League of Legends Wins

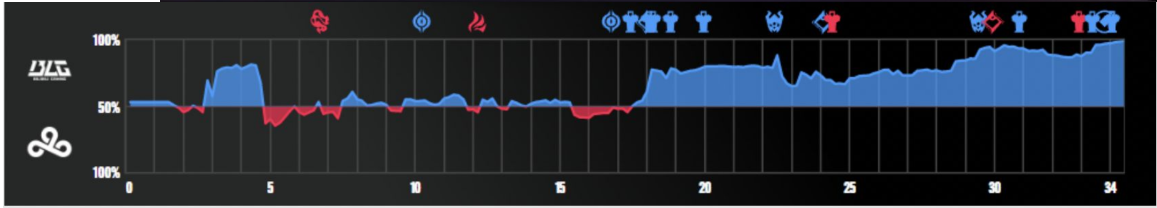
Claudia Lyu, Dhruv Chokshi,
Leona Du, Summer Wang

What is League of Legends?

- Game Structure
  - Summoner’s Rift
  - Blue and red teams
  - 5v5 MOBA
  - 3 lanes
  - Take nexus to win
- Game Components
  - Gold and Experience
  - Minions and Jungle Camps
  - Elite Monsters(dragons, heralds, barons)
  - Kills/Deaths
  - Towers

Why This Question?

- Popular Game
  - Millions of players
  - Billions in revenue
- Win percentage charts during pro games
- Determine how win percentage is calculated
  - Most important factors
  - Gold graph typically shown



Data From 9000+ Games

- Around 10k Games in 2020
  - Unique gameID
  - 38 features, 19 per team
  - blueWins target variable
  - No NULL values
- Ranked Diamond I to Masters ELO
- blueGoldDiff, blueExperienceDiff
- BlueDeaths, blueKills

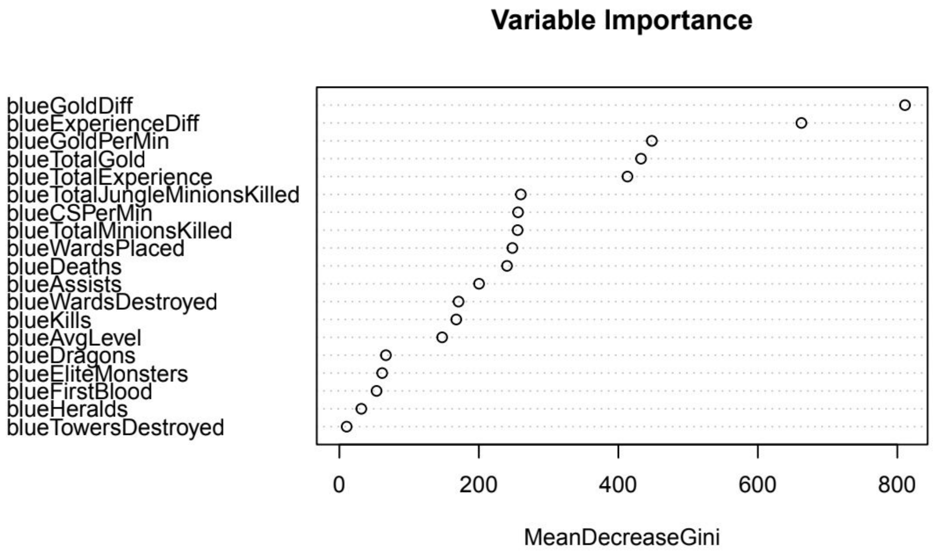


Figure 1: Variable importance for blue team

Exploratory Data Analysis

- Histograms
  - Bell curve/normally distributed
- Box plots
  - Kills, gold diff, experience diff higher when blue wins
  - Deaths higher for loss
- Correlation plots
  - Some correlation
  - Not too concerning

