

Final Project Proposal

Data 620 (Summer 2021)

Team Members:

Stefano Biguzzi, Ian Costello, Dennis Pong

Goal and Motivation:

We were inspired by a recent article in [Towards Data Science](#) (which was in turn inspired by a [Wired](#) article) which demonstrated that network analysis could be used to detect fake reviewers of Airbnb stays.

We took a different approach exploring a 2015 law mentioned in the article, "...which allows Londoners to let their home on a short-term basis without the need for planning permission, for up to 90 nights a year." This law, while providing a great economic benefit for residents opens the door for potential abuse through "systematizing" a single listing by signing up for multiple accounts.

Hypothesis:

Our project plans to demonstrate that network analysis through connection between guests and hosts can detect hosts who use multiple accounts for the same unit to skirt this law and rent out their space for more than 90 days. Other geographies could reapply this method to understand if systemization is an issue.

Data Sources:

The primary data source for this project is [Inside Airbnb](#), a repository of publicly available data that is cleaned and aggregated for open use. The site contains many geographies to use. We have selected London because of this distinctive law that seemingly eases up on restrictions when other cities are cracking down. Additionally, London has many more listings than any other location initially reviewed, which improves our chances to find what we are looking for.

<u>Date Compiled</u>	<u>Country/City</u>	<u>File Name</u>	<u>Description</u>
April 11, 2021	United Kingdom / London	listings.csv.gz	Detailed Listings data for San Francisco
April 11, 2021	United Kingdom / London	reviews.csv.gz	Detailed Review Data for listings in San Francisco

Restrictions of the data:

1. Hosts with multi-listings that are greater than or equal to five
2. Reviewers who reviewed multiple hosts (greater than two)
3. Hosts and reviewers where the reviewers left multiple reviews for a host

Work Plan / Roles and Responsibilities for Group Members (weights):

- Network analysis – 60% (Dennis Pong / Stefano Biguzzi)
- Text mining – 10% (Stefano Biguzzi)
- Sentiment analysis – 15% (Ian Costello)
- Video production – 10% (Ian Costello)
- Project proposal drafting and edits – 5% (All team members)

Concerns / Caveats:

- Type 1 errors (i.e., false positives), where hosts who are flagged are legitimate
- We selected a city where English is the primary language spoken so that it minimizes non-English reviews for text and sentiment analysis
- Lack of robust and identifiable data, such as IP addresses, emails, unit street addresses etc. that would assist in confirming that flagged hosts are indeed the same person(s) listing the same unit
- How the law is implemented to monitor and enforce the 90 days limit and identify those trying to skirt the law
- Only covers Airbnb data – if hosts use other platforms than we would not know if they are listing their units there