

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.

A system that could verify the business license validity is to create a validator that would validate your status, similar to UniDays, which validates university student status to give people discounts on a lot of different websites. This system could not be as accepted by Airbnb.com because while UniDays would connect you to your school's email and verify you in that manner (no data holding required, other than the name of the schools), this system would need to hold all of the data from the city and verify it, which could lead to some problems in not only getting all of the data of the city (permits and the scope and logistics of getting the data) as well as the concern for a leak in data. It would be very easy for this information to get leaked. If someone were to obtain data on all of the short-term rental registrations in San Francisco, it could lead to a lot of security issues. Also, because the government is not running this validator, would it be plausible to obtain all the data in a timely manner? Would the government give all of this data to a private company?

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.

A research question I would pose would be to determine the price of rent and compare that with the price of using airbnbs. A lot of people are not able to pay a whole month of rent, which leads them to find more temporary housing solutions, such as airbnb. However, the issue with this is that short term housing solutions are very pricey. By looking into this and comparing the data we have collected and comparing it with rent prices, airbnb can better regulate these prices to accommodate for people and help fight against housing insecurity.

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 implicit bias. While violation of the computer fraud and abuse act is considered illegal, it is also important to point out the issue of implicit bias in certain websites. If people create accounts with identities that associate them to being a minority, it can show an algorithm's implicit bias. People with ethnic backgrounds have been shown to be selected less often for a job application on a certain website. While some forms of web scraping is illegal, if it is found out that their algorithm is doing something inherently racist or discriminatory in any manner, it should be able to be found as an issue and expose these problems. It is a question of company privacy and present biases in their product, which can create a gray area for if web scraping should be allowed or not.

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?

There are some guidelines to consider when deciding to use or not to use public data. The best way to do this would be to make sure that when you are scraping data, that the data being used can have a form of being anonymous. If you are taking some things from the website that include names or private information, make sure to avoid these altogether or to change them to different things so people's data can be protected. Also, you need to think about your intentions for web scraping. If you are doing this for research or price monitoring, it would be fine to do so. However, if this is going to impact the company or their stakeholders, it is unethical as you are infringing on people's data.