

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 I would offer a system that acted as a database that would store and verify policy numbers. Another function of the software could notify the user that their policy is expiring ensuring that the businesses update their policies as they need.
- The first argument from the SFPO side lies within the issue of letting the user or business have the power. There could be issues with user error and policy numbers having typos. My retort would be that individuals are responsible for updating their driving information each year. As time goes on this system could grow to be as reliable as the DMV.
- Secondly, there may be an issue with enforcement, what does enforcement look like on the SFPO side? Would putting the responsibility of adding policy numbers to a database directly by the user make it trickier to hold the business accountable?

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.

- This data could give insight on the amount of rentals that have patterns showing long periods of vacancy. These trends may have unhappy airbnb owners tied to them. This could be an opportunity for the owners to switch to another platform like padsplit where the property is rented as long term rentals and supports people who may be at risk of housing insecurity.

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 issue that I found was important is a trespass to chattels especially if certain democratic systems move online like voting. If this action was made available online and if ill intended organizations made a ton of requests to that site and the site crashes that could impact the outcome of an election or cause distrust within voting remotely.

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?

-The issue of web scraping is not a black or white situation, each web scraping mission should be analyzed to mitigate any harmful outcomes it could have. The first Guideline to keep in mind when using someone else's data is if there is a public API then use it and avoid scraping all together. If an API is not supported then the programmer can check the robot.txt file to find out whether or not they are permitted to scrape data from the site (Dilmegani, 1). Secondly, implementing a User Agent String is a great way to practice transparency anytime you're scraping. Within the UAs the person scraping can include their intentions behind using the data or provide a way to contact the programmer if the owner of the site has any concerns.

