Project 2 Reflection Questions

a. If I am a software developer at Airbnb.com, I would develop a plan to automatically decline any rental whose policy number does not follow one of the two standard formats. Because all of the legal short-term rental options follow one of these two formats or say pending or exempt, it is highly likely that the properties that do not have this format made up a policy number and are illegally operating. An argument that I may hear at my organization is that this system is too harsh, but I would explain that this new system is a way to strategically discourage people from illegally operating their short-term rentals. Another concern and argument against this system that may arise is that the new system could make a mistake and decline people who actually do have a valid policy number or are in the midst of the process of obtaining a number. If this were to occur, Airbnb.com would ensure that there was a way to dispute this rejection, and if a rental owner is able to provide evidence that they were wrongly declined, Airbnb could reevaluate posting their property on the website.

b. One research question that we could answer using the data that I have found from web-scraping would be how does the cost of housing offered on a rental property website compare to one another? Data scientists who are working with a housing activist organization that works to fight against housing insecurity could see the large fluctuation of prices from just the first page of Airbnb as the results found in the csv_out function reveal that prices go from $82 to $399 for a night in San Francisco. This is a large range with the majority of options costing over $125 per night. The pricing of these rentals would be representative of the population and indicate that many houses in San Francisco would be quite expensive. The data found could indicate that housing insecurity could be a severe problem, and finding less expensive options is essential for the housing activist organization to solve this challenge. To find the most accurate answer to this research question, the process of web-scraping should be continued on many more pages to further see the fluctuation of prices with more data points in San Francisco's short term rental property market.

c. One factor that I believe is important to consider when discussing the legality of web scraping is the fairness aspect. For example, web scrapers may have an unfair advantage when utilizing their knowledge and skills on a certain website. In the Wikipedia article, a case called eBay vs Bidder's Edge involved web scrapers being able to access, collect, and index auctions from eBay. They were also able to place automatic bids or auction sniping. This provides the opportunity for people to take advantage of other bidders as well as those who are auctioning their items. Other bidders who participate in the auction the legitimate way will not have a chance to beat the web scrapers as their automatic bids can immediately take over the auction. Due to this issue, the legality of web scraping puts certain individuals at an unfair disadvantage and should not be allowed within the US.

d. One guideline that should be used when deciding to use or not use public data should be that if it is traceable to a specific individual or population, it should not be used. For

example, any identifying data points like location (city), defining characteristics, names, or contact information should not be used in any circumstance, even if it is found in public data. A second guideline to consider would be why a person is using this data. If a person is using personal data without one's consent for their own experiment or a class project, it can be allowed as nothing harmful will be done as a result of this. On the other hand, if someone is using the data to provide targeted information for each individual, or monitor their activity, it is unethical to use this personal data regardless of it being public. If a person is trying to make a profit on data that does not have consent, this is completely unethical.