

NOT-Project 1

Peter Eng

2025-02-07

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.

matchId	blueTeamControlWardsPlaced	blueTeamWardsPlaced	blueTeamTotalKills
EUW1_6882489515'	2	23	4
EUW1_6882416210'	2	46	12
EUW1_6881092720'	2	19	13
EUW1_6879405717'	3	28	8
EUW1_6879389461'	3	30	11
EUW1_6879371828'	2	24	3

blueTeamDragonKills	blueTeamHeraldKills	blueTeamTowersDestroyed	blueTeamInhibitorsDestroyed
2	0	0	0
1	0	1	0
0	0	0	0
0	1	1	0
1	1	1	0
0	0	0	0

blueTeamTurretPlatesDestroyed	blueTeamFirstBlood	blueTeamMinionsKilled	blueTeamJungleMinions
7 0		388	106
10 0		348	98
11 0		331	96
9 1		355	84
7 0		370	100
16 0		369	80

blueTeamTotalGold	blueTeamXp	blueTeamTotalDamageToChamps	redTeamControlWardsPlaced
23923	29798	21842	6
27852	30530	35817	6
28126	29501	38039	3
27333	30466	31338	0
27772	31066	26676	7
24011	26872	22414	1

redTeamWardsPlaced	redTeamTotalKills	redTeamDragonKills	redTeamHeraldKills
34	16	0	0
26	17	0	0
241	18	1	0
23	10	1	0
26	7	1	0
33	15	2	1

redTeamTowersDestroyed	redTeamInhibitorsDestroyed	redTeamTurretPlatesDestroyed	redTeamMinionsKilled
2	0	1	400
1	0	4	373
2	0	3	340
0	0	5	378
1	0	5	382
3	0	3	357

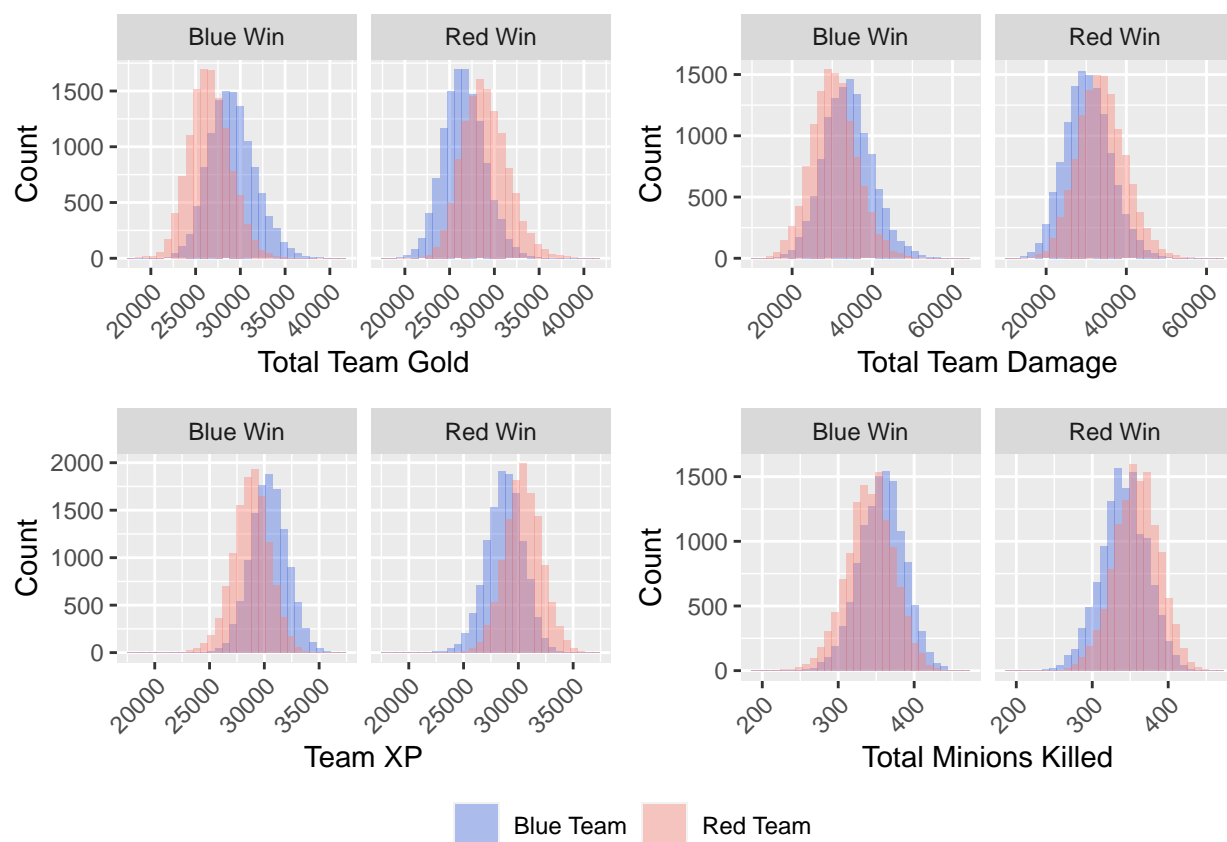
redTeamJungleMinions	redTeamTotalGold	redTeamXp	redTeamTotalDamageToChamps	blueWin
94	29637	32613	27239	1
76	28970	30320	36585	0
84	30510	29464	42984	0
87	25157	31069	28229	0
96	26052	29475	19245	1
70	28368	30422	30967	0

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.

