

- 1. 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.**

If I am the software developer, firstly I would make sure that only license numbers that are in valid format could be displayed on site, and it is required for every lister. This action would show customers that all the rental places on Airbnb are legal and safe, so that they would be more likely to choose listings on it. Also, to make this process more efficient, Airbnb could collaborate with SFPO in a way that automatically transfer the license number from SFPO to Airbnb's database once the system detected the same personal information of the lister. However, this might be hard to achieve since the SFPO side may reject such data transfer to protect their user privacy. At the same time, forcing every lister to report their license number may create delay on new rental listings, so that the listers may turn to other rental platforms, resulting in a loss of revenue, which may cause this system be opposed by Airbnb managers.

- 2. 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 reading through the site, I found there are many useful information about housing security across the country. Firstly I learned the definition of an insecure house, which describes those inadequate houses with problematic structural conditions like "no running water", "electricity is not used", "holes on the floor" and so on. One question that could be answered from the data is the relationship between house renting status and households insecurity in different states in the US. As the data shows, rented

households are more likely to report a higher percentage of insecurity, especially in those states that bedroom rentals are extremely expensive like California. The site also introduces several factors that could impact people's housing insecurity feelings, including race and these could be used to investigate these circumstances further.

3. 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.

Based on my understanding, I think one important factor to consider when trying to use online data is intent of using online data. Just like what was discussed in the article, sometimes, violations of a website's terms of service cannot be ground for criminal liability. In this case, researchers created multiple test accounts with fake information, but they did it for studying the online algorithm discrimination problem, not the worst cases that CFAA meant in their terms of service. Therefore, to better acknowledge the web scraping rules, the website owner should make their terms of service more specified and brief, while the researcher could also access the valuable information with protection of it.

4. 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?

Firstly, I think it is necessary that to let people know if their personal data will be used in researches, even if they posted it online does not mean that they would like to let it interpret by many others. Also, if people found their information was used without consent, they should have the right to ask the researcher to remove their data with no difficulty. Before scraping the site, it is important to make sure that the collected data that is protected by copyright. All the scraped data should only be used for personal use and analysis purpose. If the researcher would like to publish the collected data, they need to request the permission from the data subjects and check the site policy, making sure they did not violate any of the rules.