1. **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.**

One system that could verify that the business license is valid for short term rentals in San Francisco is to have a program that uses regex like we did in web-scraping when the lister on Airbnb.com enters the information. Similar to how programs on websites force users to create passwords that follow their guidelines, there can be a system that uses regex to ensure that policy numbers follow the proper formatting when being inputted by the host. I believe it should be Airbnb’s duty to validate their policy numbers, and thus, this system should be implemented on their website.

However, one argument that I might hear from Airbnb against adopting this system is the loss in revenue for Airbnb with such corrective procedures. One way the company makes its revenue is from charging hosts service fees. As a result, even if a listing is illegal, Airbnb is making money from the listers using Airbnb’s platform. Therefore, by reducing the number of illegal listings on the platform, Airbnb will take a loss in revenues.

Likewise, another argument against this system that I may hear from Airbnb is the customer backlash that could ensue. With such systems, customers may choose not to use Airbnb to post their listings if the process is too complex or time consuming for them. Moreover, we saw how there are some listings that have pending policy numbers or are exempt. Thus, it is important to make sure our system can capture all the various forms of the policy number a lister may enter. This may not only be complex in implementing, but if not done properly, can lead to poor customer satisfaction as hosts’ legal listings won’t be able to be posted.

1. **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**](https://en.wikipedia.org/wiki/Housing_insecurity_in_the_United_States) **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 that I could answer/explore with our current data is “What regions of the US are associated with low housing affordability?”. In our data, we are able to view the prices of listings on Airbnb. We can sort such data by the locations in the US and by pricings in order to see where housing insecurity tends to be the greatest in the US. Such insights can help housing activist organizations to know where they should be focusing their time and resources, as well as aid people in either being able to afford such housing or influence policy makers in lowering housing prices.

1. **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**](https://en.wikipedia.org/wiki/Web_scraping#United_States) **and** [**this article about the legal issues with the Computer Fraud and Abuse Act**](https://www.eff.org/deeplinks/2020/04/federal-judge-rules-it-not-crime-violate-websites-terms-service)**, and describe at least one factor you believe is important to consider when discussing the legality of web scraping and why.**

One factor that is important to consider when discussing the legality of web scraping is the purpose for why someone is using web scraping. In our case, web scraping on Airbnb was to make sure that the business is not contributing to housing insecurity. I believe if the purpose for web scraping is ethical, such as to prevent discrimination or hold corporations accountable for their actions, then it should be allowed. However, if web scraping is being used in harmful ways such as “the defendant's unauthorized use cause[s] damage to the plaintiff” (Wikipedia), then web-scraping should be punished. Overall, I believe the legality of web scraping depends on the circumstance and if it does more good than harm for society.

1. **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 whether to use public data is to be clear of what the purpose of scraping public data is. If the purpose is for the good of society, then I believe it should be considered ethical. However, if it is for personal gain such as to make money or invade privacy of others, then web scraping becomes unethical.

Another guideline to consider is the type of data being scraped and the privacy guidelines already in place. For example, if the data consists of confidential patient information, then web-scraping of such data is unethical as this data is private to the patient only. However, if the data being scraped does not have high levels of privacy and consent to use the data is publicly provided, then the data becomes more ethical to scrape.