

SI 206 Project 2 Reflection

- 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. As a developer at one of these organizations, propose a different system or policy to your organization that verifies if the business license is valid for short term rentals in San Francisco. List at least two arguments you might hear at your organization (either SFPO or Airbnb.com) against adopting your system/policy and how you would respond to these arguments.

Before being allowed to rent, businesses must be approved by Airbnb to check for validity of the license. To do this, we could potentially use an automated API that checks this validity. The San Francisco Planning Office (SFPO) may argue that this is time consuming. However, in the long run, this is the best option to ensure renter protection and to avoid related law issues for Airbnb.com. SFPO may also argue that this implementation is costly, as adding a new program is not cheap. Nonetheless, the cost of the system can break even if the companies such as SFPO or Airbnb.com consider the money that can be lost when third parties register complaints of illegal short-term rentals.

- 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 the dataset you just created if you were a data scientist working with a housing activist organization to fight against housing insecurity.

If we were data scientists working with a housing activist organization to fight against housing insecurity, we could explore the use of hotels as a solution or problem to homelessness. Using the location id's and rate of each Airbnb listing, we can identify if such costs are viable for low income people to house in, or if these are exorbitant prices that further fuel the issue of housing insecurity.

- 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.

A very important factor to consider when discussing legality of web scraping is the purpose behind it. This is because to regulate powers and ensure transparency from businesses, it may be necessary to allow web scraping. However, if a third party is trying to leverage from the website information, then web scraping should be scrutinized under law. The intent and purpose matters.

- d. Scraping public data does not always lead to positive results for society. For example, look to Facebook–Cambridge Analytica data scandal - Wikipedia or Clearview AI - Wikipedia. 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. Create at least two guidelines that we must consider when deciding to scrape or not to scrape public data.

One guideline to consider when scraping is retaining privacy and being responsible with the data acquired. For example, if there is personal information of citizens, it is crucial to consider keeping that information safe or redacted. Another guideline to consider is the fairness in capitalizing off of the information. How much will you gain and how much will the original data-holder lose? This is important because it can avoid potential lawsuits if the data scraping is contentious.