

6. Questions

- a. A different system that can verify that the business license is valid for short term rentals in San Francisco is by creating a page that runs through the policy number and checks to see if the policy number matches with all the other information within the San Francisco Planning Office database. In addition to that, we would try to match the images on Airbnb of the exterior of the house to images of the exterior images on Google Maps to determine whether it is the correct house being posted. If the policy number leads to no information being found on the database it means that the business license is invalid, if the policy number leads to information that does not match the data on the Airbnb website it means that there could potentially be fraud. I think one argument that I may hear from my organization SFPO is that they believe that it will be less costly and more efficient to match the numbers and determine if they match the policy number format; however I believe that there could also be a possibility that even if the policy number matches the correct format the policy number could potentially be invalid as well. Another argument that I might hear at the organization is that attempting to match images of the exterior of the house on the Airbnb website to that of Google Maps may be difficult as building images on Google Maps are updated every few years. Therefore, the images on Google Maps may not match the images on Airbnb even if the apartment building has a valid business license.
- b. After skimming through the Housing Insecurity in the US Wikipedia page, some research questions that I could explore using this data if I were a data scientist working with a housing activist organization to fight against housing insecurity is “Are urban areas/city areas/more populated areas more likely to be fighting against housing insecurity as prices for rent in those cities tends to be higher?” or “Is there a correlation between ethnicity and housing insecurity and would this be due to historical or economic disparities?”
- c. One important factor in determining the legality of web scraping is the “Computer Fraud and Abuse Act (CFAA)” as well as “trespass to chattels.” I think it is important to address that the CFAA criminalizes the violation of websites’ terms of agreements as well as unauthorized access. And so, this is an important factor to consider because there are situations when unauthorized scraping of data and content is occurring and the user of the web scraper is exceeding their authorized access. This is difficult in determining the legality of web scraping because it also poses a potential threat to the first amendment by criminalizing protected speech. In addition, I think it’s important to consider “trespass to chattel” such that the user of a scraper may be trespassing data by accessing and collecting information or interfering with other people’s property. This brings into question whether there was intention, when conducting web scraping, to interfere with a company’s computer system and the amount of damage the interference has caused.
- d. One guideline that we must consider when deciding whether or not to use public data is the purpose and intention for grabbing data from websites. In terms of purpose specifically, some questions that should be answered include whether the purpose is for research, data monitoring, and improving services or products for a positive impact on society or revealing critical information that the public should be aware of. Overall, I

believe that we should use public data and web scraping when it deals with supporting public oversight over legal and illegal acts and reducing corruption, ultimately promoting more transparency. Another guideline that I would consider when deciding whether or not to use public data is if the data is non-discriminatory or license-free data so that it does not involve any legal consequences so you are free from copyright, patent, tradeseecret, trademark infringements. If there are no restrictions that are stated, I think it verifies that this public data can be used for any purpose.