The world of E-Sports is growing at a very rapid pace and many young prospects are looking towards gaming as a potential career choice. However, how does one person decide if they have the potential to compete with the best in the world? How much time should be invested or achieve a certain skill to be better than an average player? What is the threshold to decide whether one person is suitable for this career choice?

The dataset I will be using is the SkillCraft1 Master Table Dataset Data Set from the University of California, Irvine Machine Learning Repository dated on September 20th, 2013.

I do not have a specific client, rather, my target audience is towards all younger generation considering professional gaming as a future career. Based on my analysis, I hope to enlighten them on how much minimum should be invested to be able to compete against the best in the world, or, if they are talented, what minimum skills must be achieved. Without the knowledge, the younger generation will never know the difference between their own talents versus the world professionals already in the business.

With the dataset provided, Players are separated by LeagueIndex, 1 being the lowest league and 8 being professional players. I will be focusing APM (actions per minute) as the skill level that much be achieved to compete against high level and professional players and using TotalHours and HoursPerWeek as indication to how much time casual players compare to high level players differentiate from each other. Professional player’s TotalHours and HoursPerWeek are not listed, so I will be inputting the numbers myself from assumption. Professional players most likely practice every day, which would be between 10-12 hours per day or 70-84 hours per week. TotalHours is impossible to compute because there isn’t a category of when each player started. In order to input the data, I will be using the game’s release date, multiply by the TotalHours by the years of the game release date to the dataset recorded. Both input values will be inputted through a random number generator between the listed numbers for variation. This is a categorical regression problem, so I will be using R and SVM (Support Vector Machines) to create the analysis. I will be using a RMSE (Root Mean Squared Error) and Cross-Validation with 5 folds.

My deliverables would be the source code, a word document with the explanation of my findings and a Powerpoint slide deck to present the key points of my findings.