

Name: Peng Yu

ID: s4663588

After intense discussion, our group finally decided to choose the 2019 League of Legends World Match data as the data set for our analysis. Because, first, our team members are all-league of Legends game fans, have a deep understanding of the game, we can more easily analyse the data set. Second, although League of Legends is the most popular game at present and has become one of the events in the 2018 Asian Sports Games in Jakarta, e-sports is still not accepted by the mainstream media. We hope that we can introduce the League of Legends so that more people can feel the charm of e-sports.

After obtaining the data, we conducted simple processing of the Data, including Data Dimensions, Missing Data, Data Cleansing, and Data Imputation. Then we divided 93 groups of data into five groups, each group of data to cross-verify the winner. After the cross-validation results were obtained, our analysis reached a deadlock, most of the results could not accurately predict the winner. After our discussion and analysis, we found that this is a reasonable phenomenon. If a single data can predict the winner and the loser, then the game is unbalanced in terms of this data. Therefore, we finally discussed the result, and integrated the two groups of data with the lowest error rate among the 93 groups, as two-dimensional data, and then analysed the winner and loser. Repeated failures in data analysis have led us to a profound understanding of the meaning of the phrase 'Data won't tell you what's right, but it can tell you what's wrong.'

The data7001 is a strong guiding course, teacher guidance, and education, let me as a new scientific data, a good understanding of the charm, and each set of data about data cannot evaluate the value of the shock, I don't know before I learned knowledge, are of great help to me, for my future in data science has laid a solid foundation of the study, thanks for all the tutor data7001 and section.