

1. We could develop a database that keeps track of all short term rentals in SF. This database would have to be regularly updated and would basically avoid a situation where people would have to do web-scraping to validate the policy numbers. The database could be an automated checker that validates the numbers and if it isn't valid, then the system can send a notification to the SFPO or Airbnb notifying them. It could also be manually controlled and kept track of, but this would prevent due diligence work from having to be done all of a sudden. Rather, there would be someone responsible for the validation and verification as their main role. One argument I might hear against this system is that it could be expensive and time consuming to maintain the database and keep it up to date. Another argument would be how easy it is for there to be errors especially if it is maintained and created by only one person/AI. If we use AI, there are other potential issues like issues within the code leading to a serious malfunction and approval of invalid/illegal rentals.
2. If I were a data scientist working with a housing activist organization to fight against housing insecurity, one question that I would want to explore using this data is "do education related impacts of housing insecurity vary across each state?". It looks like there are cases of housing insecurity in all 50 states, but I would be curious to know how differently it affects students in different states. Do students do better in schools if they are in a state that tends to prioritize education more? Are students more prone to succeed if their state has a mix of secure and insecure housing – is it better if the student is surrounded by counties/areas with affluent housing? From the article, housing insecurity takes a hit on students' educational experiences, but I would like to explore that in more detail and see if the level differs based on the surrounding neighborhoods and the U.S. state that they live in.
3. One factor I believe is important to consider when discussing the legality of web scraping is the terms of use and publicness of the website because websites cannot expect that their data will not be taken and used when they are completely public. Unless they have terms of use explicitly stated and copyrighted their content, user data and publicly available information are likely going to be used and scraped. In my opinion, companies should not file lawsuits complaining their data has been used when they have not taken the necessary steps to protect it via copyright or making a private website. The article states that "For several reasons, the court ruled, violations of a website's terms of service cannot be grounds for criminal liability under the CFAA" so not even terms and conditions make a website completely safe. I do think terms of use is a fair factor to be considered when thinking about the legality of web scraping, though. If a website explicitly states conditions, it can be argued that they did specify it in the conditions and expectations are that the data won't be extracted. It is quite arbitrary though and terms of use can be hard for consumers to understand. Legal courts and lawyers really have to set a precedent so companies are more aware and more likely to protect themselves in a way they know will be more likely to be safe, legally.

4. One guideline we must consider is whether or not the data is personal in nature. Though it might be public, it may not be intended to be public or used for data collection and analysis. Anything with people's face or personal information should be avoided, in my opinion. Another guideline to consider is whether or not the use of this public data will actually be beneficial. If doing this web scraping does in fact lead to positive results and accountability, I think it may be worth doing.