

- A. A different solution to verify the business license validity would be to create an AI screening through Airbnb that would cross reference the city planning office's database, and only push the Airbnb onto the website if it passed all the screening criteria. The validity would be verified by having access to the SFPO database of business licenses and there would be an alert to the developers for the listers that are not verifiable. An argument we could hear from Airbnb.com is that the cross reference and screening could take up a lot of time, and screening millions of listings would cause less listings for customers to book (less traffic on site). Another argument would be that having access to the SFPO database means that SFPO would have access to our database, potentially allowing for a greater chance of leaks and misuse.
- B. A research question we can explore as a data scientist with this data is the relationship between the house inadequacy and insecurity in correlation to the amount of illegal short-term rentals and falsified licenses. We could see how an invalid listing being available affects the area it is listed in within the context of housing insecurity.
- C. For me, the biggest factor when it comes to the legality of web scraping is finding the balance between a company's privacy vs a company's transparency. If a company chooses to hide data that should be available to the public behind their terms of service and policies, they should be held accountable. The problem is that web scraping can be used for that hidden data to be brought to light, but it becomes a battle between violating a company's privacy vs a company hiding things that the public should know. The most important thing is to honor the company's policies/privacy, and scrape the data without trying to crash one's site or uncover hidden details. It is up to the court's to crack down on a company's accountability rather than those who web scrape as they can put themselves in a dangerous position, like Andrew Auernheimer who was convicted of hacking because of web scraping (Roberts, Sep 17 2018).
- D. One guideline that must be used when deciding whether or not to use personal data without consent is the company's terms of service and the conditions they have on their page. Whether this is copyright infringement, trespass to chattel, or violating the Computer Fraud and Abuse Act, web scraping should heavily consider this guideline to avoid facing repercussions. Another guideline is not using the data in a negative fashion, but rather the way it was presented in the first place. Data is one thing, but interpretation is another. There needs to be context given behind data if it is pulled and brought into the limelight, rather than it being leaked and left to other's interpretations. This can lead to a lawsuit that most would want to avoid, and this requires an active consideration of what a company does not want to be portrayed as through their data.