Project 2 - SI 206

Alana Vang and Jessica Cho

alanav@umich.edu   jesscho@umich.edu

Question A

The issue with the current system is that airbnb is not validating policy numbers when a listing is posted. The validation of policy numbers should be automatic and if the number is not valid, the listing should not be able to go live on airbnb.com. A system that would verify that the business license is valid for short term rentals would have to sift through the SFPO valid policy numbers through the airbnb website when the user wants to post their property/unit up on the website to make sure the license is valid. This would need SFPO and airbnb to be in contact with one another and share data with one another. This new system would quickly eliminate illegal listings from people that don't have a license to short term rent their property. One complaint that might occur from both organizations is that there might be a privacy issue with allowing the organization's data to be accessed by another organization. Although the new system would automatically eliminate invalid policy numbers, the SFPO would need to have their system with the policy number data updated so the user can list their unit right away. If the SFPO system is not updated, airbnb listers may complain that their policy number is not read as valid when it actually is.

Question B

One research question that we can explore using this data about housing insecurity in the US is why certain states have a higher percentage than others, and what factors play into those rates. With this question, we can explore the relationship between short term airbnb rentals, and rates of housing insecurity within a single region. Then, we can dig deeper into the causes and circumstances of a region's housing crisis. Rather than focusing on Airbnb's actions in this context, we could further expand our understanding of the implications of short-term rental units and use that knowledge to fight the crisis.

Question C

When discussing the legality of web scraping, it's important to consider first and foremost the privacy of the individual's personal information. As a consumer, it's hard to sympathize with a company's hesitance towards allowing "scrapers" to access their *public* information. In my opinion, as long as the consumer's privacy is protected, scrapers can scrape away. As demonstrated in this project, web scraping can be a useful tool, especially when it comes to holding enterprises accountable. In this case, we aimed to hold AirBnb accountable for adherence to the local policy established in the San Francisco region. We utilized purely publicly-accessible information, and I see no reason why this would be a concern for AirBnb. Again, it's hard to sympathize with a company who's already utilizing their user base data to profit.

Question D

Some other guidelines to consider when deciding whether or not to use public data is the implication of using said data, and the user's consent to such data being shared. First, it's important to consider what the application of scraped data would look like. Is it used harmfully, like in doxxing? Or is it being used as a means of gentle, non-invasive accountability. Is the data being shared with a large number of people or users? Or is it being used as a single discussion point within a larger text? Secondly, we must consider what data users have *willingly* made public, and whether or not their privacy would be at risk if the data were to be shared. In relation to doxxing, people don't often opt-in to displaying their information on WhitePages, yet it's automatically shared. A web scraper could use this *legal* information to expose the home of a particular person. On the other hand, I see little potential for harm with scraping something like a public Instagram account. In this particular case, the AirBnb listers have agreed to having their listings promoted publicly, and thus any information that could be easily accessed by an AirBnb user should be permitted for public-consumption.