

- 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.**

A different system that verifies that the business license is valid for short term rentals in San Francisco could be to create a website where listers can enter their businesses' licenses numbers, and the site will automatically validate the policy numbers, which would save time and human effort. One argument that I might hear at my organization against adopting this system is that the computer could experience errors in validating due to unforeseen glitches, edge cases that weren't accounted for, or a lack of updates in the system. Another argument I might hear at my organization against adopting the system is that the website may be prone to hackers or data leaks, which would pose a major issue for multiple reasons. For one, someone could hack the site to make sure their policy number is validated. Another instance could be that personal information gets leaked and used for malicious purposes like blackmail.

- 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.**

One research question I could answer or explore if I were a data scientist working with a housing activist would be "Is there a correlation between racial identity and housing insecurity?" I would want to explore this because systemic racism can make it more difficult for some racial groups than others to achieve secure housing, and I would want to figure out where to target housing assistance efforts.

- 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 that I believe is important to consider when discussing the legality of web scraping is scraping personal data or intellectual property. As discussed in the Wikipedia article about web scraping, "In the United States, website owners can use three major legal claims to prevent undesired web scraping [such as] copyright infringement." I think this factor is most important because people can build robot scrapers which can trespass and steal data that can violate a site's terms of service and violate the trust of users. Many websites have APIs that they use that you have to pay to use that will return useful data for similar analysis to ours. Web scraping bypasses the need for APIs, and this could be considered as unethical. Companies spend a lot of time and money on maintaining those APIs for their uses, and this could be considered "pirating" a paid service. In addition, APIs allow companies to control what data they want to be accessible by others, and web scraping circumvents these restrictions.

- 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?**

Despite the web scraping of public data being legal, it should be approached with caution. One guideline that should be considered when deciding to use or not to use public data is the risk of hackers committing phishing attacks. Hackers could use web scraping to create convincing messages that scam victims into revealing information. Also, gathering data for the purpose of blackmail or extortion could occur, and measures need to be put into place that protects the security of consumer data. Another guideline that should be considered is that web scraping should not be used to duplicate the business model of an existing service. Copying an existing business by using their exact data is not only plagiarism, but unethical as the data is not being used to productively contribute a new idea.