Project 2 Questions (Yun-Jung Chang)

A. Since the policy numbers are entered by the lister, they could easily type in a random number or make a mistake. Currently we are using web scraping to verify the policy number. However, sometimes we are not able to get all the data we want by using web-scraping. Another way that can be used to verify the business license is to create a filter for the current data entering system that will only accept the valid data with the lister's information. The filter should be applied before the listers enter their license's number. This can be done by using regular expressions to check if the data entered are the correct data type and format. One argument might appear from the lack of specification for data entered that are checked with regex since regex might not be able to go beyond a basic level of complexity. Another argument would be that regex is probably not good for parsing HTML because it contains too many optional parts and special rules.

B. According to the Wikipedia page, housing insecurity is the lack of security in one's shelter that is resulted from high housing costs relative to income, poor housing conditions, unstable neighborhoods, etc. One research question that I can answer or explore to fight against housing insecurity could be "Are the houses overpriced in this area?" by analyzing the costs of the place that we have scraped. This is important because if the houses are overpriced in the area it generates housing insecurity issues that leads to lower affordability for more people.

C. One factor that I believe is important to consider when discussing the legality of web scraping is when the person web-scraping from others is making profit out of the data. Web scraping itself isn't illegal. However,scraping confidential information to make profit is illegal. I think it is important to consider this factor because the laws have not been very clear and strict on protecting these intangible properties but businesses are still suffering from intellectual property abuse.

D. I do agree that using someone's personal data without their consent is unethical. One guideline that needs to be considered when web scraping is whether the data is being spreaded, especially personal information. Not only does it lead to privacy issues, it can also lead to safety issues as the general public can gain access to others' information and can possibly do harmful things with it. Another guideline with web scraping is that it could potentially lead to plagiarism as one can copy others' business models when using their data.