Abigail Finn, Kiran Krishnan, Jake Hays

a) One potential solution for the problem of validating short term rental licenses in the San Francisco area, would be for the SFPO to develop a database of approved short term rental licenses for Airbnb to utilize via API when a new listing is posted to determine the legitimacy of the license entered. Airbnb could also navigate the issue of exemptions and pending licenses by requiring these special cases to be verified by either Airbnb or SFPO representatives. One potential argument against this idea is that the database would have to be updated regularly which could delay the process of getting a license and thus negatively affect the rental market causing Airbnb to lose revenue and increasing rental prices for consumers. Another potential issue with this process is that it would require the verification of pending and exempt listings which would require attention and resources from both Airbnb and the SFPO. This also has the potential issue that either Airbnb or SFPO would be accountable for developing the verification system for new Airbnb listings and would have to develop systems to reverify licenses as they expire.

b) A data scientist could combine this data with data on housing insecurity in the US to research the effect of short term rental license requirements or exemptions on local housing insecurity, safety and affordability rates. They could also determine if there is a correlation between the rate of false short term rental licenses one Airbnb within a given area and the rate of housing insecurity, safety, and affordability rates.

c) I think that it is important to consider why companies and commercial organizations or websites might not want people to be able to scrape their data and what they could be hiding if their sites are not allowed to be scraped. In the article about legal issues, researchers, computer scientists and journalists were looking to scrape website data in order to test the equity of outcomes when algorithms are being used. If commercial sites are able to hide this data its likely that they will not be held accountable for intentional or unintentional discrimination in algorithms. Thus, the decision regarding legality is critical for protecting beneficial research and journalism and is also important in the ongoing fight for algorithmic transparency and accountability.

d) It is important to consider the privacy of individuals when we are using public data. Although everyone has access to this information, web scraping includes taking their data to make our own inferences on it. Since they did not consent to have their data being used

in that way, we must consider how we are portraying them with our findings and how our conclusions could affect them. It is also important to consider how we are going to display the information that we have collected. Passing off the data as our own would be unethical. One last guideline to consider are social implications of our discoveries. In the first project, we worked with SAT data as well as census information. If we were to publicize that information without any background context, harmful conclusions could be made against marginalized groups in society. When web scraping, we must be thoughtful in our portrayal of groups of people, and how their own information can be used against them.