What is the Optimal Team Composition of Skill Level for 5v5 Sports?

Jhet Cabigas, Jesse Coulson, Saul Mooradian, Adam Sanden

College of Engineering, and Computer Science, California State University Chico

CSCI 385: Introduction to Data Science

Professor Tillquist

December 13th, 2023

Abstract

This article explores what makes the best team compositions of player quality in esports and traditional sports, comparing Counter-Strike: Global Offensive(CS:GO) and basketball (NBA). Both are team based games with 5 players and have a multitude of metrics that contribute to the outcome of a game. Players were categorized into below average, average, and star players based on a rating metric. For CS:GO the rating metric used was from hltv.org. For basketball the rating metric used was plus minus, a statistic that represents how many net points the team the player is on scored while the player was on the court. Logistic regression was performed on a dataset that had per match team compositions and outcomes. Team compositions were represented by one categorical variable to address multicollinearity, and this variable was analyzed against the baseline of 5 average players to see their impact on winning. Some insights from the model showed that a team of 5 star players is 2.81 times more likely to win than 5 average players for basketball, but 4.92 times for CS:GO. The study also found some less intuitive but still effective team compositions. Overall, teams generally do better with higher overall skill levels, and having more below average players tend to lower the chances of winning.The insights found in the study could be applied to creating teams on a budget and how to optimally use your resources to build the best team combination possible and can also be applied to further understand why certain teams perform better than others.

Introduction

The purpose of this study was to do a comparison of player quality on outcome for Esports(CS:GO) vs Sports(basketball). CS:GO is a first person shooter video game that is round based where the first team to win 16 rounds wins the match. basketball is a game where the team with the most points scored through shooting a basketball through the other team's hoop wins. Both basketball and CS:GO are 5v5 games that rely heavily on teamwork, so it lends itself well to comparison.

Our key objective is to determine how team compositions of player skill influence the likelihood of winning. We want to answer questions about composition such as: how much does each star player increase the odds of winning? We wanted to compare CS:GO to basketball because esports and CS:GO are a relatively new and evolving field, compared to basketball which has a well established history. This comparison will give insights into how a newer esport compares to a well studied and more solved sport such as basketball. By creating a model to analyze team composition based on rating, we will provide insights that could help with strategies for team assembly and further insights into why certain teams perform better in sports and esports.

Related Works

We couldn't find any previous studies that directly addressed our specific data science question on team composition. However, in our research we did come across a study that covered a different aspect of team composition, that being the optimal distribution of team roles. In a study titled: *What Makes a Good Team? A Large-scale Study on the Effect of Team Composition in Honor of Kings,* a 2019 study conducted with the assistance of Tencent, focuses on the impact of team role distribution, and role diversity on the likelihood of winning a game of Honor of Kings. Honor of Kings is a 5v5 multiplayer online battle arena (MOBA) where each team member can select 1 of 5 pre-defined roles. Each role has its own strengths and weaknesses and you can have any number of team members pick the same role. Their study chose to focus on role diversity and its effect on team success, whereas we chose to focus on the skill of the individual team members and their effects on the outcome of a game.

Data

We used multiple datasets from Kaggle for both the NBA and CS:GO analysis. Regarding the basketball data, after joining a data set with the games and a data set with the player performances in the games and removing non-starters from the data, we had 256,104 observations where an observation was a player in a game. There was no missing data for the basketball data set.

The CS:GO data included 45,775 professional games. Unlike the basketball data set, it already included a player rating variable from hltv.org. The CS:GO data also included best of three matches. We did not include these games in our analysis due to missing information. After we created our composition variable (we will look at this variable in detail in the methodology section), we made it so each observation in our data was a game within a season. This is true for both the NBA and CS:GO data.

Methodology

To compare the effects of team composition of player skill on the outcome of a game between esports (CS:GO) and sports(basketball), we first had to explore these relationships individually. To accomplish this, we used a categorical variable with all possible team compositions as an independent variable and a binary win/loss variable as our dependent variable. We created the team composition variable by splitting all the starters into three groups based on plus-minus. These groups were created by first finding the mean plus-minus in a season. If a player's plus-minus in that season was more than one standard deviation below the mean, he was classified as below average; within one standard deviation of the mean, he was classified as average; above one standard deviation of the mean, he was a star. To be clear, the player ratings were calculated using yearly plus-minus, not career plus-minus. This was done to help mitigate the impact of career length on performance rating.

The classification of a star player is fairly subjective. This made it difficult to find literature reporting how to classify a star player in the NBA and in CS:GO. We then used these three categorical variables to create a new categorical variable with all possible team compositions. This categorical variable was used as our independent variable of interest within our model. After looking at the number of observations for each team composition, we noticed specific compositions for both the NBA and CS:GO data had very few observations. For example, a composition of four-star players and one below-average player was rare in the NBA and CS:GO. Because our dependent variable, game outcome, is recorded as a win or a loss, the assumption of normality for linear regression is seriously violated. Instead of using linear regression, we examined the relationship between team composition and game outcome using logistic regression. We set our reference group to a composition of five average players. Thus, when we compare various compositions, they are relative to our reference composition. To test the quality of our models, we used a confusion matrix to calculate accuracy. We did not look at R^2, RMSE, or MAE because they are not great metrics for assessing the quality of a logistic model.

