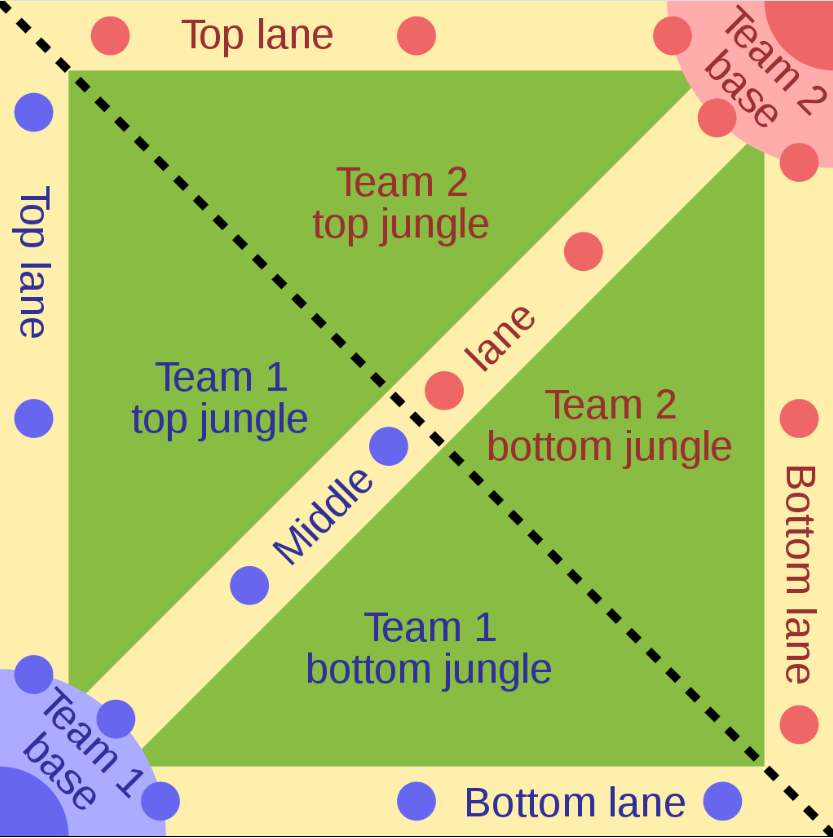
I remember 11 years ago being a freshman in college and a new, free video game came out called ‘League of Legends’. I always enjoyed a good multiplayer strategy game with friends, and had a good time with other’s playing it. As time went by and I graduated, I lost the time for playing video games, but I always had fond memories of it. Recently I was speaking with my neighbor who competitively plays video games, and he mentioned he plays League of Legends and that it is still strongly supported. I looked up the tournaments online, and saw there are now multimillion dollar events.

This inspired me to perform an analysis on this game for my semester project. What I was curious about, was if by performing analysis on certain variables early in the game, the outcome of the game could be predicted. I acquired a dataset which included 41 variables of 10,000 diamond level ranked games. An average game lasts 35 minutes. I decided to analyze the 10,000 games worth of data and create a model to generate a prediction at the 10 minute mark on which team will ultimately win.

For readers of this report not familiar with the game, I will provide a brief overview. There are two teams, red and blue. Red is on the top right corner of the map and blue is on the bottom left. Each team has a ‘base’ and the object of the team is to destroy the other team’s base.

See figure 1.0.

****

**Figure 1.0**

Notice the lanes and jungles. Lanes are visible to the players by default. The jungles are not visible unless that player buys an item called a ward, and places it in the jungle. The ward has an expiration timer.

As a mentioned the red team is on the top right. For a first basic test, I was curious if being on the top of the map versus being on the bottom had a psychologically effect on the players regarding perceived dominance. Over the 10,000 ranked game analyses the players on the red team, which was top, won 50.05% of the time, and users on the bottom blue team won 49.95% of the time. From this, I wouldn’t be concerned whether you end up on the top or bottom team.

