Aishani Agrawal and Blair Bocklet
11 November 2022
Project Two

Once you've completed the coding portion of this assignment, answer the following questions using the following information.

Turn in your answers to these questions as well as your code.

1. **Throughout this project, we acted as investigators to uphold the system of accountability created by the San 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. Imagine you're a software developer at either the San Francisco Planning Office (SFPO) or Airbnb.com. 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.**

   Our recommendation is to create a system in conjunction with the San Francisco government that allows Airbnb users and listers to type in the business license number or paste the link to the rental. Upon typing the license number, if the policy number is registered with the city planning office, the screen will be green. If not, the screen would be red. With this system, Airbnb users can check the listing before they book to ensure the rental will not cause any problems. The first argument that my organization may raise in my attempt to adopt this system is that it rests entirely on the user of Airbnb to go through the verification process themselves which could be deemed unnecessary or time consuming thus leading to the step being ignored. Another possible argument my organization may raise is that it would require our business to work closely with either the San Francisco government or Airbnb. If the business we were collaborating with was running behind and the business license number was not entered into the database, the verification system would be ineffective despite the license number being processed.

2. **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.**

   If I were a data scientist from a housing activist organization working with the data from the Housing Insecurity in the US page, one question I could explore would be: What drives an increase in housing prices? This question could be answered and explored by inspecting the various risk factors, impacts, and locations that are provided by the database.

3. **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.

When discussing the legality of web scraping, especially in terms of the Computer Fraud and Abuse Act, one factor that is important to consider is how the webpage uses user data. In the article about the legal issues with the Computer Fraud and Abuse Act, the author discusses how, in Sandvig v. Barr, web scraping was used to create tester accounts and determine if discriminatory results appear based on the characteristics of the different users. In Sandvig v. Barr, the tester accounts were used on job posting sites, a style of webpage that takes specific data about the individual user. When we scrapped Airbnb, however, the site is not taking the idiosyncratic data of each user, rather it is just taking broad data such as the location the user decides to search and the duration of their stay. The legality of web scraping heavily depends on how the user's identifying information gets used by a website.

4. **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 to use public data?**

When deciding whether we should use or not use public data, there are several factors pertaining to the individuals for which the data was collected that we must consider. First, we should consider where the public data was collected. If the data was collected from willing participants who are aware that their data is being circulated and have approved this process, then it is okay to use it. If the data was collected years ago without the user's direct knowledge of how the data would be circulated at the present time, then it would be unethical to use the public data. A second factor of deciding to use or not use public data is thinking about how the use of the public data would affect the people from which the data has been collected. If the data had names attached, we must consider the repercussions that analyzing and studying the data would have on a person whose data was utilized.