

PROJECT 2 Questions

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.

As a software developer at the San Francisco Planning Office (SFPO) another system that could be used to verify that the business license is valid for short term rentals is by requiring listers to go through a short verification process with their policy numbers prior to listing their listing on air bnb. By requiring a "holding period" for listers SFPO would likely weed out listers who are attempting to list an illegal short-term rental and provide time for SPFO to more thoroughly validate policy numbers prior to listing. Two arguments that may work against this system is the complaint that because of the urgent need for housing in San Francisco it is not beneficial for the city to have to wait prior to new listings being uploaded as well as the system being unappealing to new listers who are looking to get their listings up quickly (which in turn may drive away new users).

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.

After skimming through the Housing Insecurity in the US Wikipedia page one research question I developed is how housing insecurity relates to safety rates by state. I would be interested in scraping the site to collect information about the rates of housing insecurity in each state compared to the average rates of safety per state and see how the two are connected and how working to increase safety could potentially minimize the housing crisis. .

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.

One factor that I believe is extremely important to consider when discussing the legality of web scraping is what the web scraping is being used for and if its benefits outweigh the risks of violating terms and services. In the example of “Sandvig vs Bar” at the ACLU web scraping was being used as a means to defend against intentional and unintentional discrimination, a deeply important issue. When examining web scraping in cases in which the data that is scraped is used to protect rights in fights such as the ongoing fight for algorithmic transparency and accountability, I believe it is far more important to look past the legality of web scraping in favor of using data to support necessary causes.

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?

In an effort to increase privacy in the realm of web scraping and reduce the risks of unethically sourcing information I think it is essential for researchers to examine whether or not users were explicitly told their personal data may be used and consider reaching out to users again if their personal information is needed to conduct research. I think the biggest issue with web scraping public data is the lack of consent and awareness of what the data is being used for from a user perspective. By better communicating intent I believe web scraping of personal data could be more ethically used.