- 4/21/2021: I have finished the select and training a model section, I have found out that my data somehow produce a model not as good as the book’s. The book’s best model was Random Forest but mine was a Linear Regression! Perhaps due to the data collection or my data are relatively small to theirs could be a reason. Or Perhaps due to the nature of the data itself? Maybe I have to constraint it (regularize it) or get more training data. Also trying out more models might work as well. I also tried out some sort of things: theoretically, stratified sample can produce a better model or show a better representation of the population. However, when I used the random sample, it produces overall better models for all of the algorithms. (I have tried Linear Regression, Decision Tree Regression, Random Forest Regression).