

Question 1:

If I were a software developer for Airbnb, then I would propose a system in which each Airbnb house owner had to upload a picture of their business license. After the image has been uploaded I could create a program that uses computer vision to verify that the license is valid for short-term rentals. This could potentially be more effective than the current system as there is no possibility of user error when typing in the license number. Additionally, it would prevent the owner from simply typing in a random string of numbers that follow the same structure as a business license but are not legitimate. There are, however, many possible arguments I might hear from Airbnb about this system. For one, this makes our setup process slightly more complicated for the Airbnb owner. This is because they have to take the time to find and upload a picture of their license rather than just the number. To some, this process could take time as they may have to print or find their license. Secondly, implementing this system may cause fewer Airbnb's to be on the marketplace. As of now, though it's illegal, several of the houses on Airbnb are operating without proper licenses. This means that Airbnb as a company is still driving revenue from having these homes on its website. Although they will eventually take the illegal listings down, they will likely be able to gain extra revenues from the houses while they "verify" the listings. If this photo process were to be implemented, then the only way to get an illegal listing on airbnb.com would be to go through the effort of faking an entire business license. Overall, this system has many pros in terms of preventing illegal listings but it has cons in terms of profit loss for the company.

Question 2:

Using the data provided on the Wikipedia page one research question that I can ask is which states have the highest housing insecurity rate. This would be ideal for a housing activist

organization to know and understand because it will help them to learn and understand where they should focus their efforts. I can scrape this information by gathering data from the chart that is posted on the Wikipedia page regarding the housing insecurity rates by state. Additionally, I can also gather information on how incomes relate to housing insecurity. This can be done by scraping the "Rates" paragraph for various values having to do with income. This can help me to provide the activists with further information on target demographics to assist when trying to lessen the housing insecurity issue. Finally, I can scrape data regarding the average household size of people who are living in insecure housing. This data could be vital as it would provide information on specific niches that need the most assistance. It could be that in the scraping we find that those who have children are more likely to be in insecure housing. This means that perhaps the best route for the activists would be to try to get legislation passed regarding an additional income for people who have at least 1 child.

Question 3:

When discussing the legality of web scraping the most important factor to consider is the reason that one is choosing to web scrape. For instance, if the programmer has created code that web scrapes data and is then using that data in a for-profit manner. In the case of FareChase, they were taking website data from various airline websites and selling software that took advantage of those prices. To me, without permission from the airline company, this type of web scraping should be illegal. It causes financial harm to the company that is being web scraped from and capitalizes on that in order to make money for themselves. On the other hand, in terms of conducting research, I believe that web scraping should be both allowed and encouraged. For instance, in the second article, I believe that professors and data scientists should have the right to web scrape in order to provide evidence for algorithms that

discriminate. This should be allowed as it is making the internet a better place, and not taking advantage of anything. It is merely identifying and sharing the knowledge in a not-for-profit way.

Question 4:

When we are deciding whether or not to use public data the first guideline we should consider is how the data was collected in the first place. For instance, were the subjects of the study told that their data would only be used for one specific purpose? Or was the study clear in that any researcher could use this data in their own way? This is crucial to determine as it helps us to identify the intent of the people who were used in the data set. If the data was collected in a way where the people were told their information would be used in one way, it would be unethical for someone to scrape the data and use it for its unintended purpose.

Another guideline for us to consider when web scraping is how inclusive the data set is. Since we are not collecting the data ourselves, we must ensure that the data collected is representative of the entire population before we scrape it. If this is not done, we would run into an issue seen today with many facial recognition algorithms. These algorithms are often not inclusive as the datasets they are based on do not contain enough diversity.

Finally, we should consider if the initial data set was collected in an ethical way. This means that we would have to read through the study on how the data was collected. If we find that the data was sourced or collected in an unfair way, it would be immoral for us to use the data for our own purposes. By using the data we only perpetuate the unfair collection practices.