

**Caleigh Crossman**  
**SI 206 Project 2 Questions**

- 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 the Airbnb 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 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 against adopting your system.**

As a software developer, I could set up a system that allows listers to enter their business license and check to make sure it is valid and authorized for short term rentals. After the license is verified it could give the lister a code to enter on the website they are listing their property on. This would take place before the rental is listed on a website such as Airbnb.com and the rental couldn't be listed without a code. There may be pushback from the organizations about adopting this system because it would be a longer process for someone to list their rental on their website. In addition, Airbnb may lose money because no one would be able to list a property that doesn't have a proper business license for short term rentals. In combination with checking the business licenses, there could be a system in place that automatically removes listings (the ones who don't have a license and plan to rent for less than 90 days a year) that are rented for more than 90 days a year.

- 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 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.**

One research question that could be answered or explored using this data if I were a data scientist working with a housing activist organization to fight against housing insecurity is "Does the bathroom have a bathtub or shower?". According to Housing Insecurity in the US, people are deemed housing insecure and the house inadequate based on conditions, one of which is if they have a bathtub or shower. This question could be answered with the data by looking through the property page to find the information/description on the bathrooms in the house and then separating which houses do and do not have a bathtub/shower. Then, the housing activist organization could take action to help those who do not have a bathtub/shower.

- 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.**

It is important to consider what kind of information is being extracted when discussing the legality of web scraping. For example, there is a difference between scraping publicly available information such as a public instagram account versus private information that is protected. Information that is publicly available usually comes from a user who decided to make it public and therefore has consented to this data being extracted and shared. On the other hand, private data such as a passport number that is protected by a username and password is not something that the user has made publicly available, and therefore probably doesn't want to be extracted. Thus, I think it is important to consider what kind of information and data is being collected when discussing 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 use public data?**

Many people may argue that web scraping and using people's personal data without consent is unethical. When deciding whether or not to use public data, we should consider a few guidelines. First, we should consider what type of data is potentially going to be used. Like I talked about above, when scraping data I think it is ethical to use data that the user has made available to the public because they have consented to their information being public. It is important to consider if the user has consented to this data being public or not before using it. However, if it can be avoided and there isn't a legitimate reason to, then personal data should be avoided when web scraping. Another guideline that should be considered is the privacy policy of the website. The privacy policy for a website is it's rules for the collection and use of it's users personal data so this is important to consider and follow when scraping the website so that you are only extracting data the user consented to.