League of Legends Visual Analysis

Peter Eng

2025-03-06

Data Overview

Recently, my friends introduced me to the online computer game League of Legends. I found this game to have a very steep learning curve and it motivated me to improve and catch up to my friends' level of skill. There are numerous moving parts and nine other players constantly making unpredictable decisions that can influence the outcome of the game. Since the game changes continuously, I was interested in knowing if there was an accurate way to predict the outcome of a game.

This data was pulled from Kaggle and contains measurable factors from the first 15 minutes of 24218 ranked League of Legends games from high ranking players in Europe West and Europe Nordic & East servers. The user that posted this data scraped it directly from the game developer, Riot Games, website to obtain the game state at 15 minutes and pull the measurable factors used in this dataset. I trust this data to be real since League of Legends has publicly available data that can be easily scraped.

The average game of League of Legends lasts around 30 minutes and 15 minutes is the cutoff as the earliest time a team can surrender without a player disconnecting from the game, so players will consistently try hard which helps minimize the possible outliers in the data that may include players purposely trying to lose the game. Also, these are ranked games from lobbies of very skilled players so if they build up a lead within the first 15 minutes of the game, they are more likely hold their lead and win the game. This makes the possibility of a model viable because of the unpredictability of lower ranked players.

The data already came fairly clean, so to make it tidy I had to properly rename the predictors, factorize categorical predictors, and remove some duplicate observations. To do this, I used R and the tidyverse library to read in the csv file, rename columns, mutate the datatype of certain predictor values, and remove duplicate observations.

There are 26 numeric predictors (13 for each team) along with 3 classification variables that determine which team (blue or red) got the first kill, which team won the game, and a unique match ID. Summary statistics for each predictor except for match ID are shown in Table 1. The first kill is significant because it gives and extra gold bonus along with the normal kill gold. Some predictors such as Control Wards Placed and Wards Placed contribute to a team's vision of the map since there is a "fog of war" effect when a team does not have a character or ward placed nearby. Other predictors such as Dragon and Herald kills refer to epic monsters within the jungle that give buffs while Jungle Minions are merely used to farm gold and experience (Xp). Minions that are not in the jungle also give gold and Xp. Towers, Inhibitors, and Turret Plates all refer to damage to structures where a Tower has 5 Turret Plates, but there are a total of 11 towers for each team, so the plates destroyed may be distributed across multiple towers. Team Kills and Damage to Other Champs refer to statistics of when characters fight each other. Since killing monsters, minions, and opposing players as well as destroying turret plates give gold while killing monsters and minions give Xp, some of these predictors are correlated. Within the League of Legends community, there is a consensus that gold, xp, and kills have the most influence on the game outcome. This is likely because they can represent various other predictors in one predictor.

Data Visualizations

Figure 1 compares the histogram of values for four different predictors by differentiating whether the blue team or red team won. Clearly, the predictors have different ranges of values, but the histograms for each predictor also have very similar shapes. For all four predictors, the winning team's distribution of values peaks at a higher value than the losing team. For example, when the blue team wins, the blue team oftentimes has more gold than the red team. We see this in all four predictors; however, Total Team Damage and Total Minions Killed have much closer peaks, indicating less of a difference between team outputs and less of an association with a team winning than other predictors such as Total Team Gold and Team Xp.

Figure 2 displays the correlation plot of predictor variables that identifies which predictors may be correlated with other predictors. One notable strong correlation is that Blue Team Dragon Kills are highly inversely correlated with Red Team Dragon Kills. This makes sense since only one team can kill a dragon, so both teams cannot kill the same epic monster. Also, Blue Team Total Kills are strongly correlated with Blue Team Total Gold and Red Team Total Kills are strongly correlated with Red Team Total Gold. Since gold is given when an opposing champion is killed, these correlations once again make sense within the context of the game. One unusual inverse correlation is Blue Team Xp being inversely correlated with Red Team Total Gold. This means that when the Blue Team has more Xp, the Red Team has less total gold. This could be an indicator that the Blue Team has more time alive on the map, so they have control of most contested areas. This denies the Red Team kills that give gold as well as objectives, minions, or structures that could also give gold leading to less total gold for the Red Team.

