a. Throughout the 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 at Airbnb, I would implement a program that directly verifies the validity of a business license for short term rentals. The current system runs based on complaints made by the public, which means listers and Airbnb as a company is only held accountable when an issue is brought to their attention. This is an extremely unreliable way to ensure legal compliance, which is why a new system must be implemented within Airbnb. I propose that a program should be created to scan the policy number provided on a listing to ensure that it follows the correct format. This program should look similar to the function check_policy_numbers that was created in this project. However, this new method is not entirely sound. Other employees at Airbnb may complain that implementing this new program is a waste because if listers are aware of it, they can create fake policy numbers that match what the algorithm is looking for. Listers type these numbers in themselves, so it is very easy for them to match a fake policy number to one that looks real. Another argument that might come up against this new system is that it is invasive and shows distrust of Airbnb listers. Airbnb as a company is built around people renting out their own property, which is a business that involves immense mutual trust. If listers become aware of the new screening system, they may develop a bad image of the company and not want to list under them anymore.

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.

Throughout this project we collected many pieces of information regarding Airbnb rentals, so there are various questions that can be explored and answered using this data. One question I would be interested in understanding is "How affordable are Airbnb rentals in the San Francisco Bay Area based on the space they provide?" We collected data on the price of renting an Airbnb for one night as well as how many bedrooms it has. We can use these pieces of data to evaluate how much value people are really getting for the money they spend. We could take this question even further by looking at other rental services/properties and evaluating their prices and space provided in comparison to Airbnb. This will help us determine if using Airbnb is really the best option for those searching for affordable housing. I think this question is an important

one to explore because the context for this entire project was the affordable housing shortage that began in the 1990s in the San Francisco Bay Area. If we can pinpoint misconceptions in housing that is considered “affordable,” we will be able to move closer to solving this problem.

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 extremely important factor to consider when discussing the legality of web scraping is intention and more specifically when the intention is to promote social justice and equality. Individuals today highly value being treated as equals, and there have been major strides in recent years towards bettering racial, gender, and social equality. Web scraping can be used as a tool to contribute to these social justice movements. For example, web scraping allows people to see if certain job hiring algorithms prefer one gender or race over another. This is done by submitting two applications that have all the same information except for the gender of the candidate. Under some circumstances, this could be considered illegal because users are falsifying information. However, the intention behind this action was harmless, and the information learned from it is being used for social justice purposes. For this reason I believe intention is a vital factor to consider when discussing the legality of web scraping.

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 we must consider when deciding to use or not to use public data?

One guideline that should be considered is how much harm or good this information analysis will cause to the public and to individuals. Even if data is publicly available, this does not mean it will always have positive impacts once it is studied. For example, web scraping immigration data and putting it in an easily accessible format could create security risks for undocumented immigrants. This is just one example of how analyzing public data could harm other people, but there are countless other instances of public data that are more beneficial when left alone. Another guideline that must be considered is protecting anonymity. This goes hand in hand with minimizing harm because privacy protection is required in certain situations to prevent people from experiencing harm. I am working on a research project right now where my team and I are developing a mental health services app for the University of Michigan community. One of the first things we discussed in working on this project is how we can protect the identity of the people who use our app when we analyze their data. Mental health conditions are still highly stigmatized in certain contexts, so we want to protect the people who use our app to find resources. Even though accessing University of Michigan mental health resources is information that can become publicly available, it is not always in our best interest to use it.