

Angel Huang
SI 206
November 15th, 2022

Project 2 Questions

- a. 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.
 - i. If I was a software developer at either SFPO or Airbnb, I would suggest a system similar to our `check_policy_numbers` function. By entering these license numbers through this system, we are able to validate whether or not the business license is valid. After going through this system, we could double-check by going through a second round of verification with the other organization's database (either SFPO or Airbnb). If the license number cannot be validated, we could reach out to the homeowner for additional information. An argument that could be formed against adopting this kind of system could be the time it takes for the verification process to run. For instance, there are times when individuals could have entered the wrong number, which could slow down their process. In addition, the long waiting time for users could stop them from wanting to list their rental. Another argument could be database sharing. There could be sensitive information that organizations might not want to share with each other, which could make it harder for the validation process to go through.
- 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.
 - i. There are a lot of different states with different rates of housing insecurity. One research question that I could answer or explore using this data could be what factors are causing some states to have a higher rate of housing insecurity than others & what are some things we can do to change that rate.

- 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. I think the most important factor to consider when discussing the legality of web scraping is the violation of the Computer Fraud and Abuse Act ("CFAA"). This act relates to accessing another computer without authorization, which could also be seen as a breach of the website's user agreement. As a result, this could invade another user's privacy, which could raise a lot of legal issues.
- 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?
 - i. I think two guidelines that we must consider when deciding whether or not to use public data is the consequences it could have and whether or not the data is extremely sensitive. For instance, if accessing someone's personal information could have negative consequences, such as a breach of privacy, we might have to consider other options for retrieving similar data. In addition, if the data is extremely sensitive, then we might also have to think of other ways of accessing that information or ask for the user's consent before scraping that data.