*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 way I could ensure that the business license is valid for short-term rentals in San Francisco, as a software engineer for Airbnb.com is to validate before each transaction. Suppose I create an automated system to double-check the business license of the short-term rental a customer is trying to check out. In that case, I will have multiple touchpoints and various tests throughout the year to ensure the license is valid and the rental transaction can be approved. From Airbnb.com, I may hear the argument that this worsens customer experience. If I find that a business license is invalid, then I have to respond to the customer's needs through a refund or void the transaction, and it may take longer for someone to check out because this process to validate the business license may take longer. Customers may not be inclined to use our service if they continuously have to wait for business license approval, are continuously being sent refunds, or have to change their plans, as to where they are living. Another argument I may hear is that this process will require us to hire new software engineers, invest more money, and be in constant contact with the San Francisco Planning Office, all of which cost money and create overhead for the company. It may be seen as unnecessary since we validate before the listing is posted, and my new strategy is overkill for a problem that isn't entirely on us to solve, however, we take much of the cost and it dips into profits.

*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.*

Obviously, housing in San Francisco is incredibly expensive. Seeing that Airbnb is often a way to help pay for rent, I would love to test how much Airbnb lister depends on their room rents to actually pay their own rent. I think seeing that even renting a bedroom can cost, $200+ a night, it's clear that the value of the house itself is expensive. If I can research the percentage of total Airbnb revenue that listers take to pay their own rent, it would be useful for a housing activist organization to try and get government subsidies and government intervention into the way people pay for housing. Housing insecurity also seems to have a racial bias, so examining the makeup of the Airbnb listers demographically would be valuable research for me to understand if Airbnb listers are actually those who suffer from housing insecurity as I hypothesize. I would obviously need a bit more data, but the data I have gathered now will surely be a starting point, and scraping these HTML files further will help me gather more foundational information for my research question of how Airbnb accommodations help pay for listers' housing rent.

*As discussed in the introduction, the legality of web scraping is still uncertain in the US. Skim through the Legal Issues section of the 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 issue that I find to be increasingly important in the legality of web scraping is identifying what the definition of personal property is when it comes to the internet. Because the internet is designed to be open-source, is anything really private? Is it meant to be a public forum? I think these discussions are crucial because if we define personal property in a more strict form, then we have more violators "trespassing to chattels". Many companies and data collectors will have to change their entire business models if that is the case. However, if we define the internet as more public, less private in property form, then scraping can become a common source of data gathering for companies and people alike, without it being considered illegal.

*Scraping public data does not always lead to positive results for society. While web scraping is important for accountability and open access to 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 important guideline we need to determine to use or not use with public data surrounds the communication behind it. If web scraping is done to oust someone, or done to work as a "canceling" of someone or something you don't like, then it shouldn't be used. Web scraping can be used to collect private data, but the way that information is communicated and how it impacts entities must be thought of. This is discussing what we do with the information surrounding web scraping, and guidelines need to be produced and considered because, in the end, it has a bigger impact when the information is communicated, over someone who is just web scraping for the fun of it. The motive behind the communication of public data is crucial. Next, I think another guideline we must consider when deciding whether to use public data or not is the source. Does it come from an individual or a public entity like a government? Setting guidelines for where we can receive public data from and what we can web scrape from will help us tackle issues of privacy. What falls under a human's right to privacy vs. a society's right to know. These guidelines will help keep some websites and personal data off limits, while clearly holding certain entities accountable for ethical behaviors.