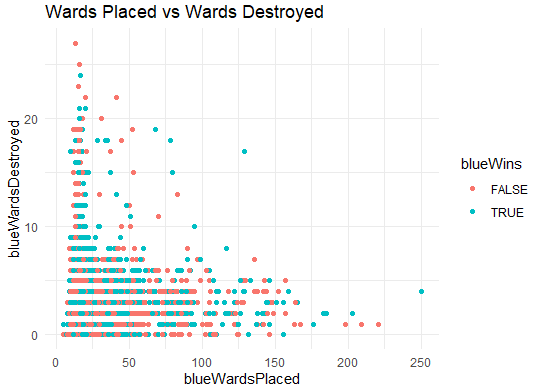
Previously I mentioned that the jungle areas of the map are not visible to the users unless a ward is placed. Online, there is resounding agreement that placing wards is for the better.

For example, on League of Legend website MobaChampion, it states that “Warding in League of Legends is incredibly important. Teams with a higher vision score and better vision control than their opponents consistently win more games. ...” .

I was not convinced that this is true, and dedicated the early steps of my project to create graphical representations to see the results.

First, I created a scatter plot of the wards placed by the blue team versus the wards destroyed by the blue team and color coded the scatter plotting by the win/loss.

You can see from this chart, that there was no perceivable affect on your win/loss for various degrees of placing or destroying wards. See figure 2.0



**Figure 2.0**

Next, according to MobaChampions placing wards better controls your opponents which would result in more kills. From scatterplot in Figure 2.1 of Wards vs Kills it can be seen that placing a minimal amount of wards did increase your chances of kills and winning. This only held true in about the first 5% of the spread. Exceeding that amount of wards, 25, starts to decrease your number of kills, and increased your chances of losing.

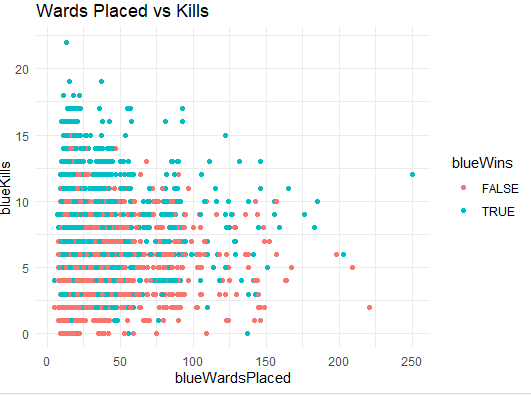


Figure 2.1

Interestingly, placing more wards does decrease your deaths, but does not increase your chance of winning. That remains consistent. See figure 2.2.

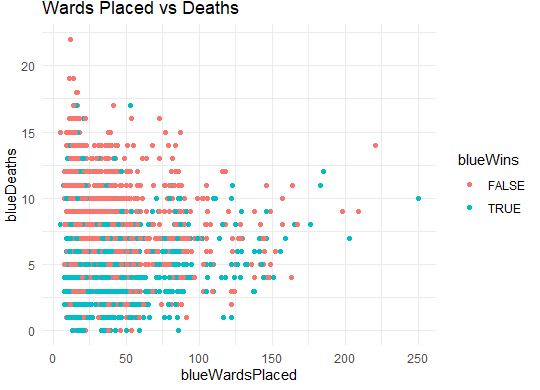


Figure 2.2

I then ran a correlation matrix, Figure 2.3, and reviewed the correlation between the variables. It can be seen that the p-value is approaching 1.0 which is results in low statistical significance, providing supporting evidence for the scatter plot interpretations. The detailed correlation can be seen in Figure 2.4.

