

## NBA FINALS PREVIEW

Within this project I have created a few data tables to rank the remaining 4 NBA teams in the playoffs by a variety of statistics. I will note that at the time I began this project the finals matchup had not been decided yet, so I decided to keep the 4 semifinalists in for the duration of the project to keep things consistent. Now we know the matchup is between the Lakers and Heat, so we will be primarily focused on those two teams.

To start, the purpose of this small project was to use existing statistics, such as effective field goal percentage, and manufactured statistics, like overall and predicted win percentage, to rank the teams and give a preview of what we could expect. I will start by saying that these statistics are not comprehensive or thoroughly researched enough and will likely not predict the game, so please don't use these to bet because then I'll feel guilty for giving you faulty statistics.

Basically, instead of giving a finals preview based solely on my opinion, I decided to try to use some data to predict who I think the winner might be. The 3 main statistics I used and/or created are:

- Effective Field Goal Percentage (eFG%)
- Overall (based on eFG%, Turnovers, Total Rebounds, and Free Throw Percentage)
- Predicted Win Percentage

I used SAS to import per game season and playoff data from [https://www.basketball-reference.com/leagues/NBA\\_2020.html#all\\_team-stats-per\\_game](https://www.basketball-reference.com/leagues/NBA_2020.html#all_team-stats-per_game) and [https://stats.nba.com/teams/traditional/?sort=TEAM\\_NAME&dir=1&SeasonType=Playoffs&Season=2019-20](https://stats.nba.com/teams/traditional/?sort=TEAM_NAME&dir=1&SeasonType=Playoffs&Season=2019-20)

First, I created a SAS data table from just the basketball-reference data and made formulas to calculate and eFG% column. Effective Field Goal Percentage is an adjusted FG% that gives 3 pointers extra weight. The equation is  $(2\text{pt FGM} + 1.5 \times 3\text{pt FGM}) / \text{FGA}$  because 3 pointers are 1.5 times greater than 2 pointers. This statistic adds extra weight to the 3 ball (which is a huge part of today's game) compared to generic FG%. I created the eFG% column using this formula and then printed a data table to rank the Heat, Lakers, Nuggets, and Celtics by eFG%.

### Ranking the Remaining NBA Teams by Effective Field Goal Percentage

Rank	Team	OVR	eFGpct
1	Miami Heat	5.491	0.546
2	Los Angeles Lakers	5.666	0.541
3	Denver Nuggets	5.701	0.535
4	Boston Celtics	6.103	0.531

The difference between eFG% for all 4 teams is negligible, so I cannot make a judgement based on this statistic. However, it is important to note that, amongst all teams during the regular season, the Heat were ranked 3<sup>rd</sup> and the Lakers were ranked 5<sup>th</sup>, while the Nuggets and Celtics, who both missed the finals, were 9<sup>th</sup> and 17<sup>th</sup> respectively. So it is possible that eFG% has some effect, but cannot be an indicator alone. I also did not use playoff data for this analysis, because I wanted a sufficient sample of games, so that might yield different and more current results, albeit less reliable.

I also created an overall rating based on an article I read about the most predictive basketball stats for wins. Here is the link for reference:

<https://www.breakthroughbasketball.com/stats/effective-field-goal-percentage.html>

According to this article, Dean Oliver, who wrote “Four Factors of Basketball Success,” claims that the 4 factors and their importance as a percentage weight are shooting (40%), turnovers (25%), rebounding (20%), and free throws (15%). To simplify things in the interest of time, I chose what I believe to be the most important or comprehensive stat within each category.

This is my equation:  $.2*TRB - .25*TOV + .15*FT\_ + .4*eFGpct = OVR$

TRB=total rebounds

TOV=turnovers

FT\_=Free throw percentage

eFGpct=eFG%

And here are the rankings:

**Ranking the Remaining NBA Teams by Overall Score (Based On Shooting, Rebounding, Turnovers, Free Throws)**

Rank	Team	OVR	eFGpct
1	Boston Celtics	6.103	0.531
2	Denver Nuggets	5.701	0.535
3	Los Angeles Lakers	5.666	0.541
4	Miami Heat	5.491	0.546

Curiously, the 2 teams that were eliminated have higher overall rankings, so maybe my system is not predictive (probably because it only includes 4 statistics). To try to resolve this, I decided to do more research and use a model created by a data scientist from over 2 decades of NBA season per game data.