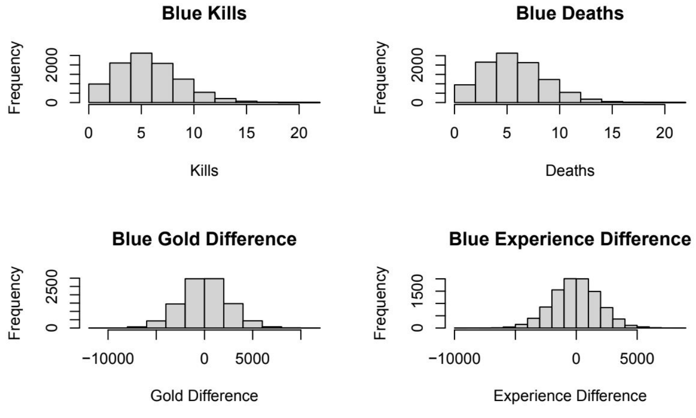


Figure 2: Hisograms of Blue kills, deaths, gold difference, and experience difference.

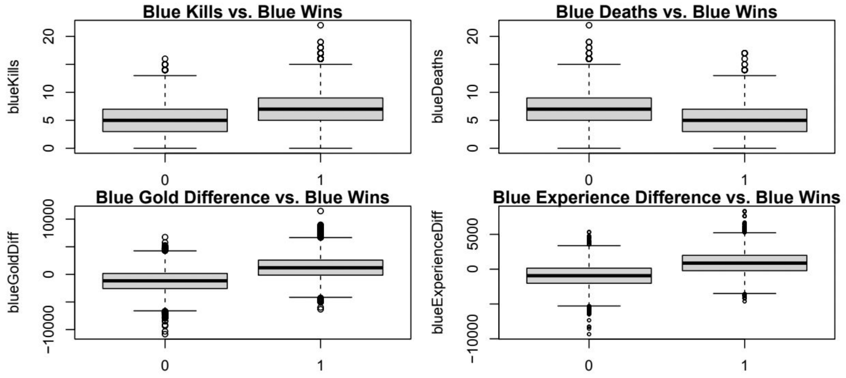


Figure 3: Boxplots of Blue Wins v.s. Blue kills, deaths, gold difference, and experience difference.

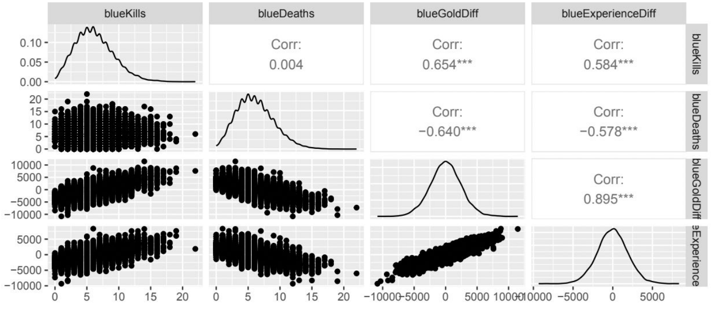


Figure 4: Correlation plots of interested variables, including blueKills, blueDeaths, blueGoldDiff, and BlueExperienceDiff

Q: How to Predict WIN Prob of Blue Team?

Hint: BlueWins (a binary variable, with 1=win; 0=not win); interested predictors

Answer: build logistic regression model! - model the log-odds of a win

- BlueWins ~ blueKills + blueDeaths + blueGoldDiff + blueExperienceDiff

Model Assumption:

- log(odds) is a *linear function* of four predictors

$$\log\left[\frac{Pr(\,blueWins = 1|\ blueKills,\ blueDeaths,\ blueGoldDiff,\ blueExperienceDiff)}{Pr(\,blueWins = 0|\ blueKills,\ blueDeaths,\ blueGoldDiff,\ blueExperienceDiff)}\right]$$

$$= \beta_0 + \beta_1 \cdot blueKills + \beta_2 \cdot blueDeaths + \beta_3 \cdot blueGoldDiff + \beta_4 \cdot blueExperienceDiff$$

How to Assess the *Linear* Assumption?

