# XSWL Group’s Group Project：Using Inside Airbnb Data to Assess London’s Short-Term Lets (STL) Market: A Comparative Analysis of Traditional and Emerging Tourist Areas

## Declaration of Authorship

We, [XSWL Group], 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: 5

## Brief Group Reflection

| What Went Well | What Was Challenging |
| --- | --- |
| A | B |
| C | D |

## Priorities for Feedback

如果我们能够提供，您是否希望在哪些方面提供更详细的反馈？

# Response to Questions

We have combined the seven questions that need to be answered with our report:

For \*Questions 1-5\*: please see \*2.1.1 InsideAirbnb Data\*.

For \*Questions 6\*: please see \* \*.

For \*Questions 7\*: please see \* \*.

# 1. Airbnb in London: Challenges and Opportunities

## 1.1 Background: Airbnb and London’s Tourism Landscape

Airbnb is a well-known shared accommodation platform that has had a significant impact on the global tourism industry and short-term rental market (Chaudhary, 2021). In London, \*\*traditional tourist hotspots\*\* such as Westminster and Camden are challenged by increasing overcrowding and housing pressures. At the same time, the rise of Airbnb has fuelled economic growth in \*\*non-traditional tourism areas\*\*, such as Hackney and East London, which were previously overlooked as major tourist destinations (Airbnb, 2023).

This shift highlights a key opportunity:

- Can Airbnb's expansion into emerging areas help reduce congestion in traditional tourist hotspots while promoting economic benefits throughout the city?

- What are the wider implications for housing availability, economic balance and tourism policy?

## 1.2 Objectives: How Can Data Inform Better Tourism and Economic Policies?

To address these questions and provide data-driven insights for the Mayor of London's office, we used publicly available data from Inside Airbnb. The analysis focuses on:

1. The \*\*spatial distribution\*\* of Airbnb listings in traditional and non-traditional tourist areas.

2. The \*\*economic impact\*\* of Airbnb on regional retail and hospitality sectors.

The ultimate aim is to assess the effectiveness of the current short-term lets (STL) policy and to propose a strategy that supports \*\*balanced tourism development\*\* and ensures economic growth and sustainable housing supply in different parts of London.

# 2. Where the Datasets Come from and How We Analysed Them

## 2.1 Data Source

### 2.1.1 InsideAirbnb Data

The data used in this analysis is sourced from \*\*Inside Airbnb\*\*, covering the past twelve months of \*Airbnb listings\* in London.

\*\*1. Who collected the InsideAirbnb data?\*\*

InsideAirbnb data was collected by Murray Cox and his collaborators. As discussed on @insideairbnb, Inside Airbnb was founded by Murray Cox, an artist, activist and technologist who conceived the project, compiled and analysed the data and built the site.

\*\*2. Why did they collect the InsideAirbnb data?\*\*

The Inside Airbnb data was collected to:

**-** \*\*Understand the impact of STL\*\* on local communities and housing markets.

- \*\*Support policymakers and researchers\*\* with transparent data for informed decisions.

- \*\*Raise public awareness\*\* and promote responsible short-term rental policies and practices.

\*\*3. How did they collect it?\*\*

Inside Airbnb collects publicly available data from the Airbnb website, including listing details, availability calendars (365 days), and reviews. The data is then:

1. \*\*Cleaned: \*\* Verified, anonymized (locations within 150 meters), and corrected for accuracy.

2. \*\*Aggregated: \*\* Summarized to provide key metrics for visualization and analysis.

3. \*\*Analysed: \*\* Occupancy models estimate rental frequency and income, with filters applied to highlight highly available and frequently booked listings.

The site uses open-source tools like Python, PostgreSQL, and Mapbox to process and present the data for public research and policy discussions.

\*\*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?\*\*

1. Completeness & Accuracy:

\*\*Sampling Limitations: \*\* Web scraping captures only publicly available listings, excluding private, removed, or hidden rentals.

\*\*Dynamic Data: \*\* Listings change frequently, so snapshots may not reflect real-time trends.

\*\*Data Quality: \*\* Errors in scraping or Airbnb's platform can affect reliability.

2. Representation:

The data highlights trends but misses local market dynamics and host behaviors. Regional variations limit broader applicability.

3. Wider Issues:

\*\*Policy Risks: \*\*Incomplete data can lead to misguided policies.

\*\*Privacy: \*\* Ethical concerns arise around data ownership and usage.

\*\*Bias: \*\* Data motivations may influence analysis.

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

1. \*\*Privacy: \*\* Publicly scraped data includes host details, raising concerns about consent and privacy.

2. \*\*Data Misuse: \*\* Misinterpretation can lead to flawed conclusions or biased policies.

3. \*\*Community Impact: \*\* Highlighting certain areas may stigmatize hosts or neighbourhoods, exacerbating tensions.

4. \*\*Discrimination: \*\* Regulatory scrutiny may unfairly target specific communities.

5. \*\*Transparency: \*\* Data collection methods and biases must be clear to ensure accountability.

6. \*\*Long-Term Effects: \*\* Policy decisions based on incomplete data can harm housing affordability and vulnerable populations.

### 2.1.2 London Boundary Data

The data used for London boundaries in this analysis is sourced from \*\*the London Datastore\*\*, which provides 2011 Middle Layer Super Output Area (MSOA) boundary files.

### 2.1.3 Economic Indicator Data

The economic indicator data used in this analysis, including turnover size bands (e.g., 0 to 49, 50 to 99, up to 50,000+ in thousands) and industries such as accommodation and retail, is sourced from Nomis. The dataset, titled \*UK Business Counts - Enterprises by Industry and Turnover Sizeband\*, provides 2023 annual data at the MSOA level.

### 2.1.4 Hotel Points Data

The hotel points of interest (POI) data used in this analysis is sourced from \*\*SLIPO\*\*, specifically the \*gis\_osm\_poi\_free.shp\* file. This dataset, updated in 2024, provides point-based data categorised under hotel for the London area.

### 2.1.5 Tourism Points Data

The tourism points of interest (POI) data used in this analysis is sourced from \*\*SLIPO\*\* , specifically the \*london\_lates.shp\* file. This dataset, updated in 2024, provides point-based data categorised under Tourism for the London area.

## 2.2 Analysis Methods

1. \*\*Descriptive Analysis: \*\* Summarizes property types and host characteristics

- Purpose: To outline the overall structure of the Airbnb market

2. \*\*Spatial Analysis: \*\* Uses GIS mapping and spatial statistics (Moran's I, LISA)

- Purpose: To identify where Airbnb listings are geographically clustered.

3. \*\*Economic Impact Analysis: \*\* Applies regression models to evaluate economic indicators.

- Purpose: To measure Airbnb's contribution to local economic performance.

# 3.