NBA Data Analysis : Assessing the current regular season standings

Introduction:

The National Basketball Association (NBA) is the most popular professional basketball league globally. Based in North America, the NBA consists of 30 teams spread equally into two conferences: East and West. From October until April, 82 games are played between teams regardless of their conference during the regular season. Once concluded, seedings are formed using the standings and dictate the initial intra-conference matchups. These take place during the latter portion of the season: the NBA playoffs.

With the current 2024-25 regular season slowly concluding, there have been dramatic comeback stories, rising young lineups, and a wide variety of elite individual and team play. However, despite all the thrilling and chaotic moments, a select few teams have maintained constant control at the apex, namely the Cleveland Cavaliers and Oklahoma City Thunder. This analysis will evaluate a range of teams based on wins using various NBA metrics to explain why some teams have achieved so much and others so little.

Data and Modelling:

Python was the primary tool used for this analysis. Data was collected from an NBA API on GitHub and multiple endpoints were utilised to collect the relevant NBA metrics. The initial goal was to deduce which metrics would be most worthwhile to explore. To determine this, a handful of correlation tests were run to establish any correlations of statistical significance. These tests used a sample size of 150 data points from the 2020-21 season until the 2024-25 season.

The NBA has 5 traditional metrics: points, assists, rebounds, blocks and steals. The Spearman's rank-order correlation coefficient test was used as each metric shared a monotonic relationship with the number of wins by a team. While each metric was of statistical significance at $p < 0.05$, points showed the strongest statistically significant correlation with a coefficient of 0.49. Assists, rebounds, blocks and steals had the following coefficients: 0.35, 0.33, 0.21 and 0.16, respectively. Further correlation tests were run with advanced point metrics to see if any stronger correlations could be identified.

Advanced metric	Correlation coefficient	p-value
Three-point field goal attempts (FG3A)	0.65	$p < 0.001$
Two-point field goal attempts (FG2A)	0.54	$p < 0.001$
Free throw attempts (FTA)	0.19	$p = 0.022$
Three-point field goal percentage (FG3 pct)	0.27	$p < 0.001$
Two-point field goal percentage (FG2 pct)	0.08	$p = 0.31$
Free throw percentage (FT pct)	0.20	$p < 0.05$
Effective field goal percentage (EFG pct)	0.67	$p < 0.001$

Once correlations were established, I sought to find which combination of metrics could be used as the best predictor of a team's number of wins during the regular season. To achieve this a multiple linear regression model was created.

A careful selection of advanced point metrics was combined with traditional metrics, totalling 9 independent variables, to build the model. The former included: FG3A, FG2A, FTA, EFG pct and FT pct. This selection accounts for all three types of shots a team can take and their percentages.

The motivation behind using EFG pct over FG2 pct and FG3 pct is that EFG pct is an advanced metric with a relationship to both. Thus, to reduce the risk of multicollinearity and the inflation of the model's r2 score, all three could not be used simultaneously. Additionally, EFG pct also accounts for the fact that a three-point shot is worth 1.5x the value of a two-point shot. As a result, a high FG3 pct will be weighed more heavily than a high FG2 pct. By being more reflective of a team's scoring ability, and therefore chances of winning, EFG pct was a more appropriate variable to use.

$$eFG\% = \frac{FG + 0.5 * 3P}{FGA}$$

Fig 1.1: Equation for effective field goal percentage, where: "FG" means field goals, "3P" means three-point shots taken, and "FGA" means field goal attempts.

EFG pct is a measure that does not consider the accuracy of free throw shooting, which is why FT pct was included in the model. The previously discussed core metrics, excluding points, were also incorporated because they improved the model's r2 score. Ultimately, the regression model achieved an r2 score of 0.78 using training data (70% of the dataset) and 0.67 using testing data (30% of the dataset).

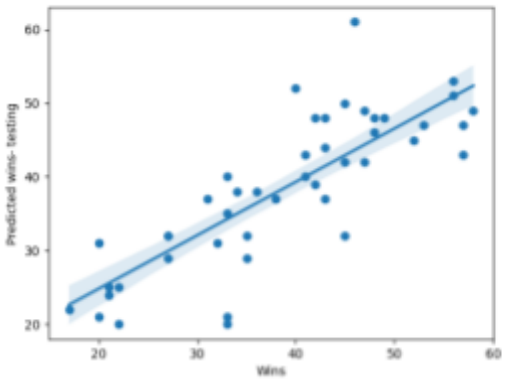

Figure 2.1: Multiple linear regression using test dataset.