Analysis

Based on my Exploratory Data Analysis, I believe that if I were to perform a one-way ANOVA test, I would find significant differences in blue and red team predictor values based on which team won for predictors such as Team Total Gold, Team Xp, and Team Total Damage to Champs.

I also chose to model the data with Elastic Net regularization since it balances the variable selection qualities of Lasso with the ability to handle multicollinearity from Ridge Regression while still being explainable.

I implemented the Elastic Net model with the caret library and used cross-validation to determine the best alpha and lambda for the model. Then after getting the probabilities for classification of a win, I measured the accuracy of the model to evaluate correctness.

```
## [1] "Elastic Net Mean Accuracy: 0.748245831939999"
```

Conclusions

As seen in Table 2, I found that there were no significant differences in means of team gold or team damage, but there was a significant difference in mean team xp. However, when conducting the same test after separating the data by which team won, I found that there were significant differences in means for all predictors. Clearly, when the red team won, the red team had significantly more gold than the blue team. This relationship was similar for every predictor tested.

This dataset has very high predictive modeling potential since blueWin is one very clear categorical response variables. I used this dataset to predict blueWin and train an Elastic Net model that predicted the winning team with 74.82% accuracy. I chose Elastic Net since it can eliminate less significant predictors since there are so many in this dataset while balancing the correlation between variables and still maintaining a decent level of interpretability. This fairly high level of accuracy can give a fairly confident prediction of the outcome of a game after 15 minutes in games from high ranking players in Europe West and Europe Nordic & East servers.

GitHub Repository

This LoLDataPackage is in the GitHub repository "PeterEng3/LoLVisualAnalysis".

Table 1: Summary Statistics

Variable	N	Mean	Std. Dev.	Min	Pctl. 25	Pctl. 75	Max
blueTeamControlWardsPlaced	24218	3.6	2	0	2	5	37
blueTeamWardsPlaced	24218	41	43	9	25	35	603
blueTeamTotalKills	24218	13	4.9	0	9	16	38
blueTeamDragonKills	24218	0.74	0.72	0	0	1	2
blueTeamHeraldKills	24218	0.12	0.33	0	0	0	2
blueTeamTowersDestroyed	24218	0.65	0.89	0	0	1	10
blueTeamInhibitorsDestroyed	24218	0.002	0.054	0	0	0	2
blueTeamTurretPlatesDestroyed	24218	8.9	3.1	0	7	11	22
blueTeamFirstBlood	24218						
... 0	12014	50%					
... 1	12204	50%					
blueTeamMinionsKilled	24218	349	31	194	329	371	465
blueTeamJungleMinions	24218	88	14	0	78	97	156
blueTeamTotalGold	24218	27831	2740	17719	25911	29573	40968
blueTeamXp	24218	29573	1879	19061	28341	30836	36801
blueTeamTotalDamageToChamps	24218	32172	6131	11022	27933	36096	62857
redTeamControlWardsPlaced	24218	3.6	2	0	2	5	15
redTeamWardsPlaced	24218	43	47	9	25	37	576
redTeamTotalKills	24218	13	4.8	0	9	16	37
redTeamDragonKills	24218	0.91	0.75	0	0	1	2
redTeamHeraldKills	24218	0.1	0.3	0	0	0	1
redTeamTowersDestroyed	24218	0.96	0.98	0	0	1	10
redTeamInhibitorsDestroyed	24218	0.0015	0.044	0	0	0	2
redTeamTurretPlatesDestroyed	24218	3.9	2.6	0	2	5	15
redTeamMinionsKilled	24218	350	32	188	330	372	464
redTeamJungleMinions	24218	89	14	0	79	98	156
redTeamTotalGold	24218	27788	2693	18247	25910	29513	41227
redTeamXp	24218	29619	1896	17602	28387	30866	36797
redTeamTotalDamageToChamps	24218	32155	6040	10383	28024	36028	62452
blueWin	24218						
... 0	12241	51%					
... 1	11977	49%					

