## 1. Who collected the data?

The dataset was collected by [Murray Cox](https://en.wikipedia.org/wiki/Inside_Airbnb) through automatic scraping from the Airbnb website, specifically for the Inside Airbnb project.

## 2. Why did they collect it?

The [Inside Airbnb](http://insideairbnb.com/about) project aims to provide an independent perspective, helping the public, researchers, and policymakers understand how Airbnb affects urban housing affordability and community dynamics. It offers insights for policy discussions and social understanding of Airbnb's role in urban environments.

## 3. How was the data collected?

[listings.csv](http://data.insideairbnb.com/united-kingdom/england/london/2023-09-06/data/listings.csv.gz) : Inside Airbnb collects its data primarily by scraping information from the Airbnb website. This process involves the following steps: \* Web Scraping: Inside Airbnb employs scripts to rapidly and extensively extract Airbnb listing data, imitating human browsing. \* Data Extraction: Information about each listing, such as location, price, availability and host details, is extracted and compiled. \* Data Aggregation: Aggregated data forms a database for analyzing Airbnb trends and insights across cities and regions. \* Regular Updates: The scraping process is repeated periodically to keep the database current, capturing new listings and updates to existing ones.

**4. How does the method of collection impact the completeness and/or accuracy of its representation of the process it seeks to study, and what wider issues does this raise?**

The dataset is mostly obtained by scraping information from the Airbnb website, so its breadth and depth of information publicly available on the site may be limited. For instance, detailed information about certain listings might not be fully disclosed, or website terms might restrict access to some data. Moreover, legal and ethical considerations in web scraping, such as data privacy and usage rights, may affect the integrity and accuracy of the data. The content of the website is constantly changing dynamically, but data scraping occurs at intervals, which means the data might not be updated in-real time, potentially leading to information gaps [@prentice\_addressing\_2023].

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

## Privacy issues

Whether the dataset has the consent of the owner to disclose its information, e.g., house location, name. Geocoded data is privacy-sensitive and highly likely to expose personal privacy when used to study demographic patterns and behaviours[@van\_den\_bemt\_teaching\_2018]. Therefore, It is crucial to obtain the consent of the owners to ensure that their privacy is not infringed upon.

## Legal compliance

Usage of the dataset should comply with laws and regulations such as GDPR, DPA and EDPS. The EDPS 2015 report states that it is not enough to comply with the law in today's digital environment; We must consider the ethical dimensions of data processing[@hasselbalch\_making\_2019]. Legal compliance and ethical considerations should be closely combined in the digital age.

## Social responsibility

It is critical to use the dataset correctly, as exposing certain data may result in inequity and bias. The Fairness and Openness Report(@DataJusticeReport) emphasizes how to use information responsibly and ethically, as well as the importance to resist the labelling of low-income communities, race, etc. For example, a significant gap in housing prices between different neighbourhoods may reflect economic differences, which may affect perceptions of the social status of those areas. To avoid unwanted consequences, it is necessary to examine how to disclose the tagged attributes of the data.

## Data security

Some sensitive information in the dataset, such as personal descriptions and geographic coordinates, must be stored securely to prevent unauthorized access and misuse. By adjusting the norms of network data use, it is possible to effectively guarantee data security and increase companies’ ethical behavior level when processing data(@СкибунО.Ж\_2022). Thus, attention to data security can prevent unscrupulous individuals from collecting housing data for profit or monitoring purposes.