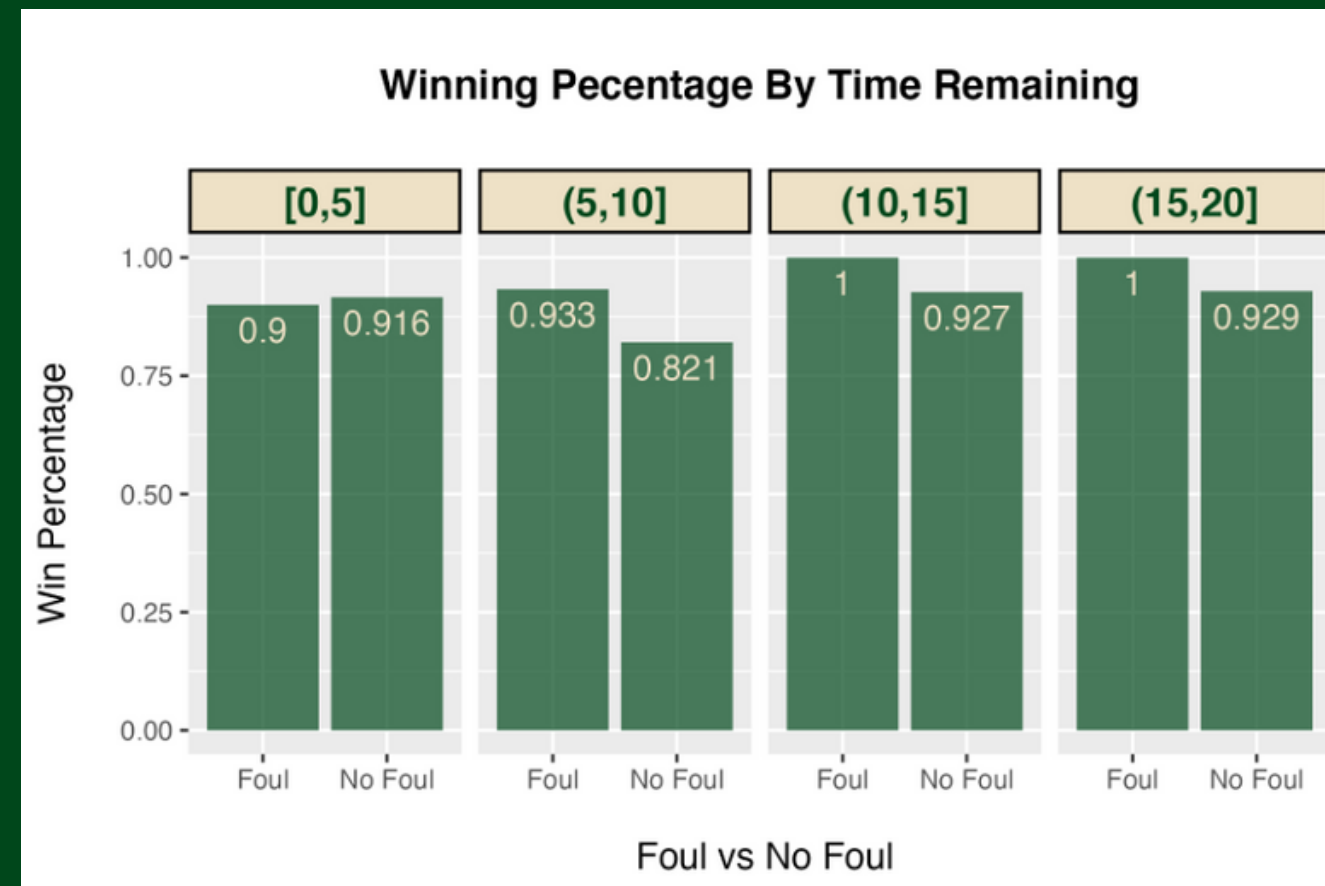


Should You Foul?