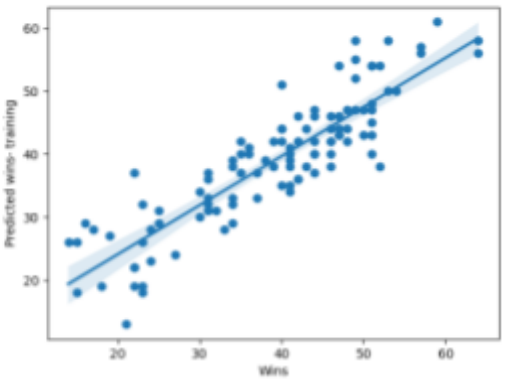

Figure 2.2: Multiple linear regression using training dataset.

The difference of 0.11 between the two r2 scores indicates ample room for improvement in the model's ability to generalise when presented with unseen data. Residual plots were generated for both the testing and training datasets to assess predictions for signs of overfitting or underfitting, neither of which were found.

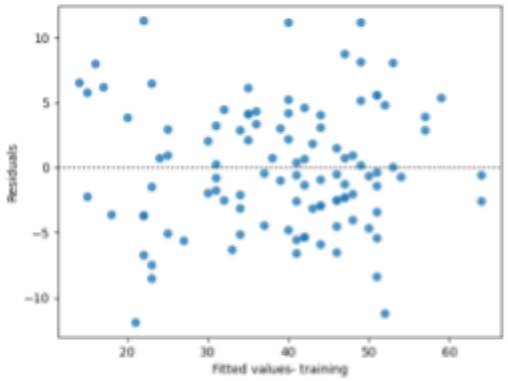

Figure 3.1: Residual plot using fitted values from training data.

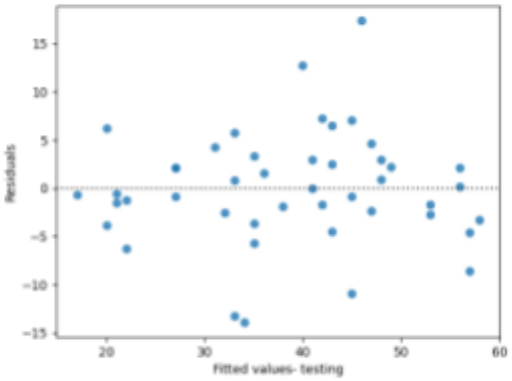

Figure 3.2: Residual plot using fitted values from testing data.

Even so, the r2 scores suggest that the independent variables are relatively effective in explaining the variance in wins achieved by teams.

The model's development primarily utilised offensive metrics which suggests that a team's offence, or lack thereof, may have more impact on their ability to win compared to other

factors. The training dataset recorded a root mean squared error (RMSE) of 5.38 while the testing dataset's RMSE was 6.54. Both datasets shared a maximum error of 15. On average, the model's predictions for wins tend to be off by 5 to 6, with a maximum discrepancy of 15. This indicates that other variables beyond point scoring must contribute to team success.

Power Bi was employed to visualise the metrics for the current 2024-25 season. The offensive metrics used in the regression model were displayed to identify outliers- instances where a team's offensive capability did not reflect their number of wins. Additionally, defensive metrics, such as estimated defensive rating and defensive field goal percentage, were also studied.

$$TDR = 100\frac{PTS_{Opp}}{Possessions_{Team}}$$

Figure 4.1: Equation for team defensive rating.

Team defensive rating is a metric that represents the points allowed by a team every 100 possessions. Defensive field goal percentage is simply the percentage of field goals made by the opposition. In both cases, a lower number reflects a better defensive performance.. The exploration of defensive aspects provides an opportunity to address potential reasons behind the regression model's limitations in predicting total regular season wins.
.

Results:

As of 23/03/25, the Oklahoma City Thunder are leading the league in wins at 58 after recently separating from the Cleveland Cavaliers, who are trailing behind at 56 wins. At the bottom of the league are the Washington Wizards, who have 15 wins, barely below the 16 wins of the Utah Jazz. The range of the current standings is a massive 43 wins, and I hope to explore some of the reasons why.

The approach I have taken in this analysis is to segment the league based on wins and in groups of 10. In each group, teams with distinctly offensive characteristics relative to the rest of the teams in their win category will be investigated further. In doing so, I hope to gain a better appreciation for why teams are at their current number of wins and find hints leading to unexplored areas where further analysis can be performed. To start, I will explore the strongest teams in the league at the 50+ wins mark.

50+ wins:

At the top of the league are three teams: Boston Celtics (51 wins), Cleveland Cavaliers (56 wins), and Oklahoma City Thunder (58 wins). All three teams exhibit impressive shooting accuracy across all types of shots (free throws, three-pointers, and two-pointers). Notably, the Thunder leads the league in free throw percentage at 82%, while the Cavaliers boast the highest effective field goal (EFG) percentage at 58%. Although these percentages evaluate different types of shots, examining their shooting efficiency strongly points to why they have been so successful this season. The Celtics, while not leading in either category, rank among the top six teams for both shooting percentage metrics. However, the true strength of the Celtics lies in the volume of three-point shots they take.

