Name: Christopher Sayah
ID: 37311035
Group Members: Jason Kemp, Remi Goldfarb, Morgan Huseby, Emily Veguilla (Office Hours)

Project 2 Responses:

- 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.
    - o Another system that could be implemented to verify that business licenses are valid for short term rentals would involve categorizing the validity of each license in accordance with its expiration date. By tracking the timeframe for expiration, this company would be able to validate the existence of each license more easily. Additionally, they would be able to push certain licenses into an inactive status upon expiration. Two arguments that may push back upon this strategy include: 1) the additional resources required and 2) possible system failure. The first issue regards the amount of time and money that would need to be invested to both gather and cross reference expiration date information with business license data. The company may not deem this issue substantial enough to employ more resources toward this situation. The second issue focuses on the accuracy of the system. By incorrectly pushing a license into an inactive status, this company could be held liable for the unfair repercussions that ensue.
- The database we've created through web-scraping is a great data source of information for data scientists 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.
    - o If I was a data scientist working with a housing activist organization, I could employ this data toward exploring the following research question: Is it more costly to rent or purchase a home in the long run? While considering the various costs of home ownership—as well as the dilemmas that may arise with it—it may be simpler and cheaper to rent for certain individuals. While renting, it is important to recognize that maintenance and overall housing stability is often independent of the renter. While buying, it is the exact opposite—maintenance and all associated costs/time are explicitly dependent on the owner. Comparing these two over time may provide a clear-cut answer on which is more effective monetarily speaking.
- 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.

Name: Christopher Sayah
ID: 37311035
Group Members: Jason Kemp, Remi Goldfarb, Morgan Huseby, Emily Veguilla (Office Hours)

- o One of the most important factors to consider while discussing the legality of web scraping is the notion of authorization. It may be reasonably argued that this gray area is based upon a lack of transparency between companies and their users. Even if a company directly states in their terms of agreement that they provide data access to third parties, it is exceedingly rare that users read and understand this information. Regardless of the statements made by the companies and agreed to by the users, it is imperative that both are equally aware of the terms of the agreement. This may include ensuring that a company is descriptive of how third parties may employ the gathered user data and ensuring that users read the terms of agreement. Additionally, it may behoove these companies to develop a detailed list of terms with the third parties accessing the information.
- 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?
  - o Two guidelines to consider while dealing with the ethical nature of web scraping are as follows: 1) ensuring consistent privacy and 2) employing a utilitarian lens. The first guideline should be employed to ensure that the data utilized cannot be traced back to specific individuals. If it is deemed reasonable to web scrape, there is absolutely no reason that certain people should be explicitly linked to the data. Connecting large groups of individuals to information is a much simpler way to provide security for the people whose data is used. The second guideline should be designed to make sure that there is a line between using public data for an increase in company profit and the overall benefit of society. Web scraping public information to better understand or improve the nature of a society is vastly different than employing this tactic for private profit gain.