Given the situation, I would tell the coach to foul. Over the last 3 seasons, teams that are in the same situation have allowed 1.472 points per possession and have a 94.4 winning percentage when they intentionally foul. Teams that do not foul in that situation allow less points per possession (0.774), but also have a lower winning percentage (89.8%). The reason why a team may allow more points on average but have a greater winning percentage is due to the defensive team preventing the offensive team from tying the game. A tie game this late in regulation will likely lead to overtime where the opposing team has a much greater chance of winning.



In addition to the broad recommendation, I would show this visual to the coach. The visual splits the given situation into four time slots and shows what the win percentage would be based on the coach's plan of action. The chart shows that the coach should foul in every situation unless there is less than 5 seconds left. There also is not a 20-24 second bin as there were no chances in the sample that ended with more than 20 seconds left in the game.

The biggest improvement to the data provided would be getting data from more seasons. Although there were 405 chances, only 36 of those were fouls in the bonus during the 4th quarter. 36 samples should be enough for a broad recommendation but narrowing down the situation to account for the game clock or start type make it difficult to provide an accurate assessment. I also did not include any data from overtime as there are confounding variables (fatigue, fouling out, etc.) which may incorrectly affect the results. Another improvement would be the having the shooting efficiencies for both teams. If a defending team was better at making free throws than the opposing team, then they might be more inclined to foul. The same could be said if the offensive team is a great three-point shooting team.

NBA Project

I largely covered the basics of PeFG%, but I’ve put the link below to an article I wrote about the project. Although I’m proud of my modeling abilities, this project was completed in January, so my skills have improved since then.

https://medium.com/@sam_auer12/predicted-efg-and-what-it-says-about-the-nba-44592879349b

One problem that I would love to solve would be creating a statistic to measure expected production or efficiency. Most professional sports have statistics that are used to estimate what a player’s production SHOULD be (e.g. xG in Soccer, xERA in Baseball, CPOE in Football). Basketball, unfortunately, does not have a similar statistic. Using the NBA’s public shot tracking data, I was able to create PeFG% (predicted effective field goal percentage). The statistic would be used similarly to xG in that players who outperform PeFG% are either exceptional scorers or are due for regression back to the mean (I also created predicted effective field goal percentage over expected (PeFGPOE) to measure this).

Top 10 eFGPOE			
PLAYER_NAME	eFG	PeFG	eFGPOE
Nikola Jokic	0.6197559	0.5264999	9.325597
Seth Curry	0.5940460	0.5137916	8.025436
Deandre Ayton	0.6391679	0.5614928	7.767506
Kevin Durant	0.5699552	0.5000155	6.993969
Chris Paul	0.5360054	0.4725805	6.342496
Jrue Holiday	0.5702306	0.5180854	5.214524
Jalen Brunson	0.5486111	0.4983880	5.022308
Tyrese Haliburton	0.5616062	0.5199814	4.162476
Mikal Bridges	0.6008159	0.5638958	3.692006
Jonas Valanciunas	0.5740938	0.5375125	3.658130

I created the statistic using an XGBoost model. I used the x and y coordinates of the shot, the shot type, and the shot’s point value to predict the probability of a made shot. I then multiplied the probability of a made shot to the point value of the shot attempt which became PeFG%. My statistic provided solid results; players who were viewed as efficient scorers ranked towards the top in the league (2022 leaders above). However, there was a major flaw in the stat. The NBA’s public shot tracking data does not have defender data. This means a wide-open corner 3 would have the same PeFG% as a double-teamed corner 3. Using an NBA team’s database, this flaw can easily be corrected. Adding a variable such as the distance of the nearest defender to the shooter would be a sufficient improvement.

This project/statistic has a lot of versatility. It can be used to inform our players which shot type and in which location they are most or least efficient compared to league average. For example: we can inform Khris Middleton that he should shoot more turnaround jump shots from the right elbow because he is much more efficient than league average. This stat can also be used to help the front office target players in a trade. For example: the Bucks should look to trade for Saddiq Bey because he is performing well above his PeFG% over a large sample size. The final way I can see this stat being used is for scouting upcoming opponents. For example: we should allow Walker Kessler to get the ball in the 4th quarter because tends to take poor PeFG% shots late in the game.