This is very useful :

[Data Critique – Inside Airbnb](https://airbnbdh.humspace.ucla.edu/data/data-critique/)

**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?**

**Scraping** – no guarantee of any accuracy is provided or implied – could be an incomplete picture of the data, especially as the script is no longer kept up to date and AirBnB changes its website – impacts results as they might not be accurate. It is also impossible to go back and re-scrape the results making repeatability impossible.

**Co**-**ordinates** – not Inside AirBnBs fault, but with co-ordinates being only approximate then results at best, can only be considered approximate.

**Information Processing** – it is not clear what processing or data cleaning Inside AirBnB have performed on the scraped dataset. *Data Feminism – tidying data as a destructive act, removing data from context.* Makes the extraction exercise difficult to repeat. How many cases have been removed? Why?

**Snapshot** – the data is scraped for each city on a quarterly basis using a snapshot on a given day. This means that the data could vary on a day to day basis and any analysis based on it might not be considered robust. Very difficult to assess any trends over time.

**“Strangers in the dataset”** – see Data Feminism – the dataset is not ours – we did not extract it, clean it, store or present it. As such, we are strangers to it, and the context in the process behind it is missing to us – we are coming at it with an incomplete knowledge and not all the information.

**In terms of impacts** – correlations not causation – see Batty’s comment on the complexity of the London Housing market – Brexit, Covid, Recession etc. Hard to disentangle what is what. Without community voices included it is just a data exercise and narrative, not necessarily grounded in reality.

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

**Privacy –** particularly true for those premises which are not entire houses, it is safe to assume that the host lives in the property. The data can (potentially), publicly identify people and their properties (given the description and approximate co-ordinates). *(St. Pauls example – a flat around the corner from where I live, co-ordinates are slightly off, but description actually says where it is – id 47311407 – there are only 5 flats).* This is a possible breach of privacy, and a breach of GDPR / Data Protection – this depends of course on what consent on AirBnB is used on registration and advertising. There does not appear to be any suppression or anonymisation used on the data. Especially concerning if mixed with other publicly available datasets such as the electoral roll.

**Use of personal, identifiable information** – the dataset includes the names of the host and approximate co-ordinates and description of the properties. Could breach the GDPR principle of confidentiality. Any local authority analyst should be able to work out who these people are, especially if using augmented datasets.

**Collection Use –** under GDPR (and DP) information can only be used for the purpose that it is collected for – in this case, the advertising of properties for rent on the AirBnB platform and the management and maintenance of that platform. I imagine this is incompatible with legislation and counters the consent and possible Fair Processing Notice from AirBnB. Name (and location, address of host) are considered personal information.

**Consented Use** - The use of the data for other purposes other than which consent is provided is counter legislation. The consented use is unlikely to include scraping, and any subsequent analytical or publicity relating to the data.

A host would be mad to provide consent to use their information to potentially put themselves at risk of going out of business.

**Agendas / Bias –** the collectors have an agenda and bias. To their credit, this is not hidden – “We work towards a vision where communities are empowered with data and information to understand, decide and control the role of renting residential homes to tourists”. The use of the word “impact” in their mission statement suggests they consider AirBnB to have detrimental effects on local communities.

Two key contributors are both described as activists who use data for social change, and develop communities of interest around the use of such data. These suggest that the presence of AirBnB is not considered “a good thing” and something that should be examined, analysed and resisted. Past collaborators show similar traits and interests and study the “impacts” of AirBnB.

*Data Feminism links to – reflexivity – understanding and publicising your own position and biases in analytical work.*

**Bias – anchoring –** the very fact that the word ‘impact’ is used suggests that AirBnB is having a negative effect on something, somewhere. Even before we have started work on this project we have all decided it is a bad thing.

**Incomplete picture** – *partial knowledge from Data Feminism.* – their agenda is clear, but this represents part of the picture around the business. There are articles which either suggest that AirBnB have had no or little impact on local housing markets. Inside AirBnB’s website does not mention any possible benefits of AirBnB – local employment, increases in tourism.

**Legality of scraping** – is it legal in the first place? [Is Screen Scraping and Web Crawling Legal in the UK? - Paris Smith](https://parissmith.co.uk/blog/web-crawling-screen-scraping-legal-position/), [Data Mining & Data Scraping | Free Guide - Rocket Lawyer UK](https://www.rocketlawyer.com/gb/en/business/data-protection-for-businesses/legal-guide/data-mining-and-data-scraping).

**“Do good with the data”** – who determines what is good? Good for who? [Data Policies | Inside Airbnb](https://insideairbnb.com/data-policies/). Doing “good” for local communities could well be a detriment to the hosts and the AirBnB business model, and those in the tourist and support industries that have built up around it.

**Competition** – whilst Inside AirBnB are not aligned to AirBnB or any of their competitors, there is nothing stopping competitors analysing their data (other than a claim on non-commercial use). Whilst Inside AirBnB might adhere to this in their activist, public interest role, competitors might not. How do they govern secondary uses?

**Gap of who is using it** – Inside AirBnB assumes that it is tourists using these properties – but no evidence in the dataset for that. Warning – approaches and views on housing change over time – a few years back HMOs were considered inappropriate forms of housing, they are now considered by some as legitimate.

**Supply side only** – to understand a housing market you need to understand the supply (in part supplied by AirBnB) but also the demand which is missing. Some of this will be collected by AirBnB in terms of who uses the properties, or who has searched for a property in a certain area – possibly Google Searches?

Common sense wise, the argument should be that house prices are too high, that pushes up private rents (and increases their demand), which is exacerbated by the presence of AirBnB; which then pushes demand down to social housing.

Consider a whole system approach.