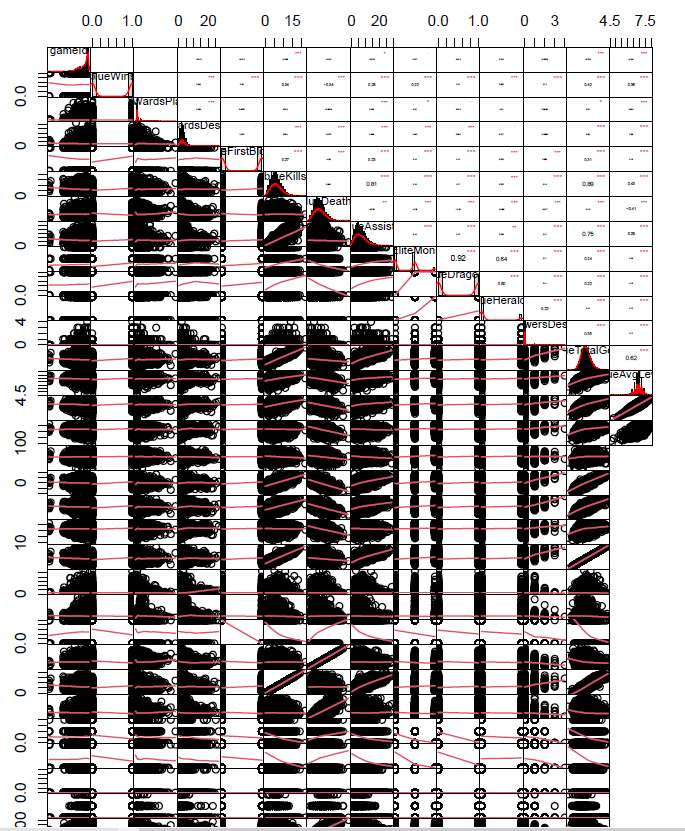


Figure 2.3

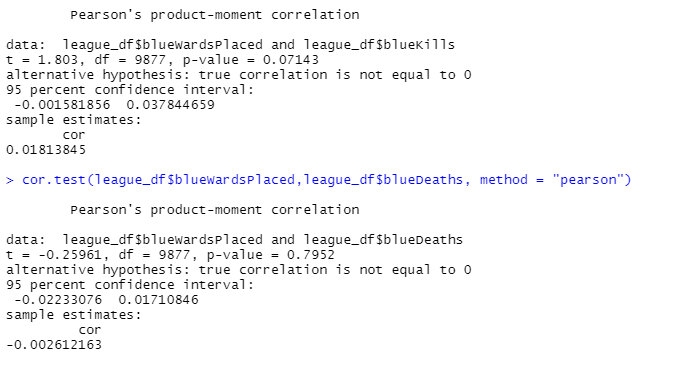
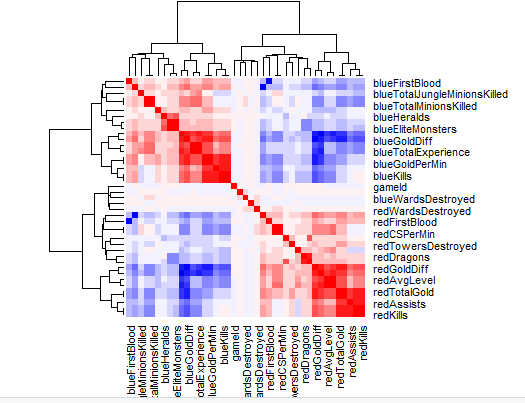


Figure 2.4 Pearson’s correlation

To view all the correlations in an easy to interpret visualization, I created a heat map figure 2.5.



Next, I wanted to create my generalized linear model to predict the winning team outcome.

The first thing I did was created my training and test datasets.

I started out by counting the total number of rows in the data frame.

I took 80% of the total data and used it to create my training dataset to train my model. I chose the 80% by creating a vector of indices which is an 80% random sample. The leftover 20% subset of data will be used for my test data to verify the accuracy of the model and see how well it can predict the outcomes.

This resulted in my training data sample size to be 7903 and my testing data sample size to be 1976.

Below in figure 3.0 the summary of my generalize linear model can be seen with the lowest AIC I was able to create for an optimal model selection.

