a. I think I can use JSON for scraping the data since I used HTML parsing and regex for my project. JSON is also use the HTML code for the data gathering, but we will use linked data for the JSON. We should use different notation for the JSON. I think the organization can argue me about the privacy issue because JSON is also using the public data on the public website. However, those public data can be the part of private personal data. Second thing that organization can argue is since there are still a few websites that are not support the JSON format, then if I use the JSON to parsing them, there will increase the security risks.

b. I think we can explore how the housing affordability affects to the price of the Airbnb houses using this data. I want to check that if the price of houses is expensive, there are cheaper Airbnb for the people to give a chance to rent a short-term housing without heavy money pressure. Or since the owning house is expensive, the Airbnb is also expensive to rent. Since our database have the cost and type of accommodation, we can compare the cost with the house in certain area.

c. I think using the real legal name is important because there are some people or group that are using the information in harmful way hiding behind the anonymity. Since ordinary people are believing and put their personal information to the website, I think government and organizations are considering web scraping as a illegal action. However, if we set the policy and regulation to use only real name for web scraping and letting people to know their purpose of web scraping, I think the fear of fraud or other harmful act will be solved. So, I believe that using the real-name system is the one factor that is important to consider when discussing the legality of web scraping.

d. As I told on part c, there is the big concern about the privacy when there is public accesses to web scraping. For the first guidelines, I think we need to consider about the real name as I said on the part c. Because if people use real name for web scraping, they could not use the information in the negative way which can harm the society. For the second guideline, I think it is important to make the people to read the possibility that their information can be scraped for other people’s various purpose and they need to accept or not for the accessibility. If they do not want some people to use their information, their information should protect in other way. Since lots of people always click the agree and most of the website always ask to click the agree to see their information, it should be changed.