Priya Ganji
Project 2 Reflection Questions

Q1. *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.

Another system that could verify that the business license is valid for a short rental property is by having some sort of documentation upload feature. This will force the users to upload proof of the policy numbers. We can then use AI technology to scan the documentation provided to be approved for the listing to be active. This way, the chances to enter random digits or numbers for policy numbers will decrease

However, there could be a couple of arguments against this idea. For one, it will require a lot more software development or money to buy the software to cross-check every single proof provided. This may not be in the budget for my organization.

Another argument against this methodology could be the constant waves of checks that need to be performed but every new listing. This could act as a barrier for many individuals to add their listing to Airbnb as their listing may not be improved instantly, has the chance to be rejected, and overall another step in the process before earning money. As this would impact the profits made by my organization, they would likely be against the use of proof of documentation and AI checks.

Q2. The database we've created through web scraping is a great data source of information for data scientists 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.

After skimming through the Housing Insecurity in the US Wikipedia page, one specific question we can explore is, "does not have a bathtub or shower". If we changed

our code just a little bit to have it scrape data specific to the number of full bathrooms, instead of bedrooms, we could answer this question. This information could be very useful as the housing activist organization can predict the safety and quality of a house to see how many people it can house comfortably. This can not only help fight against housing insecurity but also during natural disasters as well. This code can help quickly generate suitable places for people to live during emergencies, for children in foster care, help homeless shelters, etc. The possibilities are endless!

Q3. 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 factor I believe is extremely crucial in discussing the legality of web scraping is the freedom of speech. Reading through the article regarding legal issues with the Computer Fraud and Abuse Act, I realized there is a fine line between freedom of speech, and causing harm by using freedom of speech. We must consider the effects web scraping has and if it falls under our first amendment rights.

It is also important to keep in mind if something is published on the internet for the general public to view, the information present could likely be retrieved in some way or another. Hence, bringing up the point about what information we want to put on the web itself.

Q4. 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 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 guideline when web scraping is asking ourselves the question of, what are the implications and possible uses of this data? This will allow us to see if the information we are pulling out could have negative consequences or the chance for immoral/unethical activity. If the answers to these questions are looking positive, then I believe it is alright to move forward with web scraping, with consent.

Another guideline to follow regarding personal data could be to ask the question of whether I would want my data of whatever I am web-scraping to be played around with and looked at by others. Asking myself this question will push me to think in the

shoes of the person's data, helping me understand the importance and intensity of the situation at hand. Sometimes this could be a big thing, and sometimes it could be information we wouldn't mind others having.

Overall, I am sure there are many guidelines to consider. Still, I would focus on the implications of the work we are doing to see if it is ethical and morally valid, as well as think about how we would feel if our own personal data were used in such a way.