Kristen Huang

SI 206

SI 206 PROJECT #2 QUESTIONS

1) They could create a system that will double-check the validation of the policy numbers rather than having a third party register a complaint. Every day, they should check the policy numbers to make sure that they match and once they don't, they should automatically contact the city planning office. One argument that the organizations may bring up is that that may take too much time to search through all of the listings but it will be safer since there will be no need to have another party filing complaints. Another argument that they might bring up is that they will need to create another system that will alert the city planning office that the rental is illegal which is more work for them to do.

2) A research question that I could explore with this data is how housing insecurity is affected by the location of the home and the annual income. Based on the graph they showed with the key to the house income, we can see that the most populated areas are usually in the color dark purple to indicate that the income is high. The housing insecurity there will be lower than in places that are shaded in yellow.

3) It is important to make sure that people understand the effects of web scraping and why it may be considered "bad" and why it could be "useful" in some ways. This is important because people could take advantage of web scraping and be violating copyright laws.

This can also lead to hackers having access to the information on a website which can be dangerous for the owner. However, we need to establish to what extent those who want to web scrape can. There should be an indication of what you can do with the information if one chooses to web scrape.

4) As I mentioned in the previous question, one guideline that we must consider when deciding to use public data is what we are going to use it for. Someone wanting to use the data publicly versus someone wanting to use it for a school project or for practicing their skills should have different rules they must follow. Another guideline that we should consider is to what extent one should be able to use the information that they web scrape. It should depend on what they decide to do with the information.