Results

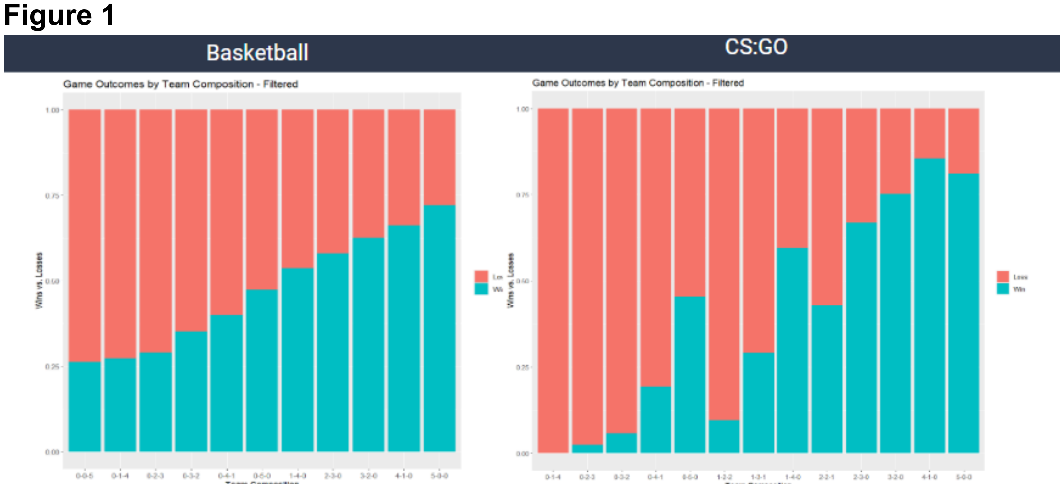
After looking at our results, we noticed a few general patterns. Compositions with more star players had higher odds of the game outcome being a win, and compositions with more below-average players had higher odds of the game outcome being a loss. This was true for both our NBA and CS:GO models, but the magnitude in which the odds changed differed greatly. In other words, the amount star players could positively impact the game outcome was greater for CS:GO. The opposite was also true; the amount below average players could negatively impact the game was greater in CS:GO (Figure 1). The model interpretations help us quantify these differences. The NBA model predicted that a team composition with two average players and three below-average players has 0.45(0.41, 0.50) times the odds of the game outcome being a win compared to a team composition of five average players and that a team composition with five star players has 2.81 (2.59, 3.06) times the odds of the game outcome being a win compared to a team composition of five average players. The CS:GO model predicted that a team composition with two average players and three below-average players has 0.04(0.01, 0.09) times the odds of the game outcome being a win compared to a team composition of five average players and a team composition with five star players has 4.92 (2.71, 9.71) times the odds of the game outcome being a win compared to a team composition of five average players.

Table 1

Key Differences		
Composition	**Basketball**	**CS:GO**
0-2-3	0.45	0.04
0-3-2	0.61	0.08
0-4-1	0.74	0.28
1-4-0	1.32	1.81
5-0-0	2.81	4.92

Our models suggest that the impact a single player can have on a game is greater in professional CS:GO relative to the NBA. Based on our analysis of data from the NBA and CS:GO, the optimal team composition for 5v5 sports is five star players.

Figure 1



Conclusion

Given our data, our findings were very linear and reasonable: Teams in both sports at a higher skill level are more likely to win than teams at lower skill levels. However, the results also demonstrate that CS:GO is a much more volatile game than basketball, as a single player's performance is much more likely to alter the course of the match. This difference in variability could be due to many different factors: Basketball is a game where players can be subbed in and out frequently, which could allow a team to minimize the losses caused by an underperforming player. In CS:GO, if a player makes a mistake in a round, they likely won't be able to play until the next round, increasing the damage caused by underperforming players as well as the effectiveness of an overperforming player. Furthermore, the round outcomes in a CS:GO match are greatly affected by momentum due to both psychological and in-game economic factors, whereas basketball is very back and forth, as the possession of the ball is swapped whenever a team scores. That said, these results and methods can be refined and further applied to other contexts and more specific performance measures in future work.

Future Work

To continue this research, the next steps are to reduce our limitations (such as sample size and composition variety) and begin to evaluate different performance measures. Furthermore, there is a wealth of potential areas beyond CS:GO and basketball to explore further in future works, the most obvious being other sports. Do the relationships remain in games with larger teams like soccer or smaller teams like Rocket League? What happens to the impact of an individual player when we scale down the basketball and CS:GO games to 2v2 matches? Given more granular data, we could also explore variables such as substitution players and coaching staff, the effects of home-field advantage, or how to maximize a team composition's skill level with a given contract budget.

Contributions

Our team consisted of four members: Jhet Cabigas, Jesse Coulson, Saul Mooradian, and Adam Sanden. We split into pairs, where Jhet and Saul worked on the basketball dataset, and Jesse and Adam worked on the CS:GO dataset. Both pairs collaborated regularly to ensure that the final versions of each dataset were organized the same and that the statistical methods employed were both consistent and statistically sound. For this report, the abstract and introduction were handled by Jesse, the discussion of related work was done by Adam, the methodology and results were created by Saul, and the conclusion, future work, and contributions sections were written by Jhet.

Works Cited

Cheng, Ziqiang et al. 2019. “What makes a good team? A Large-scale study on the effect of team composition in Honor of Kings”arxiv, https://arxiv.org/pdf/1902.06432.pdf

Machado, Dauernheimer Mateus.(2019). *CS:GO Professional Matches.* https://www.kaggle.com/datasets/mateusdmachado/csgo-professional-matches/

https://www.kaggle.com/datasets/nathanlauga/nba-games/

Lauga, Nathan (2022). *NBA games data*. https://www.kaggle.com/datasets/nathanlauga/nba-games/