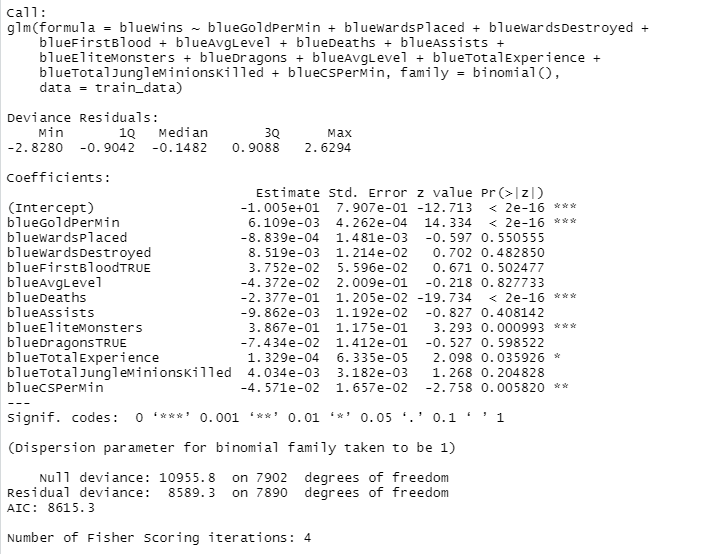


Figure 3.0

Once my model was built and statistical significance codes review per variable, I was interested to looking further by creating an Average Marginal Effects (AME) table which reports AMEs, standard error, z-values, p-values, and confidence intervals. What is of interest to the player here is that by increasing your elite monster kills, you will increase the probability of winning by 7%. AME table figure 3.1.

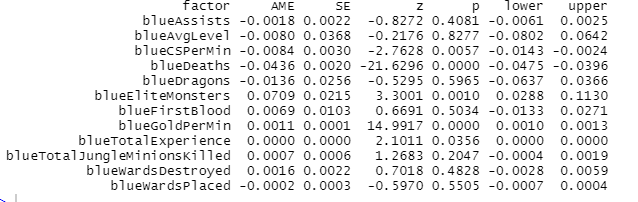
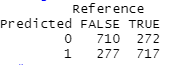


Figure 3.1

I then trained my created GLM model on my 80% train data that I split earlier. Once my model read the data and was trained, I ran it against the 20% of test data and allowed it to make predictions of the winning team.

I built the results into a contingency table. See figure 4.0.



Figured 4.0

From the table we can see that my model predicted 710 losses accurately and predicted 272 losses that were actually wins, a false negative. It then predicted 277 false positives and 717 true positives accurately.

I ran this model against game selections twice independently and consistently get the final result:

**My model can accurately predict the winning team in a League of Legends ranked match, 72.2% of the time at the 10:00 minute mark in the game.**

I think this is impressive results for a real world scenario, and I have gotten as high as 73% accuracy.

**Project Requirements:**

1. Overall, write a coherent narrative that tells a story with the data as you complete this section.
   1. Completed Above
2. Summarize the problem statement you addressed.
   1. Completed Above
3. Summarize how you addressed this problem statement (the data used and the methodology employed).
   1. Completed Above
4. Summarize the interesting insights that your analysis provided.
   1. Completed Above
5. Summarize the implications to the consumer (target audience) of your analysis.
   1. There are a number of implications to the consumer that I outlined throughout my analysis but I want to repeat, that in this analysis I have determined that warding is not as important as the Internet research states, and that the elite monsters kills play statistical significance in the outcome of the game.
6. Discuss the limitations of your analysis and how you, or someone else, could improve or build on it.
   1. My analysis is limited to ‘diamond’ ranked players due to limitations in the dataset. If we were to improve on this data, I would use the developer API and build a data set that has a sample which is mixed across various ranked skillsets. In this analysis, we are constrained to only determining how diamond players perform versus a more diverse sample size.
7. In addition, submit your completed Project using R Markdown or provide a link to where it can also be downloaded from and/or viewed.
   1. See attached.