Dylan Siewert

**Question 6 Responses**

* 1. **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.**

In the new system, renters will have four options to submit their business license number (BLN): input their BLN using the format “20##-00####STR,” input their BLN using the format “STR-000####,” claim their BLN is “Pending,” or claim their BLN is “Exempt.” If there is a different format for the BLN in the renter’s city, these options would change accordingly. The process would be split up into more steps for a homeowner to be able to rent their home, adding a “request to register” form before actually registering. Next, a request would be sent to the San Francisco Planning Office to verify that their BLN is valid (or that their BLN is pending, or they are exempt). One argument against this new system is that it would take a long time to verify each renter’s BLN. A second argument could be that renters might be discouraged from listing their property because of the lengthy registration process.

* 1. **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**](https://en.wikipedia.org/wiki/Housing_insecurity_in_the_United_States) **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.**

By web scraping the data on this website, a data scientist could explore the relation between house price in a state versus the percent of owned and rented households facing housing insecurity. By exploring this question, a data scientist could figure out whether house price by state is an indicator of high housing insecurity is. The data about rented households facing housing insecurity could also be compared to the listed prices of Airbnb’s in the same area.

* 1. **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**](https://en.wikipedia.org/wiki/Web_scraping#United_States) **and** [**this article about the legal issues with the Computer Fraud and Abuse Act**](https://www.eff.org/deeplinks/2020/04/federal-judge-rules-it-not-crime-violate-websites-terms-service)**, and describe at least one factor you believe is important to consider when discussing the legality of web scraping and why.**

One factor to consider is research when it comes to web scraping. It’s important that researchers are protected. Take the second article, where researchers were accused of fraud for collecting data about race and discrimination. Because researchers have no monetary or immoral goals when it comes to web scraping, they should be granted access to this data and protection from the law.

* 1. **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?**

One guideline to consider is the goal of collecting this data. Will it benefit the people whose data is being collected? Will it benefit the general public? If not—if the goal is instead to gain an industry advantage or gain a profit, the idea of web scraping is unethical. Another guideline to consider is how accessible the user has made the data. If a post has been archived, deleted, or hidden behind a private account, the user does not want that data to be accessed. If the user has hidden it from the public, I think that data should not be scrapped.