

- A. Throughout this project, we acted as investigators to uphold the system of accountability created by the San Francisco lawmakers: listers must register with the city's planning office and put the business license's number on Airbnb's website, Airbnb must display some effort in validating these policy numbers, and third parties can register a complaint of illegal short-term rentals with the city planning office. We used web-scraping to do the latter using several hours of our personal time.

Imagine you're a software developer at either the San Francisco Planning Office (SFPO) or Airbnb.com. Describe a different system that verifies that the business license is valid for short term rentals in San Francisco and list at least two arguments you might hear at your organization (either SFPO or Airbnb.com) against adopting your system.

A different system that verifies the business license is valid for short term rentals in SF would be for SFPO to develop a database of approved short term rental licenses for Airbnb to reference via API when a new listing is posted to determine the legitimacy of the license entered. Airbnb could try to navigate the issue of exemptions and pending licenses by requiring special cases to be verified by Airbnb or SFPO reps. The first argument I might hear is that the database would need to be updated often and untimely updates could delay the process of getting approved, which would have a negative impact on the rental market, causing Airbnb to lose revenue and increasing rental prices for consumers. Another potential issue with this solution is that it would require verification of pending and exempt listings which would require additional resources and attention. There is also a potential issue in who is accountable for developing the verification system for new listings and there would also need to be some sort of system or protocol for reverifying licenses as they expire.

- B. The database we've created through web-scraping is a great data source of information for data scientists in order to answer and explore research questions. Skim through the Housing Insecurity in the US Wikipedia page and describe at least one research question that you could answer or explore using this data if you were a data scientist working with a housing activist organization to fight against housing insecurity. I've heard that some argue that Airbnb's take up housing from people who need stable housing and can drive up prices. This is problematic because people who can afford housing continue to get richer off their Airbnb properties while those with housing insecurities are left without homes. It could be interesting to compare data regarding the number or percentage of people who face housing insecurity to the number or percentage of Airbnb's in that same area.
- C. As discussed in the introduction, the legality of web scraping is still uncertain in the US. Skim through the Legal Issues section of Web Scraping in the US on Wikipedia and this article about the legal issues with the Computer Fraud and Abuse Act, and describe at least one factor you believe is important to consider when discussing the legality of web scraping and why. I think that it is important to consider why companies and commercial organizations or websites might not want people to be able to scrape their data and what they could be hiding if their sites are not allowed to be scraped. In the article about legal issues, researchers, computer scientists and journalists were looking to scrape website data in

order to test the equity of outcomes when algorithms are being used. If commercial sites are able to hide this data its likely that they will not be held accountable for intentional or unintentional discrimination in algorithms. Thus, the decision regarding legality is critical for protecting beneficial research and journalism and is also important in the ongoing fight for algorithmic transparency and accountability.

- D. Scraping public data does not always lead to positive results for society. While web scraping is important for accountability and open access of information, we must also consider issues of privacy as well. Many argue that using someone's personal data without their consent (even if publicly provided) is unethical. Web scraping requires thoughtful intervention, what are two or more guidelines that must we consider when deciding to use or not to use public data?

One guideline that we must consider is the privacy of individuals when using public data. Web Scraping could threaten privacy and security of individuals and would allow people to make their own inferences from web data. Since individuals have not consented to their data being used in that way, we must consider how we are portraying them without findings and how the conclusion we draw from their information could affect them. Additionally, we should consider how we will display the information we collect. Passing off data from the web as our own would be unethical.

The last guideline we should consider are the social implications of our discoveries. For example, in our first project we worked with SAT data as well as census information. If we were to publicize that information without any context or explanation, harmful conclusions could be made and the data could be manipulated or twisted to push harmful rhetoric about marginalized groups in society. We must be thoughtful in our portrayal of groups and how their information could be used against them when web scraping.