Mari Jaoshvili
Project 2 Reflection
11-8-2022

    a. Another way to verify that a listing's business license is valid for short term rentals in San Francisco is by using San Francisco's database of licenses. Right now, in order to check the policy number's validity, we simply check if the format of the number matches the correct format. However, it is possible that people could make fake numbers using the correct format, so this may not work how we want it to. So, it would be more accurate to actually look through the license data. We can use regex on each of the policy numbers we found to look through the license data and make sure the policy number matches a value in the license data. An argument we might hear against this system is that getting permission to use this data from San Francisco may be *very* difficult, if possible at all. This type of data is often protected and not given out to the public, but I believe it is possible that the city will want to join with AirBnb to make sure housing units aren't taken away from the local population because of illegal rental units. Another argument we may hear against this system is regarding its lack of efficiency. It will most likely take a really long time to cross-check every single policy number against the license data, which is a major downside of this system.

    b. A research question regarding housing insecurity that can be answered from our database is: What proportion of the available housing in San Francisco is being occupied by AirBnb? Because we have the get_detailed_listing_database() function, we are able to use the data returned from this function to see how many AirBnbs there are in this area. We can then find the proportion of AirBnbs in this area by dividing this number by the total number of housing units in the area. We are then able to use this information to determine whether or not too much of the available housing in San Francisco is being occupied by AirBnb, and the city can act accordingly.

    c. When discussing the legality of web scraping, it is important to consider the privacy of a website's users. Just because someone's information may be out to the public, that does not mean that just any person should be able to access this information and manipulate it. For example, I do not think it should be legal for someone to web scrape Facebook and get a ton of personal information on its users. However, if this were to be legal, I then believe there needs to be some sort of transparency law. This law would force businesses and websites to inform its users during the sign-up process that there is a possibility that their data could be extracted from the site from third-party webscapers, and that there is nothing the company can do about it. This way, it is up to the user to decide whether they are comfortable with this.

    d. I think it is important to consider the ethical implications of why someone might be web scraping. If someone is web scraping to help the public in some way, or improve something in society, then I believe it would be appropriate to web scrape. However, if someone is web scraping simply for fun and to get personal information from random, innocent people, then they should not be web scraping. So, some questions someone can ask themselves before using public data are "Am I helping anyone by doing this?" and "Would I be comfortable with a stranger using *my* data like this?" While the answers

to these questions are subjective, it still prompts the person to think about the ethical implications and the consequences of their web scraping.