Tamariah Davis
Project 2
11/10/2022
Professor Barbara

<div align="center">Project 2 Reflection Questions</div>

GitHub Link: https://github.com/SI206-UMich/proj2-fall2022-tamarida

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.

a) 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.

   i) A different system could be a field in the AirBNB host input, where the field is required and must be in the correct format. I highly recommend this option because it would not allow fake or invalid policy numbers to be listed. The field would have back end programming that would make sure the user enters the policy number with correct formatting. One argument I might hear at my organization is that the system would just allow invalid policy numbers in the correct formatting. My response would be that the field would not explicitly tell the user that the system is looking for a formatting. It would just deny them until the formatting is correct. Another argument is that if the user doesn't know that the system is looking for particular policy number formatting and the user's policy number is incorrect. The user could possibly get frustrated and our organization would lose business. I would respond by acknowledging that user experience is important to our organization, but also state that there should be a system in place to prevent illegal listings. For these issues, we would provide a phone number for the user to talk to a representative about why their policy number isn't working. I do not believe this would be a repetitive issue for our representative and believe it could put AirBNB hosts who have unverifiable licenses on our organizations radar.

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.

i) One research question I could explore is how AirBnb contributes to the housing insecurity problem with the amount of illegal listings in San Francisco using this data if I were a data scientist working with a housing activist organization to fight against housing insecurity. To explore this question I would look for the amount of invalid policy numbers, assuming all that we find are illegal because the illegal listing in the project was the only invalid policy number. I would compare the amount of assumed illegal listings with the total number of Airbnb listings in San Francisco to the percentage of housing insecurity in the city of San Francisco.

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.
   i) Web scraping is a form of copying in which specific data is gathered and copied from the web, typically into a central local database or spreadsheet. One factor I believe is important to consider when discussing the legality of web scraping is copyright. In the United States, website owners can use copyright infringement (compilation) to prevent undesired web scraping. Copyright is important because it protects creative work. This is especially important in the digital age because it is so easy to copy things and claim them as your own. There should be more laws about copyrighting while web scraping a site.

   Source:https://en.wikipedia.org/wiki/Web_scraping#:~:text=terms%20is%20unclear.-,United%20States,(3)%20trespass%20to%20chattel.

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 to use public data?
   i) Web scraping could be harmful because of the kind of data the organization has collected. Many argue that using someone's personal data without their consent (even if publicly provided) is unethical. One method is using a public API when available and avoiding scraping all together if the data you're looking for is available through the API. This would help with ethical issues when it comes to web scraping by just avoiding doing it in general. Another method I would consider when deciding to use or not to use public data is to pass your data through a user agent string to identify who you are.

Source:

https://www.meritdata-tech.com/resources/blog/data/web-scraping-best-practices-ethical-data-collection/#:~:text=Ethics%20of%20Web%20Scraping&text=Use%20a%20Public%20API%20when,number%20of%20requests%20per%20second

People I worked with: Isabel Lopez, Nylah Omar, Nyia George