

Project 2 Questions

- a. If I was a software developer at Airbnb, a new system I would suggest to put in use to ensure that all policy numbers posted to the Airbnb site are valid is a 24-hour verification period after a user posts their listing where Airbnb employees will check with the city's planning office to validate the policy number provided by the user. This would ensure no invalid policy numbers sneak through the system and would force Airbnb to do their due diligence before allowing rentals to be posted. An argument that Airbnb might make against this system is that it delays rental listings from hitting their site which slows down the pace at which users can make transactions and the pace at which Airbnb makes money. Additionally, they could argue that it creates a need for increased manpower to verify all these listings which would require the company to divert more resources to policy number validation.
- b. A question I could answer using this data is: How does the average long-term rent cost compare to the average short-term rent cost available on Airbnb?
- c. A factor it's important to consider when discussing the legality of web scraping is what the scraper intends to do with the data. If they mean to test algorithms for discriminatory tendencies, that adds value to society and serves as an important check against "invisible hands" that guide websites. Scraping for this purpose should be allowed far more often than scraping that has malicious intent and is motivated by selfish or potentially damaging purposes.
- d. Guidelines that should be followed regarding web scraping:
 1. Personal information posted by a user of a website that isn't publicly available and easily accessible to the average platform user should not be scraped. However, if it is public in their profile that should be permitted because they are putting it out there for anyone to see.
 2. Data gathered from the scraping of private personal information should not be sold by the scraper to a third party. Scraping should not take advantage of how a website displays their data to profit off of users' private information.