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. The first kill is significant because it gives an extra gold bonus along with the normal kill gold. Some predictors such as

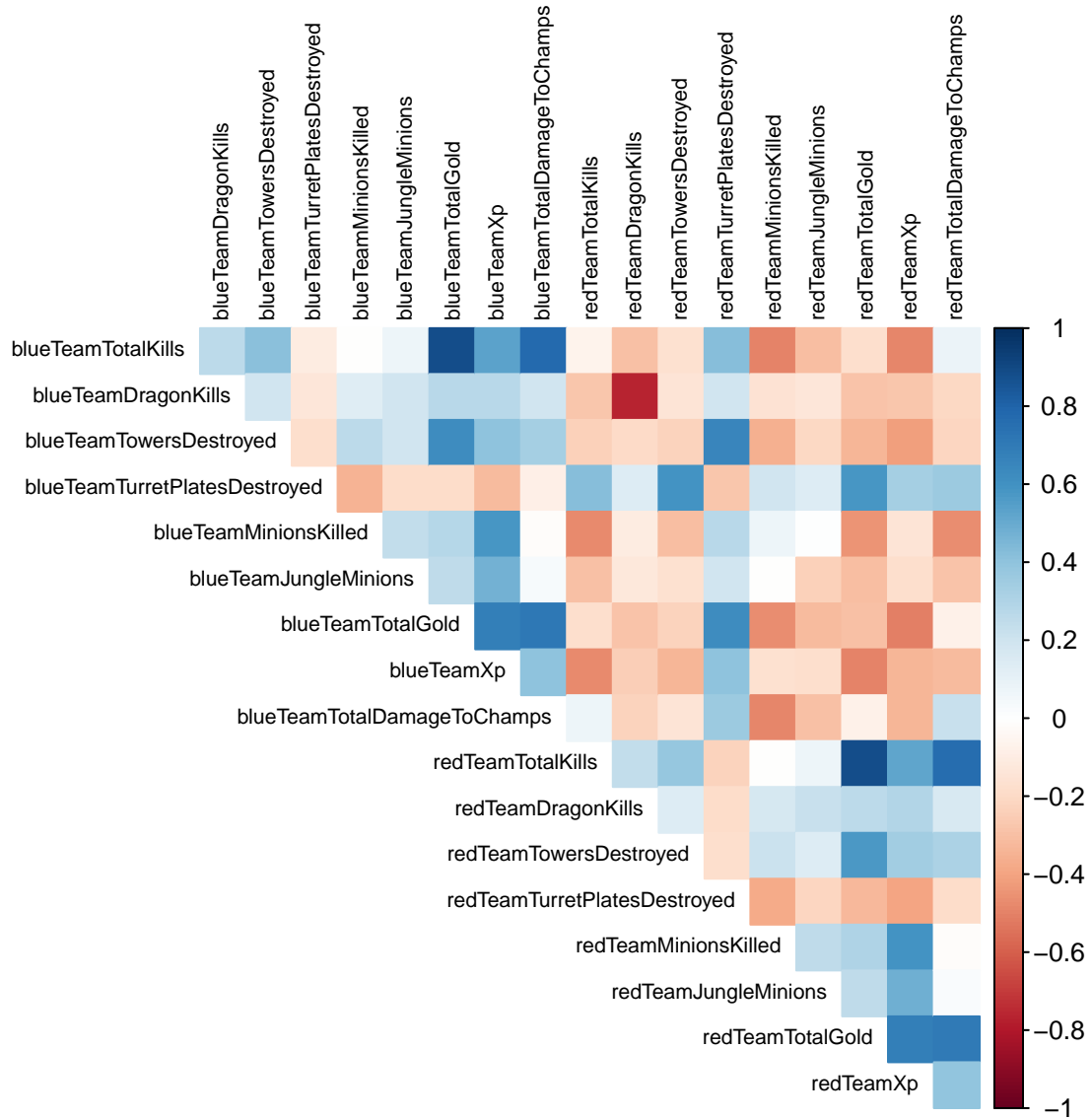
Control Wards Placed and Wards Placed contribute to a team's vision of the map since there is a "fog of war" effect when a team does not have a character or ward placed nearby. Other predictors such as Dragon and Herald kills refer to epic monsters within the jungle that give buffs while Jungle Minions 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



This figure 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.

Correlation Plot of Predictors



The correlation plot of predictor variables 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.

Conclusions

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.

This dataset has very high predictive modeling potential since blueWin is one very clear categorical response variables. In my Statistical Computing class, I used this dataset to predict blueWin and train an Elastic Net model that predicted the winning team with 74.7% 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.