

Roger Barber

SI 206

11/2/2022

### Project 2 – Critical Thinking Questions

a.)

If I were to work for Airbnb, I could think of a way to pre-emptively solve the problem of invalid listing numbers. Like how we used regex to check if the formats were valid, it would be a relatively simple fix to use JavaScript to check for the correct format within certain locales (in this case, San Francisco). It would require cooperation from the city in order to evaluate which zip codes or areas are required to have licensing, but it would make the company more compliant with local regulations and also be the ethically right thing to do. It also makes sense to have an automatic flagging system for anything that does not match the required format, and further sorting those flagged listings into items that can be automatically assumed as exempt (they are a known hotel location, i.e.), items with incorrect formats, and pass those along to a human being for review. The most ideal solution would be to link a database of valid license numbers and their locations between the city planning office and Airbnb and have a real-time connection that validates business numbers as we go, preventing a listing from going live if their ID number doesn't match an up-to-date business number and periodically checking that currently valid listings have not let their license expire.

b.)

I think exploring the effects of the density of Airbnb properties within certain neighborhoods and how that can further affect the cost of living and availability of housing within those neighborhoods would be a very interesting topic. The Wikipedia page describes over 50% of housing insecure people are single and have no children, meaning they are looking for properties about the same size as a lot of Airbnb rentals – one to two bedrooms, small space. Airbnb properties directly affect two key factors of cost: price and availability. The more investors come and purchase small living spaces, the less there are available for local housing insecure individuals; furthermore, the less housing there is available, the more valuable the remaining housing becomes, allowing landlords to charge an arm and a leg for what remains. On top of that, many Airbnb homes/apartments are expensive to live at long-term and disrupt the formation of communities by breaking up neighborhoods with temporary housing.

c.)

I think the most important factor when considering whether or not web-scraping is appropriate, especially as someone that may be designing software or systems that do this, is the potential harm you may cause the organization or individuals involved in the site you are scraping. On one hand, I don't believe web scraping is violating an organization's privacy in the least – this is data that they are putting out in the public domain. Being able to collect and use that data, inherently, is not a criminal activity. However, the intent of collecting that data, the end result of using that data, and the effects that you may cause while scraping are all important aspects of the perceived legality. If you are harvesting sensitive or protected data to sell personal information to another business, that is dubious at best. It would be the same if you were sending requests several hundred times an hour, slowing down a website

or causing it to crash or operate at lower levels; you are obstructing someone's business! If you are grabbing information to provide a service, and just using it to organize information in a helpful way for consumers, or using it to verify that someone is being compliant with regulations or laws, that is an acceptable way to use scraping.

D.) I think my above question referenced some of these guidelines pretty well. Avoid scraping identifying information when possible; maybe avoid collecting full names with addresses and phone numbers as a cohesive unit, for example. Also, if someone has requested that you do not scrape their site, it would be a generally good thing to listen to what they say. The most ethical uses of web scraping are collecting data that doesn't spread someone's contact information, and has a clear purpose and thought behind it that does not involve selling information for money or inflicting harm on someone else either on purpose or through negligence. In essence, if it COULD harm someone, then don't do it. My final assumed guideline would be that it is generally more okay to scrape from larger companies or services; the chances of hurting them and their bottom lines is much lower, and let's face it – they're usually the ones that have more stuff to hide. Oh, and corporations aren't people.