

Turn in your answers to these questions as well as your code.

- 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.

If I were a software developer for San Francisco's Planning Office (SFPO), I would develop a system that would detect listings with invalid policy numbers or listings whose policy numbers that are "pending". Only properties with valid policy numbers or properties that have received exemptions should be listed on Airbnb's website since "listers must register with the city's planning office and put the business license's number on Airbnb's website". I would create an algorithm with similar functions to those we created for the project; however, in addition to detecting policy numbers that do not follow the correct form (20##-00####STR or STR-000####), the system would also detect policy numbers that are "pending". One argument against adopting this system that might be raised by Airbnb is that listings with pending policy numbers should be allowed on their website. Airbnb wants to have the greatest number of listings possible because the numbers of listings impact their profits. The city's planning office is the party with the authority to assign policy numbers to registered properties. Once they properly registered their property, the listers of the pending properties have no control over when they will receive their policy number. Therefore, it might be argued that properties with pending policy numbers should not be penalized (not be able to list their property) for something that is in the hands of the city's planning office. A second argument against adopting this system that might be raised by SFPO is that this system is not the most efficient or convenient based on the terms of service of Airbnb. In order to not violate their terms of service, the system is run based on several static pages that were obtained from airbnb.com legally. As more listings are added to airbnb.com, SFPO will need to ensure they are web-scraping in a way that abides by airbnb.com's terms of service.

- 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.

The Wikipedia page explains the impacts of housing insecurity and the facts about the way housing insecurity impacts health stood out to me. Specifically, "housing insecure

respondents were approximately ‘twice as likely to report poor or fair health status’ compared to those who did not report being housing insecure”. Based on this, 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 is to what extent does housing insecurity impact the health of the residents? Does its impact vary state by state? Does its impact vary depending on whether it is an owned or rented house facing 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.

One factor I believe is important to consider when discussing the legality of web scraping is the reason why companies and organization are so against it; what might they be trying to hide? The article, *Federal Judge Rules It Is Not a Crime to Violate a Website’s Terms of Service*, explains an example of using web scraping to evaluate the equity of algorithms’ outcomes. This type of work holds companies and organization accountable for things like intentional or unintentional discrimination in algorithms. Therefore, although there are privacy concerns as it relates to web scraping, it is important that it be legal to web scrape public data as it can have positive results for society.

- 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?

One guideline to consider when deciding to use or not to use public data is only taking the data that you need and always have a justifiable and good reasons for getting the content to begin with. Additionally, make your intentions clear and only web scraping when necessary. Having a good purpose can help ensure your actions are ethical.

Another guideline to consider when deciding to use or not to use public data is understand the terms and conditions of the site. Although according to the article *Federal Judge Rules It Is Not a Crime to Violate a Website’s Terms of Service* “violations of a website’s terms of service cannot be grounds for criminal liability under the CFAA” (Gilens and Williams 2020), it is of ethical practice to at least understand and acknowledge the rules.