1.1 Distribution of the Number of Listings per Host

In London, only 47.8% (45,932 listings) are owned by single-listing hosts, while the remaining 52.2% are held by multi-listing hosts.
Notably, hosts with 10 or more listings account for 20.8% (20,038 listings) of the total.

conclusion:

1. Prevalence of multi-listing hosts: more than half of all listings owned by multi-listing hosts, indicating that multi-listing is common in london.
2. Professional landlords: hosts who owned 10+ listings owned more than one fifth listings, suggesting a significant presence of professional landlords in the market.

6.1.2 Changes in the number of landlords and renters over the years



- Based on Airbnb's dataset for London, the whole story started in 2008 and the growth in hosts and listings was slow between 2008 and 2010, **accelerating sharply from 2011 to 2016**. The peak occurred in 2015, with 8,939 new hosts, while 2014 and 2016 saw increases of over 7,000 hosts each. By 2016, total listings neared 50,000.

- However, **growth slowed in subsequent years**, with 2020 and 2021 adding only around 1,600 hosts and 3,000 listings annually—nearly half the growth seen in 2019—largely due to the pandemic's impact on the rental market. From 2022 to 2024, post-pandemic recovery is evident, but growth remains far below peak levels.

6.2 Analysis of property

6.2.1 Distribution of room types of property

Room type of property is divided into four categories.

- Entire home/apt: 63.8%
- Private room: 35.6%
- Shared room: 0.45%
- Hotel room: 0.2%

Conclusion:

1. The high proportion of entire homes/apt indicates that many guests prefer independent accommodations for greater privacy and autonomy.
   This aligns with a broader shift in tourism, where more visitors are opting for alternative lodging options instead of traditional hotels to enjoy a more spacious and private environment (Zervas, Proserpio and Byers, 2017).
2. The 35.6% share of private rooms suggests that some guests are still willing to choose more affordable accommodations, even if it means sharing common spaces. These listings cater to budget-conscious travelers.
3. The low percentages of shared rooms and hotel rooms indicate that Airbnb's core market in London, a well-established market, tends to favor more private lodging options.

6.2.2 Distribution of Minimum Nights for renting property

Based on the dataset, 93550 listings have a minimum night stay of less than the STR threshold (30 days), making up 97.3% of the total. Additionally, listings with a minimum stay of less than 7 days account for 92.3% of the total. **The London rental market on airbnb is dominated by short-term rentals.**

Chaudhary had illustrated some drawbacks of short term renting (Chaudhary, 2021).

1. **Reduced long-term housing supply**: Due to higher profits from short-term rentals (e.g., Airbnb), many landlords prioritize short-term leases over long-term rentals, exacerbating London's housing crisis and driving up rents, especially for low- and middle-income residents.
2. **Community impacts**: A high volume of short-term rentals can disrupt neighborhoods, increasing noise and tourist traffic, making communities less appealing for long-term residents and undermining stability and safety.

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?

7.3 The Impact of Airbnb on London Neighborhoods: A K-Means Geodemographic Classification

7.3.1 Data processing

Classification basis

- *key variables*
  - Airbnb Average Price per Night: Reflects the economics of the short-term rental market.
  - Airbnb Density: shows how well Airbnb's are distributed in a given neighborhood.
  - Changes in Social Listing Prices Over a Five-Year Period: Indirectly reflects Airbnb's potential impact on the local real estate market.
  - Population Density: Indicates the size and density of a community's population.
  - Income Deprivation Index: Higher values indicate greater income deprivation, which may be
  - correlate with the economic status and social vulnerability of the community
  - Hotel Density: measures the competitive environment for traditional lodging establishments.
  - Attraction Density: Reflects the community's tourist appeal and potential demand for Airbnb's.
  - Public Transportation Accessibility: Demonstrates the community's accessibility to tourists.

Variables that need to be processed :

- Calculate the density of tourist attractions in each LSOA in London
- Calculate hotel density for each LSOA unit
- Calculate five-year house price changes for each LSOA unit
- Read in and process Airbnb data
- Calculate Airbnb Density and Average Price per Night for LSOA Units
- Merge all data
- Check and clean the data

7.3.2 Standardization (Z-score scaling)

7.3.3 Elbow Method and Silhouette Coefficient

- Based on the information provided in the graphs, the optimal number of clusters (K) appears to be 4.

7.3.4 K-means Clustering

- **Results of K-means clustering**
- Cluster 0: Airbnb Core Areas — "Tourist Hotspots"
  - Key Characteristics:
    * High Airbnb Density: These areas have a very high concentration of short-term rentals, showing strong demand for Airbnb.
    * High Prices: Average nightly prices are significantly higher, reflecting intense tourism demand.
    * Dense Tourist Attractions: High hotel and tourist attraction densities, making them prime destinations for visitors.
    * High Population Density: Crowded residential areas with heavy foot traffic.
    * Convenient Public Transport: These areas have exceptional public transport accessibility, ideal for both tourists and residents
- Cluster 1: Emerging Airbnb Areas — "Expansion Zones"
  - Key Characteristics:
    * Rapid House Price Growth: These areas have seen significant house price increases in recent years, signaling Airbnb's growing presence.
    * Moderate Airbnb Density: Short-term rental activity is currently low but shows clear growth potential.
    * Lower Public Transport Access: Transport accessibility is limited, which may currently deter visitors.
    * Moderate Population Density: These areas are less crowded and have space for further development.
- Cluster 2: Low-Impact Areas — "Stable Residential Zones"
  - Key Characteristics:
    * Low Airbnb Density: Short-term rentals are scarce, and the market remains underdeveloped.
    * Affordable Prices: Airbnb listings are cheaper, indicating low tourism demand.
    * Weaker Local Economy: Lower income levels and economic activity compared to other clusters.
    * Low Tourist Activity: Few hotels and tourist attractions make these areas less appealing to visitors.
    * Limited Transport Access: Public transport is less accessible, reducing visitor convenience.
- Cluster 3: Transitional Areas — "Moderate Impact Zones"
  - Key Characteristics:
    * Moderate Airbnb Density: Short-term rentals are steadily increasing, showing a moderate level of activity.

