6. Questions:

Once you've completed the coding portion of this assignment, answer the following questions using the following information. We know we want to keep airbnb accountable by checking if an airbnb does not have a policy number (a reference to the business license San Francisco requires to operate a short-term rental). Every entry in our database has a policy number, is pending a policy number, or is exempt from having one (hotels are exempt from this law). This is because airbnb requires listers to enter this information in a text box before allowing their listing to go live. However, looking through our database, there is a policy number that doesn't look like the other policy numbers. The listing id "16204265" has an unusual policy number. Using images of the exterior of the house posted on airbnb, we can pinpoint which apartment building this rental unit is located in, and check the San Francisco Planning Office to find out if this airbnb does not have a policy number and entered random numbers, or if the lister had a typo. Through this process we found that this lister does NOT have a short-term rental business license! This is an illegal rental unit that is taking a housing unit away from the local population. We can now file a complaint with the planning office to start an investigation! Note that the "Property Information Map" of the San Francisco Planning Office may not work on eduroam or MWireless. Turn in your answers to these questions as well as your code.

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.

> ➢ If I am working for the San Francisco Planning Office, I would develop a system that involves an API that would coordinate between Airbnb, the lister, and the government office. The San Francisco Planning Office would register all the listers who have obtained a policy number. Airbnb would then use the API in order to validate lister name, address, and policy number.
>
> One limitation of this system is that there could be a problem for a lister if their recorded information with the government is slightly different than the information they provide to Airbnb. For example, if a lister's address or name is written differently in the government system than in the Airbnb's system, the system might say that the lister doesn't exist.
>
> Another limitation of this system could be deciding who can get an API key because the government API key cannot be shared with everyone so that it could take some effort to verify a listing company like Airbnb.

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 the data we have collected, I can answer the question "are there adequate housing available for the local population?" The data collected from Airbnb's website shows that there are many listers who don't clearly list their policy numbers, such as "pending" and "exempt" and some listers enter invalid policy numbers. If there isn't a strict policy in place regulating how many units can be given for short term rentals, anyone can list their housing for short them rental, which will reduce the number of long-term rentals. This can result in rental housing scarcity. As a result, the local people can suffer greatly from lack of housing or they may have to live in poor quality homes because of high demand for housing.

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.

> An important factor to consider when discussing the legality of web scraping is whether the web scraping is done with the purpose of doing harm to the website owner or someone else or web scraping involves suspicious activities, such as intention to commit crimes and to keep track of someone's activity to cause them harm. If the web scraping involves criminal activities, then those people should be punished. Not everyone who uses web scraping has criminal intentions; therefore, it is important to distinguish between those with good intentions and bad intentions to treat them accordingly.

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?

> One important guideline to consider while considering to use web scraping is to read and understand the terms and conditions of the website.
>
> Another guideline to consider is to directly obtain permission from the owner of the website if the person or company can use web scraping on their website and explain the reason behind it.

The third guideline to consider is to research and understand the consequences of web scraping to help decide whether or not to use the publicly available data.