Mia Friedman
SI 206, Project 2
Ericson
November 14, 2022

Project 2 Reflection Questions

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 were at the San Francisco Planning Office (SFPO), I would propose a different system for verifying licenses instead of working on a case-to-case complaint basis. I would propose that the planning office issue an ID card with a uniquely generated barcode for each lister when they receive their business license. This way, they have an actual ID card to submit to Airbnb.com when they apply to start listing short-term rentals. In addition, instead of just having Airbnb.com verify that these business licenses are real, SFPO should have software that compares the barcode of the license in their system with the one submitted, automatically pointing out discrepancies rather than somehow expecting normal people to catch them. With the current system, there is a major lack of accountability because in order for somebody to register a complaint of an illegal rental, they have to understand the intricacies and patterns of business licenses and notice a difference while simply scrolling through the source code of the website. This is not only grueling but unrealistic for a normal person(and web scraping the actual website as a coder is illegal!). SPFO having code in place to automatically compare two barcodes will spot most if not all discrepancies and illegal licenses, but do it without requiring outside help of the community, which it shouldn't have to depend on. One possible argument SFPO might have with this plan is that issuing ID cards with barcodes to every lister is an added expense that SFPO probably doesn't have in their budget. It requires additional thought in the materials, manufacturing, and distribution of these cards which may also require extra labor. Another possible argument people at SFPO may have is in regard to privacy concerns. For SFPO to also be verifying Airbnb.com applications to list, they would need a certain amount of legal access to

other personal information about listers and information about Airbnb.com. Getting the proper agreements and contracts with Airbnb.com is not guaranteed, and may be difficult. In addition, listers will need to consent to this level of information going to SFPO, which some listers might not be comfortable with. As a software developer, I still think these considerations can be thought out and are well worth having a proper system in place to more accurately verify and catch illegal listers, but they are considerations nonetheless.

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.**

One possible research question that the database we created can help answer is "what percentage of Mission District[or insert any other area of study] short-term rentals are listed illegally?" With our database, we get a list of the illegally listed listing IDs on Airbnb's website for Mission District listings. However, this data can be applied to other areas as well. We can compare the amount of illegally listed rentals with the total rentals in the area to see what proportion or percentage of listings are listed illegally. Having this knowledge can benefit data scientists working to fight housing insecurity in multiple ways. Firstly, knowing how prevalent illegally listed units are in certain areas can help activists expose the issue to policymakers. Since we know what units are illegally listed, activists can work to get those listings off Airbnb and directed toward people who are actually in need of stable, more long-term rental units in the area. To further stop the problem from continuing to occur in the future, policymakers can use the database as a gauge to see if the problem is persisting/how bad it is over time and if additional legal action is necessary to prevent the prevalence of illegal short-term listings. Another extremely important reason it's helpful for somebody working against housing insecurity to know the prevalence of illegal listings is for safety reasons. The Housing Insecurity in the US Wikipedia page classifies "housing insecure and the house inadequate based on these structural conditions:" some but not all including "has exposed wiring, not every room has working electrical plugs, and the fuses have blown more than twice, has holes in the floor, has open cracks wider than a dime," etc. Regulatory agencies have a harder time keeping track of illegally listed units, which have the potential to be major safety hazards for renters. Knowing how many illegal listings there are, which units are illegally listed, and how to get them taken down until they have a license could potentially save renters, both short-term and potentially long-term once the units are taken down from short-term rental sites to be offered as more affordable long-term rentals.

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 an important factor in the discussion of the legality of web scraping is the scraper's intent and use of the information from the site. I found the Electronic Frontier Foundation article interesting because the scraper wasn't using the information to hack the system and benefit themselves, but gain the information to analyze and better the system. In the case from the Legal Issues section of Web Scraping in the US on Wikipedia page about Ebay and placing automatic bids to take listings off the page, the scraper is using their access to the platform with an intent to benefit themselves and get preferential treatment over other users. The goal didn't ultimately benefit anyone other than the person scraping, and interfered with the site's original intended use. In the case of Sandvig V. Barr, the scrapers were gaining information to expose discriminatory algorithms based on certain characteristics like race or gender. The ultimate goal is to create platforms and algorithms that work more accurately and fairly for all individuals in all groups. In this case, violating a site's terms and conditions is warranted in order to gain the information necessary to successfully complete this research. If every study needed permission in order to web scrape for information like this, sites would have more preference over news regarding their company and fair expository journalism would no longer be objective. The truth would be easier to hide as sites could simply deny access to information that web scraping may allow if they think the information may negatively impact their site. So, in the case of journalism and bettering the community with the scraping it seems more reasonable than in the case on the Wikipedia page with Ebay. However, there are many nuances to these scenarios that make it difficult to decide exactly when to draw the line with legality. It may be difficult to assess or justify exactly what journalism has positive social motivations and what doesn't, and whether there are still privacy or territorial limits that can't be crossed even in the case of positive expository journalism. While these questions and situations are difficult to answer, I think they are important to discuss and consider when considering web scraping like in cases of our second project.

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?**

    I think the first general guideline that we should follow when deciding to use or not to use public data is that we shouldn't use any personal information from users of the site that isn't publicly available on the site. While we can use other data from the site that isn't publicly displayed, a user of the site's privacy shouldn't be at risk so a scraper can accomplish another agenda. Names and general information that are already publicly

displayed should be allowed to be used since anyone can see them, but using scraping to take other information from users becomes a major privacy concern that shouldn't be necessary in order to accomplish a completely different goal. Another guideline that we should follow in our decision of whether or not to web scrape is that we should only web scrape if we think the information we will gain from it can benefit somebody or society outside of ourselves as scrapers. For example, in the Ebay case scrapers hacked the system so they could win and get products for low prices on the bidding platform of the site. This only benefits the scraper, and actually negatively impacts other users and listers. In the case of our database creation with Airbnb, we are helping SFPO correctly enforce their regulations, which will help mitigate housing insecurity. While we do web scrape, we are doing it in the positive interest of justice and housing stability, which will positively impact many people outside of ourselves (it actually doesn't even positively affect us much!).