SI 206: Project 2 Questions
Imani Ahmad

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.*

- A different system that could be used to verify that the business license is valid for short term rentals in San Francisco is if Airbnb were automatically required to provide a running list of all the short-term rental properties' business license numbers for the San Francisco area so that the SFPO could cross-reference the list with their existing database and identify any discrepancies. They would then be able to follow up with the rental owners immediately rather than waiting for someone to notice and a complaint to be filed. Two arguments that Airbnb would have against adopting this system are that firstly, it would discourage a lot of renters from operating in the San Francisco area because they wish to avoid dealing with so much paperwork and policy, which would cause Airbnb to lose business. Secondly, Airbnb would probably argue that it takes far too much time to register for a business license, meaning fewer available rentals on their website at a given time, which would also cost them a lot of money.

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 a data scientist working with a housing activist organization to fight against housing insecurity, some research questions I could answer or explore using this data are *"To what extent does the existence of short-term rental properties impact housing insecurity?", "How many short-term rental properties on average are operating illegally (without a permit)?"* and *"Is there any correlation or overlap between areas with high concentrations of housing inequality and areas with high concentrations of short-term rental properties?"*

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.*

- Some factors I believe are important to consider when discussing the legality of web scraping are whether or not the information being scraped is someone's opinion or original expression or if it is simply general information or fact. The reason I think this distinction is important is because I do not think people should be able to hold claim to or have ownership of facts, but rather there should be open access of information and they should be available for everyone to access and utilize as they wish. Another factor I believe is important to consider is the intention behind web scraping. This is extremely important because without the context of intention, it is possible that many important and positive initiatives, whether it be research, journalism, or other intents, are being criminalized when they are not meant to cause harm or exploit a system, but rather, are trying to make a positive statement or incite positive societal change through the data collected.

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?*

- Some guidelines that we must consider when deciding to use or not to use public data are to avoid scraping people's data that, when utilized or publicized in conjunction with or when cross-referenced with other data, could expose more about an individual than they initially intended to reveal, and to use public data only when scraping said data will not pose a threat to a site's functions or performance and is selectively scraped, not at such a large volume that it appears to be some sort of attack to the site.