Empirical Logit Plot

- log(odds) v.s. predictor
- y: empirical logit values
- x: each predictor

✓ The trend looks **reasonable!**

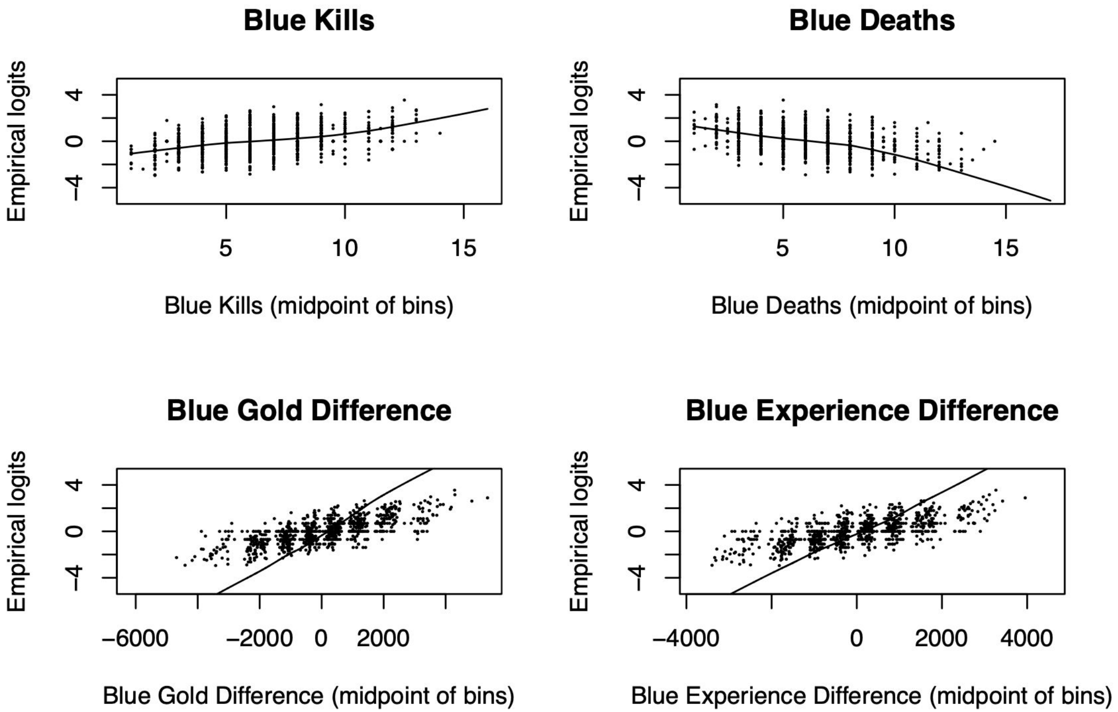


Figure 5: Empirical logit plots, where the empirical logit values are along the y-axis against each predictor (via the midpoints of the bins) on the x-axis.

Model Evaluation & Comparison

Eval: Out-of-sample Performance via 5-fold cross-validation

- $\text{RMSE} = sqrt(mean((test_y - pred_y)^2))$
- Std. Error = $\frac{SD(RMSE)}{\sqrt{k}}$

Table 1: RMSE, plus/minus one standard error, for the logistic model of blueWins on blueKills, blueDeaths, blueGoldDiff, and blueExperienceDiff.

lower_rmse	cv_rmse	upper_rmse
1.0804	1.0864	1.0923

Model Evaluation & Comparison

Alternative: logistic regression model of BlueWins on *BlueGoldDiff only*

Again, out-of-sample Performance via 5-fold cross-validation

- $\mathrm{RMSE} = sqrt(mean((test_y - pred_y)^2))$
- Std. Error = $\frac{SD(RMSE)}{\sqrt{k}}$

Table 2: RMSE, plus/minus one standard error, for the alternative logistic model of blueWins on blueGoldDiff.

lower_rmse	cv_rmse	upper_rmse
1.0813	1.087	1.0927

Model Evaluation & Comparison

Table 1: RMSE, plus/minus one standard error, for the logistic model of blueWins on blueKills, blueDeaths, blueGoldDiff, and blueExperienceDiff.

lower_rmse	cv_rmse	upper_rmse
1.0804	1.0864	1.0923

RMSE of alternative model is slightly higher

Table 2: RMSE, plus/minus one standard error, for the alternative logistic model of blueWins on blueGoldDiff.

lower_rmse	cv_rmse	upper_rmse
1.0813	1.087	1.0927

What About Uncertainty?

