ALY 6040 Data mining reflection

My role in the team is more like a coordinator. I also will document part of the work, but I will give it to Yin Jiang to audit and modify. This is because she has better English skills than me. I corporate Yin to finalize and modify the coding that after we discuss every time and not satisfy with outcomes. I give my opinions about how to do the data cleaning and modeling. I will also explain all questions for Yin about my code. So does she.

Yin is a responsible and amazing leader and organizer. In all the online meetings she is the sponsor and host. We will discuss each group assignment before, mid and after to know the schedule and adjust our plans. She is also a good thinker. She will discuss all her questions with me and we will verify them. Due to her proactivity and responsibility, I also affect by her to start to find the best solution to cater to her requirements even if the process is fairly hard and time-consuming. I learn more than ever because of her attitude with the study and work. She also gives me a lot of ideas about how to implement our data analysis why we should do this analysis. For instance, I never think about why I need do this EDA before she told me. In my opinion, I just randomly look at the data. However, she said when your report your boss, you should every single reason why you want to do this. She also told me the feature selection is only used for decision trees or random forests. I am truly glad and appreciate that she devotes herself to this project so much and gives me much help.

In the group project, my main job is to code and find a good solution to fit our model. I conduct a lot of EAD methods and plots, such as using capping to deal with outliers. However, after we discuss the house price outliers, Yin thinks we shouldn’t delete them because of the intense correlation with other variables. Thus, I have to delete it. Also, due to time-limited in the presentation, so we have to decide to omit many EDA plots. What’s more, due to the number of model requirements, we have to delete the decision tree model and feature engineering for the best subsets. To be honest, we both make a lot of endeavors to make our final project as better as possible even if it cannot be a perfect report and algorithm. Nevertheless, we both will stay hungry, stay foolish in the data analysis field. When we dive further, we will find abundant treasure and evoke our thirst for knowledge and curiosity.