Abhinav Tadikonda
SI 206
Project 2 Reflection

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.

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 was a software developer at SFPO I would implement a system that cross references the address of Airbnb listings with the SFPO database to verify that each unit listed does

in fact have a valid policy number and is therefore allowed to do short-term rentals. This would lead to more accurate and up-to-date information on if a property has a valid business license.

One argument that my organization, SFPO, might have for adopting this new system would be an issue with privacy. Since precise locations of Airbnb's are not public knowledge, SFPO may need to come to an agreement with Airbnb to get access to such information. This access to Airbnb addresses could make SFPO a target and expose them to potential litigation if their data is hacked. Additionally, since Airbnb needs to display some effort in validating these policy numbers, with this new system in place it becomes far more unclear who is accountable for invalid policy numbers.

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 was a data scientist working with a housing activist organization to fight against housing insecurity, one issue I would explore would be how housing insecurity affects people of color. By looking at factors from the *Housing Insecurity in the US* Wikipedia page such as health and education to name a few, I would analyze how people of color are affected more by where they can live compared to white individuals.

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.**

I think the concept of trespass to chattel is an important legal claim to prevent undesired web scraping. Trespass to chattel is civil wrong which involves intentionally interfering with one's movable personal property, such as a laptop. Since the owner of the device does not consent to having their data interfered with, this is a breach of privacy. If trespass to chattel was made legal, no one would be able to have any private information which could pose serious security threats for common individuals and bigger organizations alike.

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 important guidelines to consider when deciding to use or not to use public data is intent and consent. If the information being obtained is meant to benefit the public or is necessary for public safety, then accessing someone's personal data without their consent could be seen as ethical. For example, in order to stop a public attack or perhaps slow an illness outbreak, using someone's personal data could help the greater good. Additionally, it is important to consider whether the individual gave their consent for their data to be used. In a study done in 2019, IBM used images of individuals found on Flickr for a case study. While the images were publicly available, many Flickr users were upset that their pictures were used without their consent. As such IBM received a lot of backlash. Therefore, consent and intent are important guidelines to consider when deciding whether to use or not to use public data.