Data Biography

Declaration of Authorship

I, Xianlai Yin, confirm that the work presented in this assessment is my own. Where information has been derived from other sources, I confirm that this has been indicated in the work.

Xianlai Yin

Date: 21/11/2022

Student Number: 22073407

1 Who collected the data?

By comparison, this dataset is highly similar to the dataset provided by Inside Airbnb (IA), and it can be assumed that the researchers of this project collected the data ('Inside airbnb', n.d.).

2 Why did they collect it?

The IA's research aims to analyse the impact of Airbnb on urban communities to protect them from the negative effects of short-term rentals ('About inside airbnb', 2022), including gentrification and a reduction in rental housing resources (Wachsmuth and Weisler, 2018). Therefore, this dataset focuses more on collecting information on hosts to determine whether they are commercial operators, and on the availability of properties to determine the extent of their impact on the housing market.

3 How was it collected?

IA's researchers scraped the data from the Airbnb website using techniques such as python and cleaned and collated it, which allows the dataset to contain a wide range of relevant information, almost all the information on the Airbnb listing pages is covered, and avoids the official embellishment of the data (Cox and Slee, 2016). On the other hand, this does not allow access to detailed internal information, because some of Airbnb's data such as the exact location of the listing is anonymised.

4 What useful information does it contain?

This dataset contains a range of information about listings, including property details, host information, community and location information, listing availability, and listing reviews. The dataset provides host information and listing availability, which combined with community and location information can reflect the overall status and distribution of Airbnb's listings in the city, and analyse the various impacts they have on the city or community. In addition, the property details, listing availability and listing reviews could help people choose properties and help hosts adapt their strategies.

5 To what extent is the data 'complete'?

The data itself, which inevitably has many vacant values in the dataset, does not affect the overall data analysis. However, although this dataset is widely used in various studies, some scholars have questioned the validity and accuracy of the dataset, arguing that the python script for scraping data is outdated and that the scrape process may be interfered with by anti-scraping on the Airbnb website, so that the dataset may have problems such as missing and duplicate data (Alsudais, 2021).

For the researcher's purposes, the data collection is mainly complete, and the dataset covers almost all the information on Airbnb's listing pages. However, if this is the only perspective it does not give a complete picture of Airbnb's operations in London, as it is only considered from the listing perspective (supply side). However, the picture of Airbnb's operations cannot be viewed without the perspective of the renter, and a user profile of the rental groups can provide a relatively complete picture of rental operations from the demand side.

6 What kinds of analysis would this support?

Based on the researcher's purpose of collecting this data, the impact of Airbnb on urban communities can be analysed. Firstly, the gentrification process of the city or community can be analysed by considering Airbnb's rental income as the potential income in the Rent Gap Model (Wachsmuth and Weisler, 2018). Secondly, by analysing the available information on Airbnb listings, the extent to which housing resources are wasted by using homes as short-term rentals can be estimated. Thirdly, based on the number of listings owned by the host and the acceptance rate, the analysis determines whether the host is a commercial operator, which is considered to have a more negative impact on the housing market (Harris, 2018).

The dataset can also support other analyses outside the data collectors' research aims. From a host's perspective, the dataset can be analysed to understand the market situation to develop business strategies; from the perspective of the tenant, the dataset can be analysed to determine the strengths and weaknesses of different listings to choose the suitable one; from the perspective of governments and the planning authorities, the dataset can be used to analyse the spatial characteristics of cities as reflected by the distribution of Airbnb listings and the impact of Airbnb listings on urban development (e.g. the impact on the housing rental market), so that policies and plans can be adjusted to respond (Gurran and Phibbs, 2017).

7 Which of the analyses outlined above are ethical?

Firstly, the ethical issue of data collection and use involves violating the right to privacy and informed consent (Crawford and Finn, 2015). According to Airbnb's Terms of Service statement, users are not allowed to use any automated process to scrape data from the website (Stringam, Gerdes and Anderson, 2021). Using this dataset for research is an obvious violation of this clause. Also, the information was uploaded by many Airbnb users who did not permit to use the data for various analyses, and the research using the dataset is likely to invade their privacy. On the other hand, it is also argued that permission is not required to collect and analyse publicly available data (Stringam, Gerdes and Anderson, 2021), the ethical boundaries are controversial in different regions and groups.

Secondly, the ethical issue of the impact of the research (Kitchin, 2016). The analyses mentioned earlier are mostly related to urban society, and the results will inevitably directly or indirectly affect people in society. Suppose the government, having analysed the dataset, had made a policy to ban or restrict Airbnb's operations in the city to avoid its negative impact, to the detriment of the platform and the hosts, even though the dataset used in the research originated from them and was not permitted to be used. However, the policy was developed in the broader public interest so that residents could enjoy more reasonable rents and a better community atmosphere. This involves an ethical debate between the public interest and the interests of particular groups. But what if the government uses the data more for economic benefits than for the rights of citizens? Suppose the government uses the dataset to find that a particular area is popular with tourists and decides to invest in tourism in that area, leading to higher rents and living costs for the local community and the destruction of the social fabric of the community, albeit with higher tourism revenues, is this ethical?

In summary, most of the above analyses using this dataset are ethically controversial. However, the data collection process did not collect private data, circumventing many ethical issues related to privacy. The ethical issues arising from the impact of the research involved ethical thinking at a societal level and were not confined to this dataset and analysis.

References

'About inside airbnb' (2022) *insideairbnb.com*. Available at: http://insideairbnb.com/about/.

Alsudais, A. (2021) 'Incorrect data in the widely used inside airbnb dataset', *Decision Support Systems*, 141, p. 113453.

Cox, M. and Slee, T. (2016) 'How airbnb's data hid the facts in new york city', *Inside Airbnb*.

Crawford, K. and Finn, M. (2015) 'The limits of crisis data: Analytical and ethical challenges of using social and mobile data to understand disasters', *GeoJournal*, 80(4), pp. 491–502.

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.

Harris, J. (2018) 'Profiteers make a killing on Airbnb – and erode communities', *The Guardian*. Available at: https://www.theguardian.com/commentisfree/2018/feb/12/profiteers-killing-airbnb-erode-communities (Accessed: 21 November 2022).

'Inside airbnb' (n.d.). Available at: http://insideairbnb.com.

Kitchin, R. (2016) 'The ethics of smart cities and urban science', *Philosophical transactions of the royal society A: Mathematical, physical and engineering sciences*, 374(2083), p. 20160115.

Stringam, B., Gerdes, J. H. and Anderson, C. K. (2021) 'Legal and ethical issues of collecting and using online hospitality data', *Cornell Hospitality Quarterly*, p. 19389655211040434.

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.