Aaron Sanders, Drew Kearny, Matt Anderson Group 10 STAT 654

The Winning Formula: Discovering NBA Strategies that Drive Success Through Elastic Net Regression

1. Introduction

1.1 Motivation

The quest to identify the key drivers of success in professional basketball has long been a central challenge for team executives, coaches, and analysts. The modern NBA generates vast amounts of performance data, yet determining which factors truly drive winning rather than merely correlating with it is something that remains difficult. This challenge is compounded by the high degree of multicollinearity between basketball metrics as many statistics that appear to predict success are themselves interrelated, making it difficult to isolate their independent effects. Traditional statistical approaches like Ordinary Least Squares (OLS) regression struggle with multicollinearity, often producing unstable coefficient estimates that can lead to misleading conclusions. As NBA teams increasingly rely on data-driven decision-making to guide strategic choices about roster construction, player development, and in-game tactics, there is a clear need for more robust analytical approaches that can disentangle the complex relationships between basketball metrics and team success.

1.2 Problem Statement

This research addresses the following question: Which basketball strategies and team characteristics most strongly predict winning percentage in the NBA when accounting for the relationships among performance metrics? More specifically, we want to understand which offensive and defensive factors have the strongest independent relationships with team success, how much impact (in terms of wins added or lost) can be attributed to strategic choices like shot selection, pace of play, and rebounding emphasis, and how well these factors can predict team performance in future seasons.

1.3 Project Objectives

The primary objectives of this research are to develop an interpretable predictive model for NBA team success that accounts for multicollinearity among basketball metrics, quantify the independent impact of various strategic factors on win percentage, identify the most influential success factors that teams can focus on to improve performance, validate the model by testing its predictive accuracy on recent season data, and show how elastic net regression helps address multicollinearity challenges in sports analytics. By achieving these objectives, this research aims to provide NBA teams with actionable insights into which factors they should prioritize to maximize their chances of success, while also advancing the methodological toolkit available to sports analysts facing similar challenges in other contexts.

2. Data Description and Preprocessing

2.1 Data Sources

This study utilizes advanced team statistics from Basketball Reference with the initial data spanning five NBA seasons from 2019-2020 through 2023-2024. The dataset includes regular season performance metrics for all 30 NBA teams across each season, resulting in 150 team-season observations. This timeframe was selected to capture recent NBA trends while providing enough data for robust statistical analysis. The data was collected using a custom Python function that reads and processes Basketball Reference CSV files, combining data across multiple seasons into a unified dataset. Each observation represents a team's performance over a complete regular season, with various offensive and defensive metrics as potential predictors of win percentage.

2.2 Variables Collected

The initial dataset included team information (team name, season), performance outcomes (wins, losses, win percentage), team characteristics (average age of players weighted by minutes played), pace indicators (possessions per 48 minutes), shooting metrics (three-point attempt rate, free throw rate, true shooting percentage, effective field goal percentage), ball possession metrics (offensive turnover percentage, offensive rebound percentage), free throw metrics (free throws per field goal attempt), and defensive metrics (opponent effective field goal percentage, defensive turnover percentage forced, defensive rebound percentage, opponent free throws per field goal attempt, defensive rating). An initial variable was also added measuring the points per game of each team's highest scoring player to measure the value of having a super star. The target variable for our predictive model was Win percentage (win_pct), calculated as the ratio of wins to total games played (W / (W + L)). In Figure 1 below, we provide a sample of the initial dataset showing the range of metrics collected for analysis.

