

Project 2 – Reflection Questions

- A. If we were software engineers for Airbnb, to verify that the business licenses are valid for short term rentals in San Francisco, we might build an algorithm that verifies that policy numbers are valid by double checking each license with the San Francisco Planning Office's (SFPO) database. In addition we can automatically detect policy numbers that have an incorrect format (much like the way we did in the `check_policy_numbers` function), delist that rental, and reach out to the owner of the property to resolve the issue. With the SFPO's agreement to work with us, most of this process can be completely automated up to the point of contacting the property owner. One argument against adopting this system is that in order to verify each business license with the SFPO, Airbnb would need to be given access to the SFPO's database. With the increased transparency between Airbnb and the SFPO, Airbnb might open itself up to extra scrutiny from the SFPO, which the company would probably like to avoid. Another argument against using this system is that it can potentially delist some properties, that actually do have a valid license. The owner could have just submitted the wrong code or made a typo by mistake, or the SFPO could have a delay in their database and be unable to verify some valid licenses. While these situations would be resolved eventually, the time and effort necessary to resolve a situation like that would be a major cost to Airbnb.
- B. One research question we could explore using this data is what factors lead to housing insecurity and how do those factors interact and depend on one another. One factor that leads to housing insecurity is home price depending on the county that you live in. The Wikipedia article shows a graph with data on home prices for counties in the United States. Thus, one can conclude that there is a correlation between housing insecurity and the county you live in.
- C. Some factors we believe are important to consider when discussing the legality of web scraping is copyright infringement as well as violation of the Computer Fraud and Abuse Act. Web scraping can raise legality concerns because extracting and copying data from a webpage can be seen as stealing property from an organization, which raises ethical concerns as well as legal.
- D. Some guidelines we must consider when deciding to use or not to use public data is if this public data is absolutely necessary and vital to the person who wants to use it. Specifically, if a person needs to do a project for a class and has the choice to either collect their own data or scrape it from the internet, they should make a decision to either collect their own data or take it from the internet and cite it properly to make it clear that it is not their own. Another guideline would be to try to get in touch with the person who owns the data and ask for

their consent to use it and cite them correctly.