Meghana Srinivasa

Barbara Ericson

SI 206

11 November 2022

Project Questions

1. I am imagining that I am a software developer at the San Francisco Planning Office. A different system that would verify that business licenses are valid for short term rentals is a program that counts the number of Airbnb listings in San Francisco on the Airbnb website and compares it to the number of business licenses the local government has given out for short-term rentals. If the number of listings on Airbnb is greater than the number of business licenses distributed, that means that one or more listings must be an illegal rental unit. One argument the SFPO might make against this system is that it would need to be run frequently which would require time and effort costs that the SFPO might not be willing to pay. Another argument that the SFPO might make is that the system does not provide exactly which listing is illegal, just that there is one or more illegal listings on Airbnb. They might argue that it is better to create a program that might take longer to run but would provide which exact listing is illegal.
2. One research question that could be explored using the data is: How are rising Airbnb prices increasing the value of San Francisco to the detriment of the local residents? The data that would primarily be used are cost of the Airbnb rentals per night. A data scientist could track the cost of Airbnb rentals and analyze how changes in the price of rentals affect the housing crisis in San Francisco. Another research question that could be explored is: How many bedrooms in San Francisco are in Airbnb rentals? The data that would primarily be used are number of bedrooms for the Airbnb rentals. The SFPO limits the number of rental units, but they might not consider of how many bedrooms each rental unit has. One rental unit that has 6 bedrooms can have a larger effect on the housing crisis than two rental units with one bedroom each.
3. I think one factor that is important to consider when discussing the legality of web scraping is the intention of the web scraper. In order to be held liable for committing “trespass to chattels” for web scraping, the plaintiff must demonstrate that the defendant intentionally interfered with the plaintiff’s computer system and that the interference caused damage to the plaintiff. I believe it is important to prove malicious intentions because it is very easy for individuals to violate a company’s term to service or “trespass” on their computer system. Terms and services are so long and dense that it is unreasonable to expect every individual to read and follow the terms. If an individual violated terms and service accidentally because they were unaware of the terms, then I feel that the individual should not be held liable.
4. One guideline to consider is if the data can be traced back to an individual or if the data identifies an individual. I think personal data should not be web scraped because it puts them at risk. If personal data gets publicized, it opens up that individual to phishing attacks and hackers. Another guideline to consider is to make sure that the information being retrieved is mainly factual in nature. This guideline helps prevent copyright infringement. A final guideline to consider is if the information being scraped is necessary to create beneficial end product. This guideline helps to ensure that a website is not being over scraped and the data being scraped is essential. It also makes the web scraping consider their intentions behind web scraping.