Questions:

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.

An alternative system for verifying the business licenses could involve having the city provide AirBnb with a list of policy numbers it has granted, perhaps by working with me (as a software developer) to link the databases between the two organizations. Then, when a potential housing lister types in a policy number, it would have to match a policy number in the database in order to be accepted. I'm envisioning a system that works similarly to how people enter credit card numbers online: The transaction only moves forward if the numbers entered match an existing, valid card number. This system would reduce the risk of listers typing in random numbers, in the hope that Airbnb will not properly validate. It would also reduce the amount of time/energy that AIrbnb has to invest in the validation process.

However, I can envision several arguments against this approach. If I was working at the SFPO, I envision that my colleagues might be concerned about sharing or linking a registration database with Airbnb. They might fear that Airbnb (as a private sector company) has a financial motive to exploit this information in some way. So they may believe that it is ethically wrong to share policy number information and potentially compromise data confidentiality in this way. In addition, my colleagues at the SFPO may simply not want to go through the extra work of doing the additional programming and effort needed to make this happen. If I was working at Airbnb, I can envision my colleagues not wanting to make changes that could make the listing process more difficult, as this could reduce the number of listings and damage the business.

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.**

If I were working with a housing activist organization, my primary interest and goal would be protecting the rights of low-income residents. I would want to help protect these people from being forced out of their homes and neighborhoods, so that the owners can rent their flats via Airbnb.

With that goal in mind, I could potentially explore the correlation between housing insecurity and rates of Airbnb listings. Specifically, I'd look to see if patterns emerge that suggest that when the number of Airbnb listings increase, it drives up housing insecurity (most likely by reducing affordability). This information could provide evidence and ammunition for me to argue/lobby in favor of limits on policy numbers or licensees within certain neighborhoods.

I could also analyze the data gathered through web scraping to research the quantity or intensity of Airbnb policy numbers and listings in different neighborhoods. I might look to see which neighborhoods have high concentrations and/or significant increases in Airbnb listings. I could then focus my lobbying efforts on setting limits on the number of policy numbers within that area. Another approach would be for my organization to lobby for establishing protections for residents, which would help them find new homes and/or make it more difficult and less compelling for absentee owners to rent out the units.

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 potential factor to consider is whether and when web scraping constitutes a form of stealing, in the sense that it is accessing and using data/information gathered by another entity. After all, an organization has most likely invested a great deal to gather

data, either to make money (in the case of a business) or as part of its mission or responsibility (in the case of a non-profit or government agency).

Web scraping involves another person or organization coming in afterwards and gathering and using that data for its own purposes. On the surface, this seems to be a form of stealing, since the second organization is taking something that it did create and that someone else invested in. However, perhaps this is acceptable if it amounts to re-using information in order to serve a broader public interest.

With that in mind, I think that how web scraping data is being used - whether it is for private benefit or for the public good - is an important issue to consider when discussing the legality of web scraping. One related question is who should decide or determine whether or not a particular web scraping effort is justified. Should it be the government?

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?**

Two potential guidelines to consider when deciding whether or not to use public data are:

- Whether the data gathered is for a legitimate public interest (rather than personal gain)
- Whether any personal data gathered is absolutely necessary to the project

As per my answer to question #3, I do think that why web scraping is being done (and the ultimate goal of the project) is one important criteria to consider. While it may be challenging to define exactly what constitutes a "legitimate public interest," this question strikes me as a guideline to consider.

But to act ethically, we need more than positive intentions and objectives. We also need to conduct our web scraping efforts in an honest and thoughtful way. This requires planning ahead to minimize any unintended negative effects and most likely, it means gathering as little personal data as absolutely necessary. So a second guideline to consider is whether or not we are gathering any non-essential personal data, which can be linked to an individual and potentially misused (against that person's best interests).