Project Report

Question1:
Take a dataset of house price(such as the California Housing Prices) for example.

We all know the community and neighbourhood affect a lot on the price of a house. Therefore, if we add the average house price of a neighbourhood to the data, the data will include attributes that depends on other data.

Besides, there may be some extreme situations that have great effect on the house price. For example, the price of a house that has just been remodeled in a old neighbourhood can be much higher than other houses. The price of a house under a high voltage power line can be much lower than other houses. These are examples of outliners.

In conclusion, applying linear regression on such a dataset may pose an issue.

Question2:
The basis for the argument is that ethical stance comes with a method and can not be separated from the method. We have to adopt a method as well as its ethical stance.

I don't agree. I think the statistical method is ethically neutral, while our decision to choose any statistical method has some "ethical" stance. For different scenarios, we choose different method. For example, a security system of a small company can be very different from the security system of FBI. The system of FBI should be greatly penalized if it allowed non-authorized personnel in, while it is more acceptable if it denied authorized personnel to enter. Therefore, we should choose different model and loss function for different security systems. In this case, the method we choose has no ethical stance, our decision of method has ethical stance.