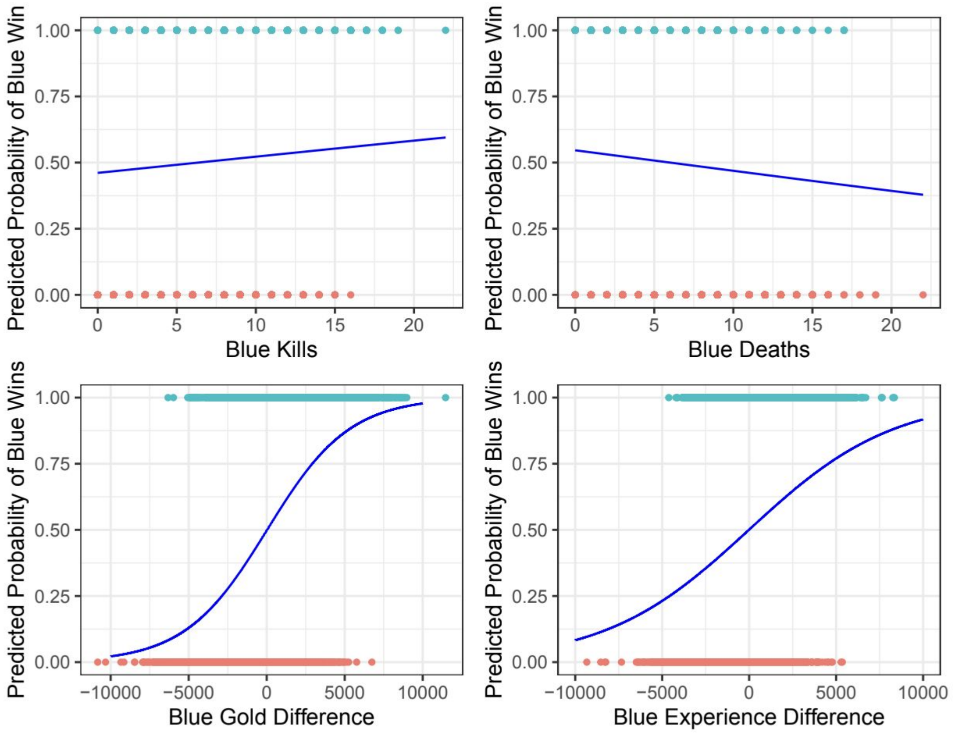
Bootstrapped (case-resampling) 90% Confidence Intervals

Table 3: Case-bootstrapped 90% confidence intervals for the coefficient estimates and average predicted probability of blueTeam win for the logistic model.

	lower	main	upper
(Intercept)	-0.071	0.037	0.15
blueKills	-0.0028	0.025	0.047
blueDeaths	-0.057	-0.031	-0.004
blueGoldDiff	0.00032	0.00038	0.00043
blueExperienceDiff	0.0002	0.00024	0.00029
Mean predicted probability of blueTeam win	0.49	0.5	0.51

Results

- Logistic Model Results: Blue team win probability predictors
- Predictors with Positive Impact:
  - Blue Kills
  - Blue Gold Difference
  - Blue Experience Difference
- Negative Impact:
  - Blue Deaths



We identified the limitations of our analysis and places for future research.

1. Large positive gold differentials
2. Large positive experience differentials
3. More kills
4. Fewer deaths

- **Impacts of findings:**
  - Helps players choose an optimal forfeit.
- **Challenge: Finding predictor values in game**
  - Further research to explore stand in values as predictors.

References

- Ma, Y. (2020, May). League of Legends Diamond Ranked Games (10 min). Kaggle. https://www.kaggle.com/datasets/bobbyscience/league-of-legends-diamond-ranked-games-10-min
- Shalizi, C. (2024, Feb). Chapter 6: The Bootstrap. http://www.stat.cmu.edu/~cshalizi/uADA/24/lectures/ch06.R
- Yurko, R. (2024, Jan). Lecture 2 demo: Building an Expected Goals Model.
- Yurko, R. (2024, Jan). Lecture 3 demo: Calibration and Cross-Validation.