

NA_Group1's Group Project

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.). 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.).

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.). Studies show that short-term rentals disrupt communities, drive gentrification, and exacerbate housing inequities in cities like New York, London, and Nanjing (Jiao & Bai 2020, Sun, Zhang & Wang 2021, Wachsmuth & Weisler 2018). To promote housing equity, Inside Airbnb focuses on: Increasing Transparency: Highlight the effects of short-term rentals on housing availability and affordability (Inside Airbnb n.d., **garcia_ayllon_airbnb_2018?**). Supporting Policy Development: Provide actionable data to regulate short-term rentals and tackle urban challenges (Jiao & 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). 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 & 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?

1. 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 (API Terms of Service, 2023). In addition, data collection is done at intervals, which means dynamic changes such as new or deleted lists can be missed. (Gurran & 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, 2019).

2.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 & 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 (InsideAirbnb, 2023).

3.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, 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. (Comptroller's Office, 2018)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.), 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, n.d.)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, n.d.). 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, n.d.) . 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 (Intersoft Consulting, 2018; Australia Government, 1988). 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 & 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?

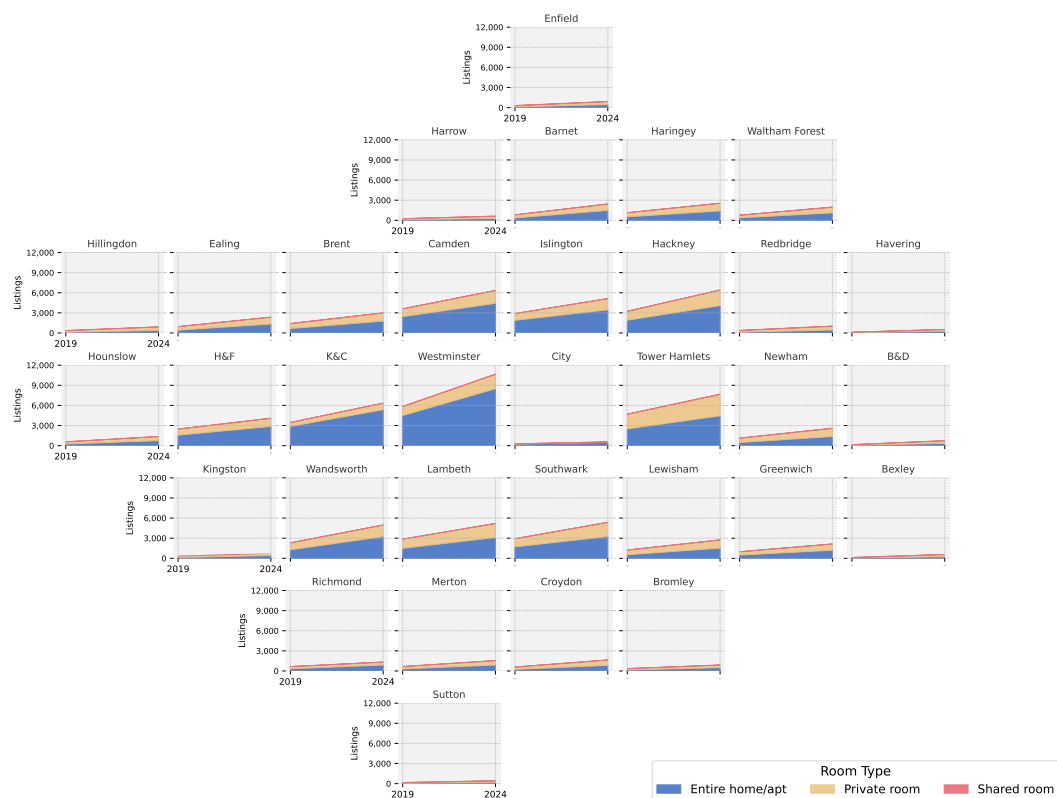
□□□□□□

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	60.11	39.25	0.57

	Borough	Entire home/apt	Private room	Shared room	Total
0	Barking and Dagenham	52	96	3	151
1	Barnet	348	462	8	818
2	Bexley	35	81	0	116
3	Brent	667	723	38	1428
4	Bromley	159	196	1	356

```
[ 'B&D' 'barnet' 'bexley' 'brent' 'bromley' 'camden' 'city of london'
  'croydon' 'ealing' 'enfield' 'greenwich' 'hackney' 'H&F' 'haringey'
  'harrow' 'haverling' 'hillingdon' 'hounslow' 'islington' 'K&C'
  'kingston upon thames' 'lambeth' 'lewisham' 'merton' 'newham' 'redbridge'
  'richmond upon thames' 'southwark' 'sutton' 'tower hamlets'
  'waltham forest' 'wandsworth' 'westminster']
```

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



Reference

Alsudais, A., 2021. Incorrect data in the widely used Inside Airbnb dataset [Online], *Tilgængelig Decision Support Systems*, årg. 141, s. 113453.

fra: <https://www.sciencedirect.com/science/article/pii/S0167923620302086> [Lokalis-
eret 15-04-2024].

Inside Airbnb, n.d. *About Inside Airbnb* [Online].

Tilgængelig

fra: <https://insideairbnb.com/about/>.

Jiao, J. & Bai, S., 2020. Cities reshaped by Airbnb: A case study in New York City, *Tilgængelig Chicago, and Los Angeles* [Online], *Environment and Planning A: Economy and Space*, årg. 52, nr. 1, s. 10–13.

fra: <https://doi.org/10.1177/0308518X19853275>.

Sun, S., Zhang, S. & Wang, X., 2021. Characteristics and influencing factors of *Tilgængelig* Airbnb spatial distribution in China's rapid urbanization process: A case study of Nanjing [Online], *PLOS ONE*, årg. 16, nr. 3, s. e0248647.

fra: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0248647>.

Wachsmuth, D. & Weisler, A., 2018. Airbnb and the rent gap: Gentrification through the sharing economy [Online], *Environment and Planning A: Economy and Space*, årg. 50, nr. 6, s. 1147–1170.
fra: <https://doi.org/10.1177/0308518X18778038>.