

Samantha Lawrence
206 Project 2
November 10, 2022

1. A different system that can verify the validity of the business license for short term rentals in San Francisco is one that will only accept valid formatted answers. For example, if the business owner types in something that does not fit a valid format, it will not go through the database or receive a license and will be illegal to run. In the example, someone input an incorrect policy number, and it was able to run undetected for a while. With this new system, property owners will only be able to get a license if their inputs are correct. One argument against adopting this system is that there is nothing to stop these people from inputting incorrect information that is in a valid format. Another argument is that it would then make it harder to find properties that are operating illegally. While this may be true, the system could also have the ability to notify the SFPO if there is a duplicate or unregistered license number entered into the system.
2. One research question that I could answer or explore using this data to fight against housing insecurity is, "What is the rate of change for the number of short term rentals available now vs previous years?"
3. One factor that should be considered when discussing the legality of web scraping is general interest if given the ability to access harmful or sensitive information. This is because you don't want information about your customer's address, bank, or phone number to be exposed. Therefore, this type of web scraping should be illegal and penalized because it harms the general public. However, web scraping flight prices should be allowed because it is not causing harm to anyone else.
4. The first guideline we must consider when deciding to use public data is double-checking to ensure that this information cannot be linked directly to a person. For example, if the public accessed our information, I would not want anyone to be able to pinpoint which data points were theirs and link them back to themselves. The second guideline we should consider after checking that this data isn't invasive is if this piece of information we are scraping will be critical to the research we are doing. Even if the public data we are scraping is harmless, we should be thoughtful about what we are using because it still is not our data, and we should only scrape it if necessary.