

a. 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.

Everytime when a person wants to post a unit on Airbnb.com, the info entered will be transferred to SFPO and automatically checked. If the info is invalid, that person will not be able to post it.

SFPO might say this system increases the burden of servers of SFPO, and Airbnb might think this design will decrease the number of listings on their website & prevent users from posting their units on the website.

b. 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.

Research Question: What percentage of housing has been provided as illegal rental?
We can figure out the total number of invalid business licenses online through web-scraping, then calculate the percentage of illegal rental services using the total number of invalid licenses divided by the total number of units in that area.

c. 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 think the use of scraped information can be important in the discussion. If the info is used to harm the interest of info providers (i.e. making the website stop responding), web scraping can be illegal. Also, we need to consider if the info obtained from the website is open to anyone or just a specific group of people.

d. 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?

1. The potential outcome caused by the use of data.
2. The willingness of data providers.
3. People should also consider if the data is reliable, or it might produce biased results.