

# SI 206 - Project 2 Questions

- a. As a software developer at Airbnb.com, adopting a system that verifies that the business license is valid for short term rentals in San Francisco could be met with some pushback. One argument that might be heard in this situation is that implementing this system could decrease business. Without this system, it is likely that Airbnb.com allows hosts with invalid business licenses, which undoubtedly contributes to Airbnb's business. Upon implementation of this system, hosts who fit this category would be unable to offer rentals and thus would no longer contribute to Airbnb's business. Therefore, an argument against implementing this system is that it reduces the number of hosts and thus the revenue of Airbnb.com. A second argument against this is that upon implementation of this system in the San Francisco area, other areas may move to require Airbnb to implement this system. While the loss of business generated by this system's implementation in the San Francisco area may not impact Airbnb too greatly, it is important to consider that hosts with invalid business licenses likely exist outside of San Francisco. Thus, creating this system with the San Francisco area in mind may create a waterfall effect where subsequent cities require Airbnb to uphold this system. This would likely have a large impact on Airbnb's business, and therefore pushback against this system being implemented is likely as a software developer at Airbnb.com.
- b. If I were a data scientist working with a housing activist organization to fight against housing insecurity, one research question I could explore using this data is whether housing insecurity has increased or decreased over time and by how much. I would like to investigate this question because although various graphs on the page show a steady increase, these might fail to consider outside factors. For example, the "1 Bedroom Rent By State" shows an increase over time, but this may coincide with the state of the economy and the average income of residents of the state. Thus, I think that while some of this data may appear alarming, further research into the actual conditions of the data could provide insight as to how serious the housing insecurity crisis is. From there, correct actions can be taken to maintain or change the outcome of the study.
- c. One factor that is important to consider when discussing the legality of web scraping is the type of access made by the scraper. This is important because the access of the scraper determines how much information they have accessed and potentially copied. If the data accessed by the scraper adversely affects the site owner's system, this should be met with consequences for the scraper. Therefore, when discussing the legality of web scraping, what should initially be considered is the level at which scrapers may access information contained on a website. By considering this factor, we can provide the scraper with the ability to retrieve data, a degree of privacy to those whose information is on the site, and security of the site owner because the scraping being done will not compromise their system.
- d. One guideline that should be considered when deciding whether to use public data gathered from web scraping is that using an API allows you to avoid the complications of conducting web scraping. Typically, websites will have APIs that let you gather data from them without scraping and potentially breaking their rules. Thus, a good guideline when seeking to use public data is to first attempt to gather this data using an API, as what you

gather from the API you will be fully authorized to use. A second guideline for web scraping and using public data is to check the copyright and data privacy legislation of the site you are scraping from. For better or for worse, publicly-available data can typically be scraped without consequences. But, to be absolutely certain, you should still refer to legal documentation to make sure that any scraping you are doing is allowed. Thus, when scraping, first check if an API can collect the data that you need, and if this is not the case, you should refer to the documentation of the website you intend to scrape in order to identify that what you are doing is legal and applies to existing laws and documentation.