# Data Biography

### Declaration of Authorship

I, Zhonghao Li, 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.

Zhonghao Li

Date of signature: 22 Nov 2021

Assessment due date: 26 Nov 2021

Student Number: 21047001

\_Please write your answer immediately below the level-3 headers and delete the guidance prior to submission.\_

---

1. Who collected the data?

The data was originally compiled by Airbnb then Murray Cox built Inside Airbnb to collect and organize the data [1].

2. Why did they collect it?

To deliver quantified data, which was filtered effective public information, in order to investigate the influence of short-term tenancy on housing and residential communities and the influence of Airbnb on the residential market, this was also a platform of promoting policy propaganda to protect citizens against tenancy impacts [1].

3. How was it collected?

The data of Inside Airbnb was collected from public information published in Airbnb web site [1]. Each city was marked as a search area, the search outcomes were deposited in a database which was manipulated by PostgreSQL [2]. A survey presented a respective collection of an individual city, and a database consisted of multiple independent surveys [2]. Since some of the properties information were repeated and being inactive, the final data was collated and processed before applying [1], [2].

4. What useful information does it contain?

It applied an occupancy model called "San Francisco Model", the benefits are: approximated bookings are converted by reviews from half number of review rate; the occupancy rate could be obtained from product of the average dwell time allocated for each city and estimated number of bookings for each listing during a period; the maximum occupancy rate of Airbnb is 70%, which is comparatively high but rational; among the indicators of high availability and frequent rentals, the number of booked or available periods per year are generally in accordance with a city's short-term rental laws designed for protecting residences [1].

5. To what extent is the data 'complete'?

Firstly, the data is based on disclosure information collected from the Airbnb data, which does not contain any private information or commercial derivation, therefore, the correctness and completeness are corresponding to Airbnb website [1].

Then in most cases, the data accuracy of listing in a city is less than 20%, which refers to necessary declaration of Airbnb as well as comparing related data. The potential reasons including unappropriated search methodology that needs to exclude unavailable sites as trip dates are not considered; searhes in main cities can transfer to surrounding listings which is possibly out of the city boundary; attitude and longitude values could be applied for increase the accuracy of spatial data, while tackling this mission is a labor-intensive procedure and is unpractical in short term [2].

Finally, it is challenging to match the context data with Airbnb data since they barely issue details on quantitative reports or key terms in terms of certain cities' listings, nevertheless, the difference between these data is within an acceptable range except a few cases, and the spread of multiple-listing hosts is in accordance with each other. Therefore, the data is sufficiently reliable for policy and social analysis [2].

6. What kinds of analysis would this support?

Firstly, the data could be used for observing Airbnb's competitive modes against the residential market, and could also investigate basic questions such as the nearby listings and reviews, whether the tenants are tourist or general residents, and comparing the profit between short- and long-term rentals [1], [3].

Secondly, it could support a city on estimating relative visits and income, since it is not practical to obtain absolute results of both [2]. From research, it could be investigated that the rate of review to visit are consistent among the examined groups, and the approximated income is proportional to the absolute income [2].

Subsequently, it could be used for monitoring irregular commercial activities of Airbnb. Research was conducted that Airbnb proceeded interventions by fudging and deleting some data of real facts, in order to describe a more attractive business with their dataset and create a better impression among media and public rather than increasing transparency as they had claimed [4].

Finally, the data could be compared with different source of data to verify the number of short-term rentals actually exist within a region, governments generally lack authentic and up-to-date data to supervise and manage the rental market, this data provides a reliable analysis for policy implementation and legal safeguard [5]

7. Which of the uses presented in Q.6 are \_ethical\_?

Firstly, whether the data analysis is ethical or not, it depends on the context of using [6]. To some extent the data is ethical since the data is gathered from public information without private information or commercial derivation, the exact location of each listing is modified for protecting the privacy [1]. However, if the data is used for vicious competition such as exaggerating Airbnb's business model, irregularly increasing the rent, or even trading in privacy of listings, these are unethical.

Secondly, applying the data for observation of visits and income is ethical in some way, in order to standardize the market and industry criteria, however, healthy competition could be broken if overuse the analysis, which is unethical. Hosts might pay excessive attention to draw an engaging picture of data rather than focusing on provide a cozy and comfortable environment for customers, and might hide the truth of discriminatory rejection of tenants and breaking community rules [7].

Subsequently, the common use of monitoring irregular commercial activities of Airbnb is ethical, as the behavior of modifying real data of Airbnb could create a "Big Dick Data" and impact policy implementation and legal safeguard within a region as well as misleading the media and people, the data cannot prove themselves is correct, the context including rules, policies and laws in a particular city or country should be considered and combined with data analysis [6]. Although local governments need relatively accurate data to supervise the illegal rentals, racial discrimination and other market-destroying phenomenon, the application of unverified open data could be unethical and problematic [8].

Finally, the data is used for comparing with other data source could be ethical, if the adopted data is for legally management of government. Nevertheless, over control of the short-term rental market could impede professional property management and could be unethical, legitimate property managers could really understand customer demands and provide higher standard service for tourists, industry standards could not be improved without the contribution of these people. Government and the short-term rental company should find a balance between them to maximize the business values [5].

(Word Count: )

## Bibliography

[1] ‘Inside Airbnb. Adding data to the debate.’, *Inside Airbnb*. http://insideairbnb.com (accessed Nov. 23, 2021).

[2] ‘Airbnb Data Collection: Methodology and Accuracy – Tom Slee’. https://tomslee.net/airbnb-data-collection-methodology-and-accuracy (accessed Nov. 24, 2021).

[3] ‘End-to-End Predictive Analysis on AirBnB Listings Data’, *Analytics Vidhya*, Oct. 05, 2021. https://www.analyticsvidhya.com/blog/2021/10/end-to-end-predictive-analysis-on-airbnb-listings-data/ (accessed Nov. 25, 2021).

[4] M. Cox and T. Slee, ‘How Airbnb’s data hid the facts in New York City’, p. 16.

[5] ‘Airbnb Data Analytics In London: Inside Airbnb, AirDNA, and Airbnb Data Sets’, *https://www.rentalscaleup.com/*. https://www.rentalscaleup.com/airbnb-data-analytics-in-london-comparing-data-sets-from-inside-airbnb-airdna-and-airbnb/ (accessed Nov. 24, 2021).

[6] C. D’Ignazio and L. Klein, ‘6. The Numbers Don’t Speak for Themselves’, in *Data Feminism*, 2020. Accessed: Nov. 25, 2021. [Online]. Available: https://data-feminism.mitpress.mit.edu/pub/czq9dfs5/release/3

[7] ‘Is There An Ethical Problem with Using Airbnb?’, *The Mediterranean Traveller*, Aug. 31, 2017. https://www.themediterraneantraveller.com/airbnb-ethically/ (accessed Nov. 25, 2021).

[8] A. Alsudais, *Incorrect Data in the Widely Used Inside Airbnb Dataset*, vol. 141. 2020. doi: 10.1016/j.dss.2020.113453.