

A software system could be easily developed to check an Airbnb listing's business license. The software would simply check that the business license matched the business license number format. If a number doesn't match the format, it would get reported. Another option would be for both organizations to work together. A software could be created that takes an Airbnb listing's reported business license number and then searches the San Francisco Planning Office's database of business numbers it has issued. If an Airbnb listing has provided a number that doesn't appear in the database, action could be taken. Two arguments against these plans for a potential software solution revolve around privacy. Some people may feel that the information they give to Airbnb shouldn't be shared with anyone. On the flip side, the San Francisco Planning Office may not want to give Airbnb access to its database because it would no longer have sole security control over the information they have about the rentals, which is a privacy concern for people that have rentals saved in the San Francisco Planning Office's database.

The data collected by this project could help answer many questions about housing insecurity in the US. One question this data could answer is whether or not there are illegal rental units in certain areas. This project just focused on San Francisco, but it would be very easy to web-scrape and collect data about any other region. The data collected from various regions could be used to identify illegal rentals listed on Airbnb.

I believe that it is important to consider what web scraping does to the website that is scraped. When a site gets scraped it has to process many requests from the web scraper to provide the information that is being scraped. This is an important factor to think about because if tons of people are scraping the same website, a website could get overwhelmed with the number of scraping requests combined with the number of regular website requests. This could cause the website to crash, which would undoubtedly hurt the business because their website would no longer work, resulting in that company losing business.

When determining whether or not to scrape public data, it's important to think about whether or not you can ethically access the data to scrape, and you need to think about the consequences of the data you scrape ending up in the wrong hands. Although data can be public, in some cases it's not legal to get that data by scraping it from the web. Therefore, as a data scientist it's important to know if the means by which you accessed the data is legal or not. Also, you should think about the potential consequences of the data you scrape getting used in negative ways. Although it's public data, it's your responsibility to keep people's data safe while it's in your hands.