

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.

A different way to verify business licenses and lessen the workload for outside parties would be for the SFPO to have to verify the license before Airbnb can list the rental on their website. Right now SFPO only checks the legality of a business license if someone reports it as possibly invalid. If I was working for the SFPO, however, I could see this idea being rejected. One possible argument against this is the amount of resources and time the SFPO has to deal with checking these business licenses. I'm sure workers already feel overworked and underpaid so they would likely not want to take on this new project that would be time consuming, and hurting new people for this task would be expensive. Another argument could be if this would even be beneficial for the SFPO to do. Although finding these invalid licenses for outside parties is time consuming, the SFPO likely doesn't really care much about this. It does not seem like SFPO faces any legal repercussions from letting Airbnb post these invalid listings as long as they check it after someone else reports it. So, why wouldn't they let outside parties continue to do the hard work.

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.

Rent in San Francisco for a one bedroom is one of the most expensive in the United States. Researchers could use our data to answer: Are companies like Airbnb taking away affordable housing from people who need it, or would these rentals be too expensive anyways to help with housing insecurity in San Francisco? Someone could look at the price per night and the location of the rentals to see how expensive they would be to rent long term for people if they weren't available for short term rentals. Looking at the ones that have invalid policy numbers would be important because these are the ones that could be reported to Airbnb because the owner did not obtain a legal business license. This could also be applied to other cities in the future if more data was collected.

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.

When discussing the legality of web scraping I think it is important to consider the ambiguity around the Computer Fraud and Abuse Act and its original purpose. The CFAA was created in order to stop criminal activity that could hurt someone in a serious way through a website being hacked. However, this law is very vague and prosecutors have tried to use it to target anyone that has done something that a company just doesn't really like or they feel may hurt their revenues or public view. Web scraping can be really helpful for researchers and journalists and they are often doing it to find information to help a group or bring problematic information to light. However, because of this blurriness of the CFAA it is important for researchers, or anyone who wants to scrape a website, to make sure they understand the rules of a particular website until there is a more clear understanding of the legality of web scraping.

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?

When deciding to use or not to use public data, one important thing to consider is how anonymous this data really is. Some people may argue that almost no data is truly anonymous because when taken with other data from somewhere else it can be easy for some analysts to figure out who someone is. Make sure that the data being used won't put anyone in danger. Additionally, data that was collected for one reason shouldn't always be used for something else. Just because someone consented for their data to be used and displayed publicly, it does not mean they want their data to be used for another reason. Getting consent is always important when collecting data.