Robert Alterman, a data scientist from the University of Michigan, developed an equation for predicted win percentage using a variety of statistics. Here is the link and equation that he developed: <https://towardsdatascience.com/predicting-nba-win-percentage-84148ae8d3e6>

**WIN % = 0.5000 – 0.0330ln(PTS) + 0.0587(FTM) + 0.0186ln(OREB) + 0.0543ln(DREB) + 0.0376ln(AST) – 0.0480(TOV) + 0.0408(STL) + 0.0186ln(BLK) – 0.0639ln(BLKA) – 0.0107(PF)**

**Points (PTS), Free Throws Made (FTM), Offensive Rebounds (OREB), Defensive Rebounds (DREB), Assists (AST), Turnovers (TOV), Steals (STL), Blocks (BLK), Blocked Field Goal Attempts (BLKA), and Personal Fouls (PF)**

I did have to adjust the equation because this equation was giving me win rates about 100%, so I removed the 0.5 at the beginning of the equation. This adjustment ended up helping because the predicted win percentage seems to correlate nicely with the actual win rates for most teams in the league this season.

To implement my adjusted equation, I merged the basketball reference regular season data with regular season data from stats.nba.com to include BLKA (because my original data set did not contain this stat). BLKA is block against, which measures the number of field goals a team gets blocked by an opponent per game. I created a column using equation above and ranked the 4 teams according to their predicted win percentage, hoping that this, being a more comprehensive statistic based on years of data, would be a good predictor for who is more likely to win. I did the same thing for strictly playoff data, and ranked the teams according to regular season and playoff data.

Here are the results:

#### Ranking the Remaining NBA Teams by Predicted Win Percentage from Regular Season Data

Obs	Team	PRED_WIN_PCT
1	Miami Heat	0.66547
2	Boston Celtics	0.65950
3	Los Angeles Lakers	0.59308
4	Denver Nuggets	0.53421

#### Ranking the Remaining NBA Teams by Predicted Win Percentage from Playoff Data

Obs	TEAM	PRED_WIN_PCT
1	Miami Heat	0.90490
2	Boston Celtics	0.65684
3	Los Angeles Lakers	0.62508
4	Denver Nuggets	0.50191

Using both the regular season and playoff data, the rankings are the same. The Miami heat have the highest predicted wins based on this model, followed closely by the Celtics in the regular season model. The Lakers fall behind the Heat by about 6% in the playoff model but the playoff model displays a massive margin between the Heat and the rest. It appears that the Lakers and Heat are very comparable using the regular season data, but in the playoffs the heat are projected to produce a 90% win rate, which is far above that of the Lakers and any other

team in this data set. Judging by regular season data, I can make no conclusion about which team is more likely to win any given game, but if we look at purely playoff data, the Heat are definitely rolling. Irrespective of opponent, the Heat are much more likely than the Lakers to win the series. Unfortunately, the opponent matters so this model does not adequately take into account the matchup between the 2 teams. One thing these results do show us, however, is that Miami is performing significantly better in the areas that this model favors. In other words, Miami is supposed to produce a lot of wins given their playoff performance. Using this statistic, the Heat should win the Finals.

Obviously, there is a lot more at play here, including depth, star power, possible injuries, luck, etc., but this win percentage statistic, at least when applied to this year's playoffs, appears to foreshadow a Heat victory if they can keep up their current pace against a powerful Lakers team led by 2 of the best players in the game.

Here is the link to my github repository with the SAS code if anyone is interested in more of the specifics. I am available to explain it and take suggestions/criticism (since I know some of the statistical methods are not sound):

<https://github.com/krodonnell7/NBA-Finals-2020>