Lauren Pyfer & Devon Vitale

Project 2 Questions

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.

The system will first verify if the home is valid for a short-term rental. The system would determine if the lease is considered short term which means 30 days or less. It would also figure out if the home is eligible to be rented, such as if it is considered, in the state of California, short-term rental property. The system would then factor in if the property is eligible for rentals by its zoning rights. Once all these criteria are fulfilled, the system would then verify the license of the home-owner. The system would see if you have a general business or short-term rental license. Once this is checked, the house will be available for rental.

Arguments I may hear from adopting this system could be the constant change of zoning laws, how will the system update so quickly if a zone is changed? Or what if there are certain permits granted to houses within a restricted rental area?

Another argument against adopting the system could be how the system will react to a person who may have a permit revoked or expired? Does the system automatically update these problems? Or must it be manually plugged in to fix these issues?

2. **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 research question that could be answered would be about neighborhood safety. Are the prices of housing affected by the location? Does a house in a defined imminent health safety location lower than a normal house? If so, how much is it marked down? What determines price differentiation?

3. 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 that is important to consider when discussing the legality of web scraping is the crucial reasons a website's information should be accessible to the public and government. Although web scraping has the potential to be used unethically, when the scraped data isn't used for harmful purposes it can provide information required for analysis. Since publicly available sites can not require a user to agree to Terms of Service before they access the data, those sites are obligated to allow free collection of data from the site. So, in order to keep web scraping ethical while legal, the user who posted the data had to have made it public and no login access should be required.

4. 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?

When considering when to use or not use public data, one should follow certain thoughtful guidelines. There is no perfect quantitative system for calculating how classified a particular data element is, so the scraper must be aware of the possible effects. The first guideline is to weigh the confidentiality level of the information. Although the accessed data is publicly available, different potential impact levels using it could have on organizational operations, assets, or individuals. If the impact is low, meaning it would have a limited adverse effect on said parties, then the web scraper can make a moral decision on what to do with the information. Moderate impacts mean that the disclosure of information could have serious effects, while a high potential impact on confidentiality would have catastrophic negative effects on organizations or individuals. The measurement of web scraping and its potential adverse effects on an organization can also be conducted based on an integrity and availability scale. The next guideline is to collect personally identifiable data only when it is absolutely necessary. The purposes of data collection should be specified prior to or at the time of data collection. By

making your intentions clear by following ethical procedures, you are more likely to make thoughtful choices when using public data.