

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 SFPO, a different system that could be implemented is one that takes an Airbnb business license number and parses it through the SFPO's internal database instead to ensure that there is a match. This system could allow for more accurate confirmation of a business license's validity.

One argument against this system would most likely be the issue of privacy and granting access to such records. It could pose risks in information security and open the opportunity for hacking and other potential threats to occur. Another argument could be the undesired ambiguity in who then becomes responsible for validating policy numbers. In terms of the current system established by SFPO lawmakers, Airbnb is expected to be the one displaying efforts in validating policy numbers. Thus, SFPO workers could be displeased with the fact that this new system blurs the defined roles that were outlined in the previous accountability agreement.

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.

Through data, history, and experience, it is known that housing insecurity disproportionately affects people of color. Taking the different factors of race and ethnicity, I want to delve into why this is. By analyzing the different factors that are listed on the Wikipedia page like health, education, and a variety of other factors that affect housing insecurity, I would research how all these issues come together to cumulatively affect people of color.

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 believe the issue of trespassing to chattels, which involves a computer system itself being considered personal property upon which the user of a scraper is trespassing, is important to consider. It would be one thing if the owner of the data knew and had consented to their information being used but trespassing to chattels is the intentional interference with another person's lawful possession of a personal property. Essentially, information is being taken without mistake and without the owner knowing. If this were to be lawful, there wouldn't be such a thing as privacy anymore as private data would belong to anyone. Although it is not stated by law that privacy is a human right, it is morally wrong and takes away a human's need for privacy.

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 important guideline is consent. An ethical action is one that best protects and respects the moral rights of those affected. We must ask ourselves if the owner of the information is aware that their information is being used, what type of information is being used, and whether they have given their final consent. Without consent, this becomes a situation in which an individual's right to privacy is violated.

Another important guideline is intent. Through the utilitarian ethics lens, if the information is being used for the purpose of the greater good of the community like education for example, it will ultimately be beneficial to a large number of people. However, if the information is being used for non-beneficial or morally invalid reasons like identity fraud, it will only hurt the owner and will not contribute any good to society.