The Celtics have taken 3,249 three-point shots this season, leading the league in this statistic and exceeding the league average by a remarkable 636 shots. What's more, the Celtics are the only team in the entire league that has taken a larger volume of three-point shots than two-point shots. The 1.5 times increased value of the three-point shot over the two-point shot, along with their impressive shooting efficiency for all three shot types, are strong reasons for their success in the league.

Defensively, OKC excels. They lead the league with the lowest estimated defensive rating of 104.4 and a defensive field goal percentage of 44%. Through the effective limitation of opposition team shooting, they have allowed 285 fewer field goals compared to the league average. They also rank first in steals and fourth in blocks. Both sides of the floor considered, OKC demonstrates that they are a powerhouse team and their numbers solidify their position as the number one team in the league.

While OKC is somewhat in a league of its own defensively, the Celtics and Cavaliers are still more than able to hold their own. Both teams concede fewer field goals compared to the league average. They are among the best 5 teams for estimated defensive rating and ranked 8th for defensive field goal percentage. This combination of a strong offence and effective defence is a reasonable explanation for their positions within the 50+ win bracket.

40-49 wins:

There are a total of ten teams in the 40-49 win range, making it the densest category in this analysis. Teams include: Houston Rockets (46), Golden State Warriors (41), Milwaukee Bucks (40), Los Angeles Lakers (43), Memphis Grizzlies (43), LA Clippers (40), Minnesota Timberwolves (41), New York Knicks (44), Indiana Pacers (41) and Denver Nuggets (44). Many of these teams display strong shooting efficiencies but the Rockets and Warriors show comparatively less success in this area.

In fact, the Rockets have the worst free throw percentage in the entire league and rank among the bottom 6 teams for EFG percentage. Given their current record of 46 wins, this finding is quite surprising and possibly explains some of the variability found in the regression model. However, their strengths become more clear when examining their defensive statistics.

The Rockets rank as one of the top three teams in estimated defensive rating and are among the top four teams in defensive field goal percentage. Compared to the league

average, they allow 77 fewer field goals, record 8 more blocks, and achieve 22 more steals. This amalgamates to a team that effectively contests shots, dominates the paint, and intelligently reads schemes made by opposing offences. With these defensive edges, their poorer shooting efficiency is not significant enough to impede their success on the court, providing more insight into their 46 wins.

The Warriors are a team that also struggles with low shooting efficiency. For a team historically known for their ability to shoot three-point shots, their rank in the bottom 10 of the league for EFG percentage is surprising. They are also among the bottom 4 teams for free throw percentage. Unlike the Rockets, their defensive numbers don't provide as strong an argument for their ability to win, although a holistic view provides more clarity when ascribing reasoning to their record.

The Warriors concede fewer field goals than the average team. This suggests that, on average, their games typically end with lower scores for opposing teams. As a result, fewer points would be needed to ensure a win. However, teams like the Clippers and Lakers concede even fewer field goals than the Warriors which reduces the likelihood of this explaining their ability to win despite their poor shooting efficiency.

However, the key difference between the Warriors and these teams is their ability to pass, rebound, block shots, and steal the ball to gain extra possessions. Unlike the Clippers and Lakers, the Warriors are above league average in all these core metrics. A more well-rounded, foundational style of play presents a stronger case for how the Warriors earned their position in this bracket.

30-39 wins:

Slightly less populated at nine teams is the 30-39 win range. Teams include: Phoenix Suns (34), Chicago Bulls (31), Sacramento Kings (35), Detroit Pistons (39), Dallas Mavericks (34), San Antonio Spurs (30), Atlanta Hawks (34), Portland Trail Blazers (32) and Orlando Magic (33). In this group, and for different reasons, the Chicago Bulls and Orlando Magic stand out.

The Magic are one of the two worst teams in the league in the context of EFG percentage. As a metric that reflects the ability to make both two-point and three-point shots, this is deeply concerning. Yet, they have still managed to achieve 31 wins. While still a record below the .500 mark, such a clear area for improvement indicates a lot of potential. However, like the Rockets, the Magic's strength primarily lies in their defence.

They are among the top 3 in the league for estimated defensive rating, concede the second least field goals (behind OKC), lead in blocks, and are above the league average in steals. These numbers cement the Magic as a defensive juggernaut in the league and have almost certainly played a significant role in their ability to win 31 games.

