Zeinab Ghandour

SI 206 Project 2 Responses

a. If I were working in one of those companies, a different system I could think of would include verification and extra caution before the listing is public. I think user's with suspicious policy numbers should be required to get extra verification before they are able to profit. This can be done by adding additional verification when they type in the policy number on the website to ensure the number is valid. A system that requires less manual effort and one that takes place prior to the property being actively used over an extended period would be better. It should make it more difficult for an illegal listing to slip through. I could imagine the organization being against this system because it may frustrate users that do not fall under the suspicious category and may require more time from them which could affect the overall opinion of the organization. It may also cost Airbnb some money and additional effort on their end if they choose to go with this system, which they may not be willing to put forward given their previous behavior towards serious problems that harm people being hosted on their platform.

b. According to the Housing insecurity in the United States Wikipedia, "57% of housing insecure households are made up of unmarried individuals that do not have children." This is also the population that would likely need and realistically searches for a one-bedroom home. The price for the rent in these types of homes has increased over the years, to a point where a single-bedroom place could cost as much as a small mansion in another state that has a lower cost of living. If I was a data scientist working to fight against housing insecurity, I would possibly explore how many of these rentals are one-bedroom places using the data collected from web-scraping, and additionally find out how these listings are impacting the population. San Fransico has a large homeless population and a high percentage of home insecurity, as well as being a desired destination for so many tourists interested in different things offered by the city. I would explore the correlation between these two events through the data collected.

c. Something that is very important to consider when addressing the legality of web scraping is: how much you're collecting, where from, and what information. There are laws implemented across the world that addresses the results web scraping can have. Like Australia, they have the Spam Act of 2003 to outlaw the collection of email addresses.

America also has a few things put in place to control web scraping, but the details of what counts are not always clear. Web scraping can also harm companies/ people if used recklessly, an example given the Wikipedia page is the case with QVC and Resultly. Yes, the information provided through QVC's website was public, however, this negatively impacted the organization by crashing the website which lead to them losing money. Resultly used "sometimes to up to 36,000 requests per minute" which I think is not ethical or fairly utilizing the tool (Wikipedia). Another bad way to use web scraping that enters the legality of things is related to social media and collecting data on the users.

d. I think if the web scraping is done maliciously or without properly adhering to guidelines can create a lot of privacy issues. Something similar that I have seen along these lines is a company taking millions of "public" photos from users' social media and including them in a database for machine learning. Even though the good intention of improving machine learning was there, they crossed a boundary by collecting this information. This was basically irreversible, and if I was a user who had their photos collected I would feel like boundaries were crossed and that my privacy was compromised. When deciding what data to collect, the company/person must put thoughtful consideration into their plan. They must understand the type of information that is being collected and what it will be used for. Public data such as the date joined by a user may be okay, however collecting pictures, or sensitive information is unethical.