Table 2: One-Way ANOVA Test Results

Predictor	Winning Team	P-Value	Difference (Red - Blue)
Gold	Not Separated	0.0817070	-42.98497
Gold	Red	0.0000000	2384.50339
Gold	Blue	0.0000000	-2523.98063
Xp	Not Separated	0.0069154	46.33838
Xp	Red	0.0000000	1693.76464
Xp	Blue	0.0000000	-1637.40085
Damage	Not Separated	0.7604305	-16.86320
Damage	Red	0.0000000	3442.34327
Damage	Blue	0.0000000	-3552.31836

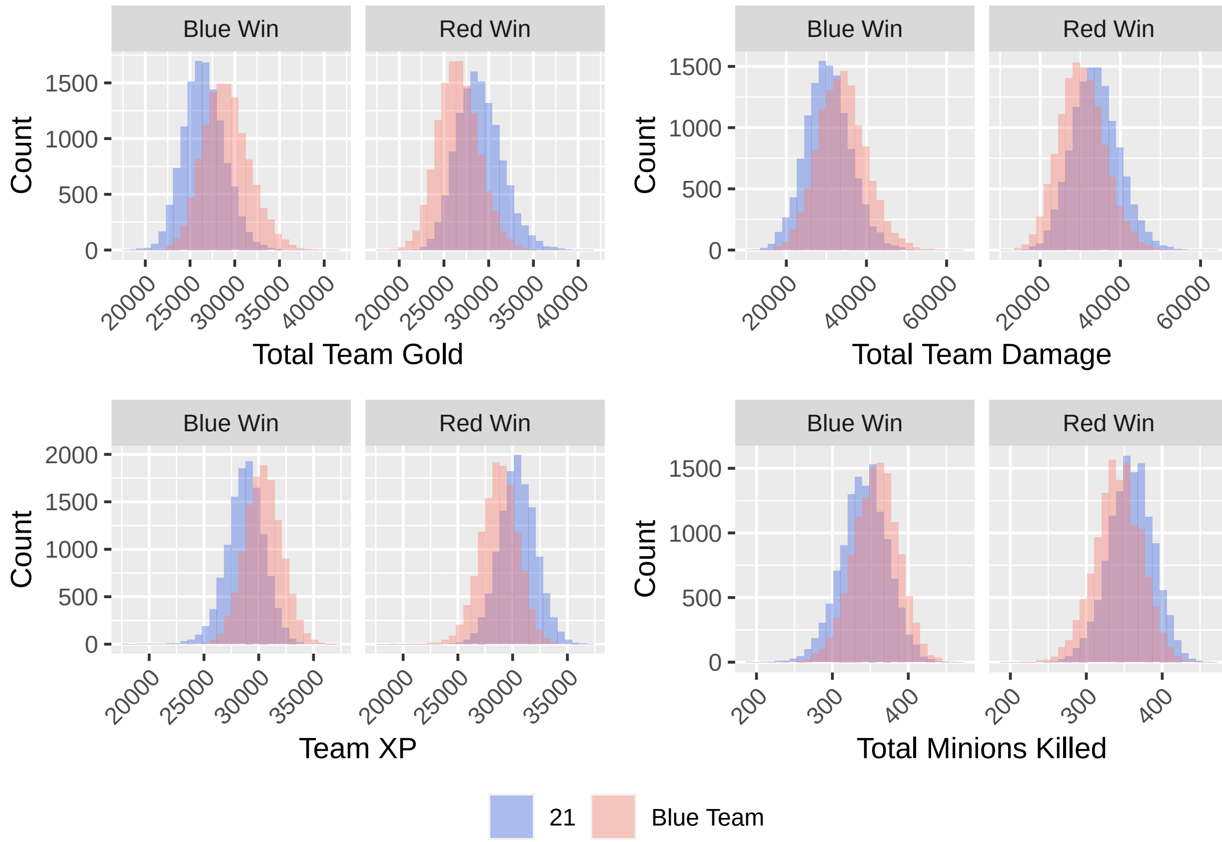

Figure 1: Histograms of Team Statistics

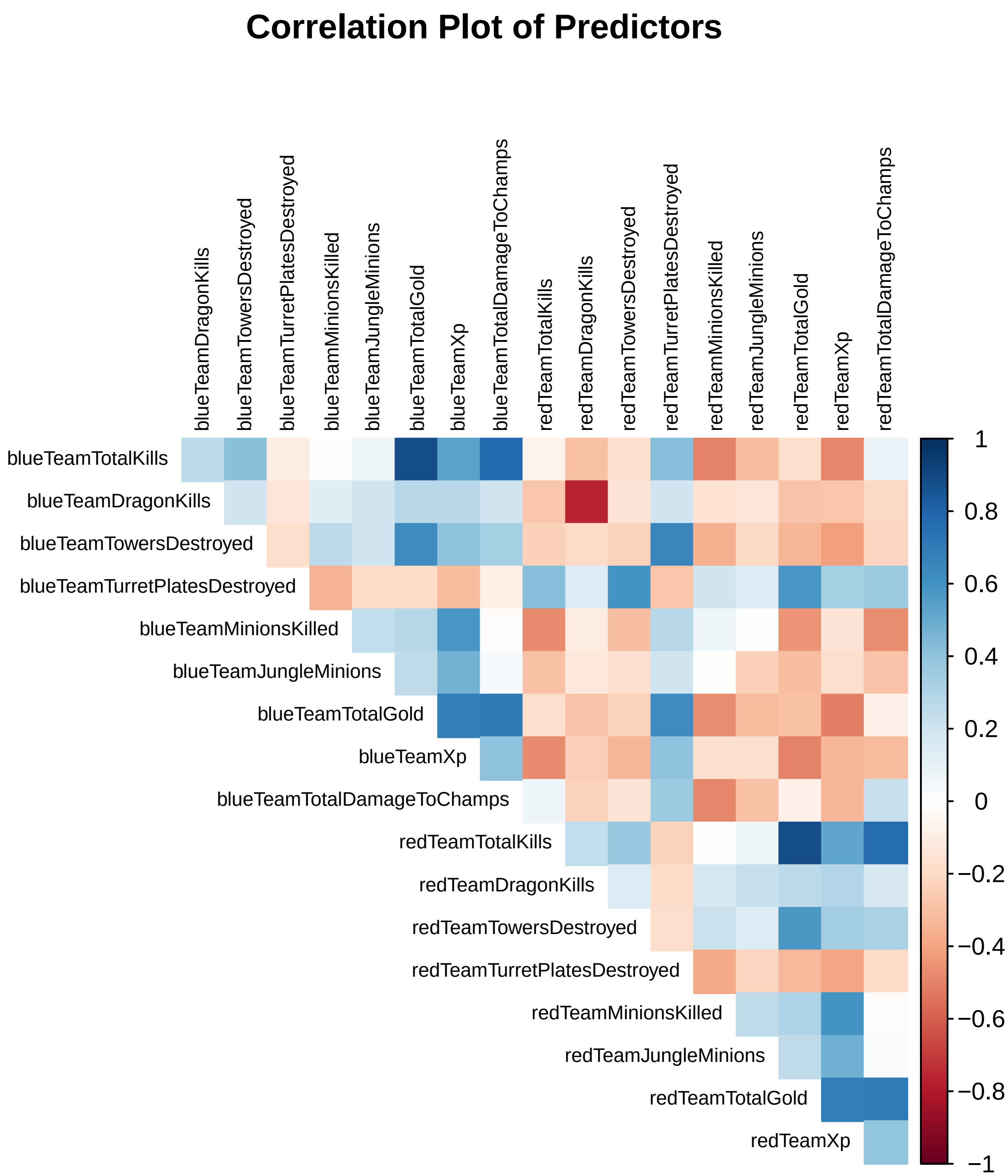

Figure 2: Correlation Plot of Selected Predictor Variables