- High Population Density: These neighborhoods are densely populated, similar to central areas.
- Economic Inequality: Income deprivation is evident, with significant disparities across neighborhoods.
- Moderate Tourist Presence: Tourist attractions and hotels are present but not as dominant as in Cluster 0.
- Good Public Transport: Transport accessibility is above average, supporting both residents and visitors.

7.3.5 Plot K-Means Clustering Result

- From the map:
- Cluster 0: Airbnb Core Areas — "Tourist Hotspots" : Primarily in the central London including the City of London, Westminster and Kensington & Chelsea , along the River Thames and near key landmarks.
- Cluster 1: Emerging Airbnb Areas — "Expansion Zones" : Located between central London and outer suburban areas, representing transitional zones. Such as Bromley, Croydon, and parts of Bexley and Havering.
- Cluster 2: Low-Impact Areas — "Stable Residential Zones" Primarily on the outskirts of London, far from the city center and main tourist zones. It includes areas like Lambeth, Southwark, and parts of Lewisham and Greenwich.
- Cluster 3: Transitional Areas — "Moderate Impact Zones" : Located between central and suburban London, acting as buffer zones.Such as Barnet, Enfield, and parts of Redbridge and Waltham Forest.

7.3.6 Summarize Airbnb Impact and Policy Recommendations:

- Airbnb Core Areas — "Tourist Hotspots":
  - Airbnb Impact:
    - Housing Pressure: Rising rents and property prices have made it harder for locals to afford housing.
    - Disruption to Local Life: Over-tourism may disturb daily life for residents.
  - Policy Recommendations:
    - Introduce tourism taxes or fees to balance economic benefits with community needs.
- Emerging Airbnb Areas — "Expansion Zones":
  - Airbnb Impact:
    - At-Risk Areas: These neighborhoods are at risk of increased housing pressure as Airbnb expands.
    - Potential Community Change: Growing short-term rental activity could alter the neighborhood's character.
  - Policy Recommendations:
    - Introduce preventative regulations, such as Airbnb listing caps, to manage future growth.
    - Prioritize local housing needs to avoid displacement of residents.
- Low-Impact Areas — "Stable Residential Zones":

  - Airbnb Impact:
    * Minimal Disruption: Short-term rentals have little effect on housing or local communities.
  - Policy Recommendations:
    * Invest in local economic development to improve quality of life for residents.
    * Monitor Airbnb trends to preempt potential issues in the future.
- Transitional Areas — "Moderate Impact Zones"
  - Airbnb Impact:
    * Emerging Challenges: Increasing Airbnb activity may intensify housing pressure and inequality.
    * Changing Community Dynamics: Short-term rentals could disrupt the balance of residential and tourist use.
  - Policy Recommendations:
    * Introduce balanced regulations to protect housing affordability while allowing controlled tourism growth.
    * Ensure economic benefits from Airbnb are reinvested into the community to address inequality.

Sustainable Authorship Tools

Using the Terminal in Docker, you compile the Quarto report using `quarto render <group_submission_file>.qmd`.

Your QMD file should automatically download your BibTeX and CLS files and any other required files. If this is done right after library loading then the entire report should output successfully.

Written in Markdown and generated from Quarto. Fonts used: Spectral (mainfont), Roboto (sansfont) and JetBrains Mono (`monofont`).

References

Brenning, A. and Henn, S. (2023) 'Web scraping: A promising tool for geographic data acquisition', *arXiv preprint arXiv:2305.19893*.

Chaudhary, A. (2021) 'Effects of airbnb on the housing market: Evidence from london.', *Available at SSRN 3945571*.

Krotov, V. and Johnson, L. (2023) 'Big data: Challenges related to data, technology, legality, and ethics', *Business Horizons*, 66(4), pp. 481–491.

Pawlicz, A. and Prentice, C. (2021) 'UNDERSTANDING SHORT-TERM RENTAL DATA SOURCES – a VARIETY OF SECOND-BEST SOLUTIONS', *Tourism in Southern and Eastern Europe*. Available at: https://api.semanticscholar.org/CorpusID:246571127.

Prentice, C. and Pawlicz, A. (2023) 'Addressing data quality in airbnb research', *International Journal of Contemporary Hospitality Management*. Available at: https://api.semanticscholar.org/CorpusID:258644931.

Sobel, B. L. (2021) 'The new common law of web scraping', *Lewis & Clark L. Rev.*, 25, p. 147.

Xie, W. and Karan, K. (2019) 'Consumers' privacy concerns and privacy protection on social networking sites in the era of big data: Empirical evidence from college students', *Journal of Interactive Advertising*, 19(3), pp. 187–201.

Zervas, G., Proserpio, D. and Byers, J. W. (2017) 'The rise of the sharing economy: Estimating the impact of airbnb on the hotel industry', *Journal of marketing research*, 54(5), pp. 687–705.