SI 206 - Project 2 Lindsay Nelson

Airbnb.com could utilize artificial intelligence in order to verify if a business license is valid for short term rentals in San Francisco. In order to do this, they would need to use training data that first shows the machine the correct format for business licenses. They would need to compile training data that show both invalid and valid licenses in order to determine machine errors. They could also implement the usage of APIs in their software. Airbnb could extract data from the San Francisco Planning Office by relaying the license number through an API and then returning the verification to the backend service. However, SFPO may be against adopting this system due to security issues and development requirements. First off, the API may pose security risks because there is a lack of end-user boundaries. A hacker may take advantage of broker user authentication in an API and acquire excessive amounts of data without authentication. Another disadvantage of APIs for SFPO is they may require development effort on their end. If SFPO is not compatible with Airbnb's API, they will need to reconfigure their website and database query.

If I was a data scientist working with a housing activist organization to fight against housing insecurity, I would be curious how Airbnb's nightly prices compare with the development of starter homes that was mentioned in the Housing Insecurity in the US Wikipedia page. Governments can regulate entry level housing that have to follow square foot requirements, zoning ordinances, and permits, which could create more affordable housing options. These units can be owned instead of leased like an Airbnb, so I would be curious how many weeks it would take to break-even cost wise between nights spent at an Airbnb and nights owning a house.

One factor that is important to consider when discussing the legality of web scraping is the difference in data available on specific websites. For example, a retailer might not be as concerned as a healthcare provider about web scraping because of the confidentiality of the data that could be exposed. It is also questionable what data is under copyright protection depending on the type of data the website provides. Furthermore, I think another important element to discuss when considering web scraping is who is using the data provided. If web scraping is being used for the purpose of potentially beneficial research, it may be a highly regarded opportunity, but if it is being done by a hacker, it may be viewed as a threat to security and therefore questionably legally.

It's important to think about whether or not using someone's data goes against ethical considerations. First off, I think we should consider whether or not there is a provided common good to a community when deciding to use public data. If using a specific dataset may negatively affect one person, but has the potential to help many people, it's important to consider why it may be beneficial then to use it for web scraping. Another guideline to follow when using public data is that the usage of the data should not provide a personal risk to the individual. If analyzing

a specific data metric, it's important not to share the individual's identifying information. Additionally, if the data relates to a highly sensitive matter, the web scrapers should consider whether or not consent from the individual should be required to use that specific individual's data.