

SI 206 : Project 2 Reflection

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. A different system that can be recommended is separating Airbnb listings within a given area by the number of rooms being provided and then providing data on how many people stayed within an Airbnb based on the size of the listing. To support the system idea that was stated, separating it by region or individual states is another great way to support the system.
 - i. One argument I might hear is that adopting this system may not be productive at all when looking at information regarding short-term rentals. This system might be more helpful in the long-term.
 - ii. The second argument being shared is that it could be irrelevant to compare data by region or individual states because every state has a different amount of listings for very different reasons (i.e. personal travel, work, weather, tourist sight seeing infrastructure).
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. I think a few things I would explore overall using this data if I were a data scientist working with a housing organization are these factors: race, regions, household wage. I think I would study these three factors the closest because they say a lot about housing insecurity overall in the United States. As we can see from the data provided, certain regions suffer from housing security more due to the cost of living within that region and what jobs people currently hold. I think race is important because a lot of people suffer from housing insecurity because a lot of people face homeowner discrimination, racial segregation and racist policies within their given region.

Source:

<https://www.brookings.edu/essay/homeownership-racial-segregation-and-policies-for-racial-wealth-equity/>

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.
 - a. When reading the second article regarding the Computer Fraud and Abuse Act one of the first things we learn is that this is a **vague** crime law, although it is a law. I think this is important to consider because I personally believe when a law is set and being implemented whether the terms of the law are vague or not it still needs to be addressed accordingly. I think creating a **“vague law”** such as this one was to visually appeal to users on the surface level to make it seem like users are protected when in reality they are not but to also support technology products or technology companies that create a substantial amount of income and data for the federal government. I don't think this law was created at all in support of the everyday user.

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?
 - a. I believe two or more guidelines that can be considered when deciding whether or not to use public data is to consider terms of agreement and if users have a passcode that is attached to their account or profile in relation to whatever web application that is being used. It is important to make sure the rules of the terms of agreement are not broken because it is essentially a contract created that users sign before using an application. The importance of having a passcode or a pin number is because both are used for security and privacy purposes. Users are under the impression that if they create a passcode or pin number for their account that their data will not be shared without their knowledge or consent but to also protect them for everyday hackers.