For this group project, I was mainly in charge of coming up with the prediction model and writing the code for our analysis, as well as contributing the same section (Section 3) for the final report. For the first part of our project, my group focused heavily on deciding which metrics to use for our model and how we will be choosing them. Firstly, we each tried to find the top-performing metrics for different datasets, and my part was on finding the top metrics for traditional team statistics. I decided to use the correlation coefficient to find the top metrics because it is a straightforward way to measure how strongly each metric is related to win percentage, allowing me to quickly filter out irrelevant or weak predictors. Later on, we just decided to just focus on metrics in both traditional and advanced team statistics instead of including other datasets such as player statistics, as it would get too complicated. We used R^2 and correlation coefficients to determine the top 5 metrics that could be used for the prediction model. After deriving the metrics, I then went ahead to construct the prediction model using linear regression. Because the 2025-26 season's statistic was incomplete, I decided to split the past 10 season datasets we had into training and testing datasets so that I can train the model and then test its accuracy later on. Next, I used the least squares methods taught in lecture to compute the weights for each metric. Then, I put weights back into the model equation together with the testing datasets and got the predicted win percentages for the latest season and compared them to the actual test datasets. Lastly, I coded the evaluation of the model by calculating the R^2 value and absolute mean error against the testing dataset.

However, after we were done with this model, we faced a challenge. My group realised that using the top 5 metrics is too straightforward and does not provide us much analytical insight. This led us to come up with another model that includes metrics with low correlation coefficients but still good metrics to help us predict win percentage. I contributed by identifying metrics that were not directly tied to winning (e.g. not using wins and losses) and by ensuring that selected metrics were not subsets or duplicates of one another. Then, we had to find out reasons why, despite using those metrics, it still gave high model accuracy, which was also quite challenging, as we were surprised by the results we got. Through group brainstorming, we eventually manage to come up with explanations in the end that it was because the previous model's metrics were redundant.

After all the codes were done, I was in charge of combining all the codes into one single script, which was quite confusing to me at first, as I had to change all the file directories, file names, and variables so that they were consistent and aligned for the code to run.

Through this group project, it has strengthened my analytical thinking and taught me not to be easily satisfied with the initial successful result. I have also learned to work closely with my groupmates and overcome challenges together as a group. Whenever I faced difficulties with my part, be it coding or interpreting the model, I would actively discuss them with my groupmates.