Neha Kotagiri
SI 206 Project 2 Questions

A. Web scraping is incredibly time consuming as it requires humans to go through and analyze the data. It would be beneficial to use an artificial intelligence algorithm to verify the policy numbers and automatically use the San Francisco Planning Office to complete the verification. This would save a lot of time for researchers who would otherwise have to spend hours and hours web scraping the data set.

B. One research question that could be explored using this data is how Airbnb is contributing to housing affordability (or lack thereof) in specific neighborhoods. According to the wikipedia article, housing affordability is defined as the ratio of annualized housing costs to annual income. An individual is housing insecure if they spend more than 30% or 50% of their annual income on housing. Though this would require more investigation, the prices of the Airbnb are a good place to start by comparing them to the rent prices of other buildings in the area and also the prices of Airbnbs in different areas/cities. An article by Travel Noire, describes the "Airbnb effect", hosts convert long-term rentals that could house local residents and families to short-term rentals for visitors, thus decreasing an already short supply of housing. When the supply of housing decreases, rent will increase which contributes to housing insecurity and affordability.

C. One factor that I believe is an incredibly important factor when discussing the legality of web scraping is the common good. In the article from the Electronic Frontier Foundation, computer scientists and journalists wanted to research how algorithms discriminate based on characteristics like race or gender. When it comes to researching the real life, tangible, and potentially agonizing effects that code and algorithms can have on people (especially marginalized communities), web scraping can be a powerful tool that could make a large difference in the way our world operates: as long as the research is done in a way that is fully ethical, carefully audited by neutral sources, and does not put the general public's data at risk of being leaked or sold to a data broker. In recent years, we have come to learn that technology is anything but neutral, and can have drastic effects on people, like when it comes to surveillance of black and immigrant communities, gender-based airport scanners, and more. However, the ethics of web scraping needs to be analyzed on a case by case basis, as some investigations could put people at risk and may not always end up advocating for the common good.

D. One guideline to follow is to use web scraping as a last resort: if there is a public API available that provides the data you are looking for, that should be used instead of going

in and scraping the data. As stated in the previous question, it would be best if web scraping was used only to help others and for the common good, like identifying illegal Airbnb properties that are taking housing away from everyday people. Another guideline to follow is "whether the action is authorized by reviewing the terms of use and other terms or notices posted on or made available through the site" as stated in the [Wikipedia article](#).