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.

I believe a more coherent way in verifying that the business license is valid for short term rentals in San Francisco is by looking directly through SFPO licensing data. In doing this, we avoid the issue of fake policy numbers – our web scraping is validating that the policy numbers match a specific format, yet it does not validate its actual validity. Two arguments against this may be that we are unable to access SFPO's licensing data or that it may take awhile to be given permission to look through. In this case, we would face a road block because direct access to SFPO's data would be our best bet to ensure these policy numbers are accurate and valid. Yet, if this did not pose an issue, then we would be able to cross-check between the policy numbers on Airbnb and SFPO. Another argument would be the time it takes to cross-check these policy numbers. We would need to come up with an alternate way to automatically cross-check these numbers to avoid this critique. These may be reasons why our organization is against this system, yet if we can overcome this controversy, we would successfully register these policy numbers.

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.

Using our data, a research question we can ask is what proportion of housing is being registered with Airbnb in San Francisco and what price are they paying in comparison to housing costs in San Francisco? We could utilize this data and define the ratio of annualized housing costs to annual income and compare this to annualized/monthly Airbnb costs to annual income. This would help us determine if Airbnb or houses in San Francisco fit the definition of housing affordability, painting us a bigger picture of the living requirements San francisco entails. This could help us navigate whether it is more affordable to utilize Airbnb or rental housing, helping consumers make a more informed decision. Additionally, this would allow us to express the total

price of living conditions in San Francisco, informing residents outside of San Francisco prior to moving there.

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.

I think it is extremely important to consider the legality of web scraping. For example, this wikipedia page discussed Farechase's usage of data from American Airline's server that collected this publicly available data. The issue with this is that regardless of who is web scraping data, the usage of this publicly available data should not be legal. Social media and technology are substantial and prevalent in this day and age, making personal data widely available. However, the usage of this data should not be allowed without definitive permission from individuals. This must be considered when discussing the legality of web scraping. It is a fine line because individuals willingly allow their personal data to be displayed online, yet in most cases, people do not realize how much data is actually accessible. Web scrapers should only be allowed to access information that is not only publicly available, but is definitively confirmed to be used by specific companies or individuals. Otherwise, it is an invasion of privacy, distributing and scraping through unwanted information.

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?

Scraping public data can be unethical. This is a huge problem in today's society as our data is circulated constantly through various companies, platforms, and individuals. However, sometimes web scraping is important and can benefit the public. It is hard to determine where this fine line is, as it is usually blurred. In order to decide whether web scraping is ethical, we should follow these guidelines. First, one should ask themself if it is benefitting or harming the individual. Then, if it is beneffiting an individual, they should ask if it is benefitting or harming the public as a whole. Following this, the last guideline to follow is whether or not this has a substantial impact, making an actual difference or if their utilization of data is irrelevant and an invasion of privacy. These guidelines could accurately help us decide whether web scraping is ethical in regard to personal data and open access of information.