The Magic and Rockets are teams with evident defensive strengths that lack offensive power. This raises the question of how a 16-win opening has been created between the two, with the Rockets leading. A possible explanation could be attributed to shooting volume. While both teams share a preference for the two-point shot, the Magic shoot this type 41

times less and the Rockets 432 more compared to the league average. This difference may be a result of the number of possessions and scoring opportunities each team generates. These are metrics that may be worth exploring in more detail.

The Bulls, on the other hand, have had no shortage of three-point shot attempts this season. They currently have the second most shot attempts of this type in the league, just behind the Celtics, 364 above the league average. With an EFG percentage of 55%, also just short of the Celtics, it is quite strange that they find themselves 20 wins behind. However, through the examination of the Bulls' defensive performance, their situation can be better understood.

Although the Bulls rank below league average in blocks and steals, the biggest culprit is likely their inability to prevent their opposition from scoring field goals. At 231 more field goals conceded compared to the league average, the Bulls rank as the worst team in this metric. Collectively, these statistics suggest that the Bulls engage in high-scoring games but end up on the losing side more often. It would be interesting to look at the average scores of games that the Bulls participate in, compared to the league average, to confirm this suspicion.

20-29 wins:

Among the teams in the 20-29 win range are the Brooklyn Nets (23), Toronto Raptors (24), Philadelphia 76ers (23), and the Miami Heat (29). The only team that stands out in this bracket are the Toronto Raptors due to their high volume of two-point shots. They rank fourth in the league in this metric, characterising an identity with a focus on controlling and scoring within the paint. This belief is supported by their above-league-average rebounding, the only team in this win range to achieve this. However, their poor shooting efficiency is a significant issue. They rank among the bottom two teams in the league for free throw percentage and also rank in the bottom eight for EFG percentage.

The Raptors appear to be average defensively. While they allow fewer field goals relative to the league average, their estimated defensive rating and defensive field goal percentage are uninspiring. They also average fewer blocks than the league average, suggesting the absence of an effective presence around the rim. Additionally, their steals across the season number below the league average which may result in a reduced ability to create extra possessions. Overall, the Raptors possess an average defence with a scoring strategy centred around the two-point shot. It would be interesting to assess their two-point shooting efficiency independently to determine how effective the Raptors are at executing this strategy.

10-19 wins:

In the lowest win range of 10-19, just 4 teams remain. These are: Charlotte Hornets (18), Washington Wizards (15), New Orleans Pelicans (19) and Utah Jazz (16). None of the teams in this bracket exhibit unexpected offensive or defensive numbers. Each team has a

considerably low EFG percentage, with the Jazz slightly surpassing the rest. This could be attributed to their higher volume of three-point shots in addition to the possibility of having made more of them. To verify this, a comparison of made three-point shot percentages should be conducted.

In terms of estimated defensive rating and defensive field goal percentage, there are no real upsides for these teams either. However, there are some silver linings. Surpassing league averages are the Pelicans in blocks and steals, and the Wizards in blocks. The Hornets have also conceded less field goals than the league average. Nevertheless, none of these findings, in isolation, indicate that any of these teams make scoring for the opposition difficult on each possession. Without a threatening offence or robust defence, by NBA standards, these teams have naturally found themselves at the bottom of the league.

Conclusion:

The best teams in the league are well-rounded, dominating with both a strong offence and defence. The teams following suit, with respectable records of their own, generally have a solid offence and defence marginally weaker than the leaders. However, there may be exceptions such as the Rockets with a defense that carries a higher proportion of the load. This highlights the versatility of the league and the significant role a strong defence can play in a team's ability to win. Going further down the standings, teams weaken on both sides of the floor and no bracket exemplifies this better than the 10-19 win range.

Furthermore, based on the different teams explored in the results section, it is clear that for more confident deductions to be made, a wider variety of metrics must be studied. For instance, the average points scored in a game. This metric could explain how a team could have high-volume shooting, and decent shooting efficiency, yet still have a poor win record. Other metrics include independent two-point and three-point shot percentages. Paired with each shot type's volume, a more comprehensive picture can be made around teams prioritising a certain shot type.

Limitations:

The data used to perform the correlation tests and create the regression model used five seasons of data. The absence of additional data could have resulted in the deviation of r2 scores between the training and testing datasets. This need for more data is compounded by the inclusion of nine independent variables into the model.

In the teams that presented as offensive anomalies in their win range, a narrow selection of defensive metrics was used to speculate their defensive ability. Logically, defence must play a role in a game's outcome as a stronger defence will reduce the opposing team's ability to score. Nonetheless, correlations between defensive metrics and team wins should be explored to determine a relationship unaffected by a team's offensive output. Further, defensive metrics could be used to predict a team's number of wins independently, then

alongside offensive metrics. Such steps would provide more confidence in the degree of value defence plays in wins throughout the regular season.