Project 2 Questions

a. An alternative system to make sure the license is valid would be one that cross references the license number put on the airbnb site that shows San Francisco actual license. This would make it easier to spot whether or not someone has mistyped their license number, or even worse, if they have made up a fake one. One argument to this may be that it would be costly to have a staff who tracks this kind of information, as it is a lot of data to have to cross reference. This may not be something the company can fit within their budgets. While having a team focus on this task could have many positives as it helps ensure a lister is not taking away a housing unit in San Francisco, it may not be enough to prevent this kind of thing from happening, and they company may find it pointless. Another argument may be the reliability of the government website to provide up to date information on the validity of the license holder. If our task of cross referencing is solely dependent on the San Francisco information, this is not concrete enough to know if these jobs will survive.

b. After skimming through the page a question that could be explored is rather or not the cost of a rental is reasonable to the area that it is in ? Reasonable being defined as a number that fits within the median range of prices offered for other properties in the area. We have access to the location, number of bedrooms , and bathrooms so through this information a pretty solid comparison can be made between the various local listings and their prices.

c. One factor that I think is very important when it comes to the legality of web scraping is ultimately the reason behind the web scraping. I am kind of in the middle of whether or not I think it is okay to sometimes violate a website's terms or not. In the situation of trying to research to find discrimination within algorithms, I feel that it is okay. If research is not done and close eyes are not often of the developers, it can be very easy for them to get away with unfair practices. On the other hand I can see how web scraping could potentially lead to the information of users being at risk and the developers may be trying to protect them. These are some factors that should be taken into account when deciding whether web scraping should or should not be legal.

d. When it comes to using information from people, one guideline should be in place to receive approval from whatever personal information is up for grabs. Some sort of contact should be made to let the person get permission to access and scrape their data. While this could cause some time delays, this term could be mentioned inside the terms and conditions which ultimately allows the person up front to decide if they would like their information to be shared or not. Another guidelines would need to be being transparent about how the person's information will be used. If sharing one's information can lead to the harm of their identity or physical safety then this should not be a route taken, but if it is used to determine an algorithm then potentially it could be okay. Something else that comes to mind is a guideline that ultimately hides the identity of the person if other information is being used. If their address and rent cost for example is a part of the data, then their name should not be included as a security measure. The most important guidelines should include things that consider people as humans and not just numbers that have been crunched onto a data sheet.