

1. A different type of system that would verify that business licenses are valid or short-term rentals in San Francisco would be one that only accepts the correctly formatted inputs. If the input is not in the valid format, it will not be put on the database. This process would ensure that business owners are confident in what they want to put into the database and it also weaves out anything that would be considered invalid based on the formatting. However, one argument that could be posed against this system would be that people could put incorrect information but have the accurate format. In this way, the system would allow the information to be put on the database even though it is incorrect. Another strong argument would be that Airbnb would need some access to the SFPO's database. This opens a lot of doors about privacy matters both for business owners and the planning office. While the verification system does help, these arguments do pose some concern.
2. After looking through the Wikipedia page, one research question that could be explored with this data would be: what factors lead to housing insecurity, and of these factors, which ones correlate to one another? The Wikipedia article shows the rent by county and the median cost of a home by state as two stances how to view the discrepancy between communities in housing insecurity, but diving into the factors will definitely open up a different perspective.
3. Examining privacy matters is also important to consider when discussing the legality of web scraping. When information is published on the internet, anyone can have access to view it and with web scraping, this information can be taken and kept by anyone. This exposure could result in harm to people with information such as phone numbers or addresses. Additionally, in some cases, information as such can be seen as property and when taking it, there are ethical concerns that should be considered as well.
4. One general guideline that all web scraping users should follow is having clear and positive intentions. With this, they should also try to determine what the implications can be as well. In this way, the likelihood of negative outcomes decreases because when people are doing something with the intent of helping users, there is no room for them to have an intention of harming them. Another general guideline that should be considered would be checking if the information that is being scraped on the web is not invasive. Again, privacy matters are extremely important to consider and if someone's address, phone number, and name are all connected, there should be no reason that anyone on the internet can scrape that information and connect it back to the individual. With these two guidelines in mind, it is important to only scrape information when necessary to prevent harming the general public or an individual.