Madison Leidner
11/9/22

a. **Throughout the project, we acted as investigators to uphold the system of accountability created by the Sam Francisco lawmakers: listers must register with the city's planning office and put the business license's number on Airbnb's website, Airbnb must display some effort in validating these policy numbers, and third parties can register a complaint of illegal short-term rentals with the city planning office. We used web-scraping to do the latter using several hours of our personal time.**
Describe a different system that verifies that the business license is valid for short term rentals in San Francisco and list at least two arguments you might hear at your organization (either SFPO or Airbnb.com) against adopting your system.

   The current system that is used is simply based on complaints made by the public. A different system that can verify that the business license is valid is by directly verifying the validity of a business license for short term rentals. Currently, Airbnb is only responsible when an issue is brought to them, which can be extremely unreliable. A system that scans the policy number provided on a listing can ensure that it follows the correct format. Although this system may seem to be more accurate and reliable, there may still be issues. Other employees may complain that creating a new program can be a waste since listers may create fake policy numbers that match what the algorithm is specifically looking for. Listers type in these numbers by themselves and can very easily create fake policy numbers that match what the real ones look like. Another argument may be that it shows distrust of Airbnb listers. The whole company of Airbnb is reliant on trust as people rent out their own personal property, but by having listers become aware of new screening, the company may demonstrate that they don't trust the listers which can cause some to not want to use this site anymore.

b. **The database we've created through web-scraping is a great data source of information for data scientists in order to answer and explore research questions. Skim through the Housing Insecurity in the US Wikipedia page and describe at least one research question that you could answer or explore using this data if you were a data scientist working with a housing activist organization to fight against housing insecurity.**

   There are many pieces of information that we collected from this project in regards to Airbnb rentals and therefore there are many different questions that can be further explored. One question that I would be interested in answering is, how affordable are these rentals based on the space they provide? The data we collected includes the price of renting an Airbnb for one night as well as how many bedrooms there are. These two pieces of information can help us understand the relative pricing for different sized rentals.

c. **As discussed in the introduction, the legality of web scraping is still uncertain in the US. Skim through the Legal Issues section of Web Scraping in the US on Wikipedia and this article about the legal issues with the Computer Fraud and Abuse Act, and describe at least one factor you believe is important to consider when discussing the legality of web scraping and why.**

One important factor I believe is important to consider when discussing the legality of web scraping is intentions and what some may intend to do with the data. More specifically, web scraping can be used as a way to promote social justice and contribute to social justice movements. For example, in some industries, people believe that jobs are hired based on gender or race, and with web scraping it can allow people to further investigate these ideas. In some circumstances web scraping may be illegal, but in most cases it is harmless and can even benefit the society from knowing the information.

d. **Scraping public data does not always lead to positive results for society. While web scraping is important for accountability and open access of information, we must also consider issues of privacy as well. Many argue that using someone's personal data without their consent (even if publicly provided) is unethical. Web scraping requires thoughtful intervention, what are two or more guidelines that must we consider when deciding to use or not use public data?**

One important guideline to consider is if the information may cause a harmful or a beneficial outcome to the public and individuals. Data that is made public does not necessarily have positive impacts when looked at further. Private information about people's background that is made to the public can violate privacy but can even cause individuals to become in danger as a result of it. Another guideline that should be considered is protecting anonymity. This is pretty similar to the first guideline, as making sure that a way to identify the individual can help with the privacy aspect of specific data. If data is collected about individuals that is later released to the public, those specific individuals should not be able to be identified as it shouldn't be able to get traced back to them. Overall, privacy is always the main issue for receiving data and by ensuring these two guidelines, it can help some to understand when releasing information to the public can be beneficial or harmful.