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.
    a. A system that verifies the business license is valid for short term rental would be a separate third party system that users would use to upload a scan/document of their business license and policy number that has been approved from the San Francisco Planning Office. As of now, renters can type in their own policy numbers and business license, which will be verified but it is not as efficient or trustworthy. However, someone could hack the third party and work around to verify their business license, so it is important to have multiple systems verifications and terms and conditions that prevents users from the possibility of hacking the site. In addition, someone could forge a scan which would result in properties not actually being verified and reliable. I also think a program to verify the format of the Airbnb.com inputs would be useful and efficient. Such as, users can only input in a format that is parallel to short term rental policy numbers. Therefore, it prevents the homeowner from inputting wrong information or falsifying information. However, some homeowners are pending or exempt so that could be a loophole for this system or cause additional issues.
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.
    a. One research question that we could answer or explore using data would be does the living space have a bathtub or a shower? There are several factors that go into housing insecurity and a house is deemed inadequate based on structural conditions like having a bathtub or shower. To answer this question, web scraping could be used on websites that provide information about houses/living situations to determine whether the living space has a bathtub and shower therefore deeming the house adequate. In addition, I could use web-scraping to find data about neighborhood quality. I could scrape different web pages to see if neighborhoods are described with undesirable characteristics such as unsatisfactory county/city services, boarded-up units, defects on the roads, no stores within fifteen minutes, and more. As a data scientist this would help to define certain neighborhoods as low quality and essentially allow a housing activist organization to prioritize working to improve the quality of these certain neighborhoods and overall fight against housing insecurity.

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.
    a. I think it's important to consider whether such action is authorized by reviewing the terms of use and other terms or notices posted on or made available through the site. It is important to understand how certain actions can have a negative impact on a company. It's an interesting concept because a lot of people think that anything online should be allowed to be accessed by anyone. Also, users don't generally look in the terms of conditions/notices because of how intricate and lengthy it is. However, in the article it discusses a browser wrap contract or license which would basically put a boundary between the scraper and the website so that only people with a contract or license are able to gain access to or use material from a specific site. I think this is important to consider because not many people even know what the legalities are surrounding what you can do with data and information of websites. I never stopped to think about how it can be illegal to pull data from a public website. However, after reading this article I can see why it can cause issues. For example, with the QVC situation Resulty's seemed to try and go undercover using an anonymous web crawler without authorization in order to crash their website. Obviously, that shouldn't be a possibility especially because there is so much competition in the online space and everyone should have equal share to the market without having to worry about another company planning to take down their site and affecting their revenue. This is why it's important to think about who can make requests to websites, how many requests they can make, and if there should be a specific type of authorization to allow people to access the data.
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?
    a. One guideline that we must consider when deciding to use or not to use public data is what is the purpose and intention behind using the data and how does it help users. We want to make sure that there are no intentions of trying to crash a website and that we have good intentions to use the data to inform us or help to solve a problem. We need to make sure that when requesting data from a website we are aware of how the amount of times we request form a website will impact the server and overall website to ensure we don't make a mistake of requesting too much and crashing a website. Another guideline we can use is to make sure that web scraping from the specific site we are trying to use is legal. Oftentimes there could be a statement in the terms of conditions or a notice that prevents web scraping and we want to make sure we are abiding by the rules. I

know oftentimes people think that they are entitled to any information that’s online because it's public data and accessible to anyone; however, that isnt always the case and we want to make sure that we aren’t breaking any rules.