Figure 1: *Sample of Initial Dataset with Basic and Advanced NBA Team Statistics*

Age	Pace	FTr	3PAr	TS%	eFG_off	TOV_off	ORB_off	FT_FGA_off	eFG_def	TOV_def	DRB_def	FT_FGA_def	DRtg	star_ppg	win_pct
24.8	100.0	0.220	0.419	0.593	0.560	10.3	24.2	0.180	0.513	14.9	74.6	0.211	107.5	32.7	0.829268
28.9	95.7	0.212	0.536	0.591	0.561	10.8	25.7	0.169	0.522	11.6	76.0	0.154	111.1	26.8	0.743902
26.6	99.8	0.241	0.457	0.607	0.578	11.6	25.9	0.187	0.528	12.6	74.8	0.181	112.2	24.0	0.780488
27.2	97.3	0.249	0.455	0.588	0.554	13.0	25.8	0.196	0.532	13.2	75.1	0.178	111.5	27.6	0.597561
29.7	97.5	0.251	0.387	0.589	0.554	13.4	24.4	0.200	0.536	13.7	77.5	0.189	110.3	22.8	0.609756

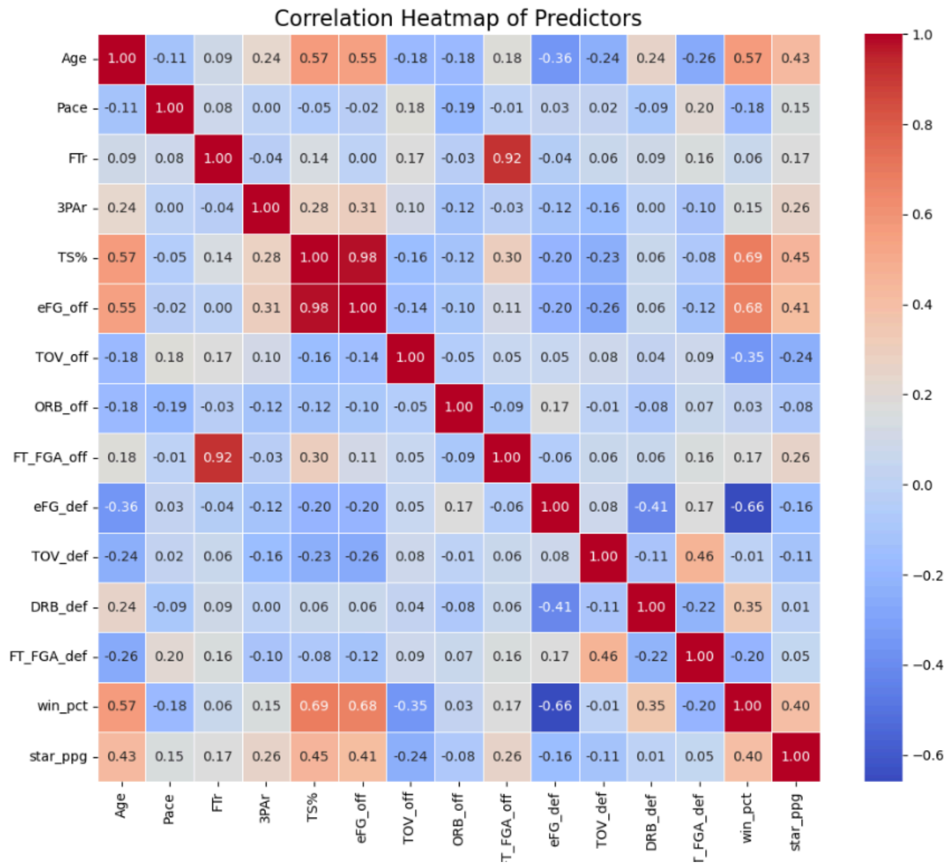
2.3 Preprocessing Steps

The data preprocessing involved several steps to ensure data quality and prepare it for modeling. First, we cleaned team names by removing asterisks (playoff indicators). We then handled missing values by dropping rows with incomplete data to ensure complete observations. Feature renaming was performed to clarify column names and distinguish between offensive and defensive metrics. We also conducted data type conversion to ensure all numeric columns were properly formatted and removed any duplicate entries to prevent data redundancy. Finally, we standardized all predictor variables to have mean=0 and standard deviation=1 to ensure fair comparison of coefficient magnitudes.

2.4 Exploratory Data Analysis

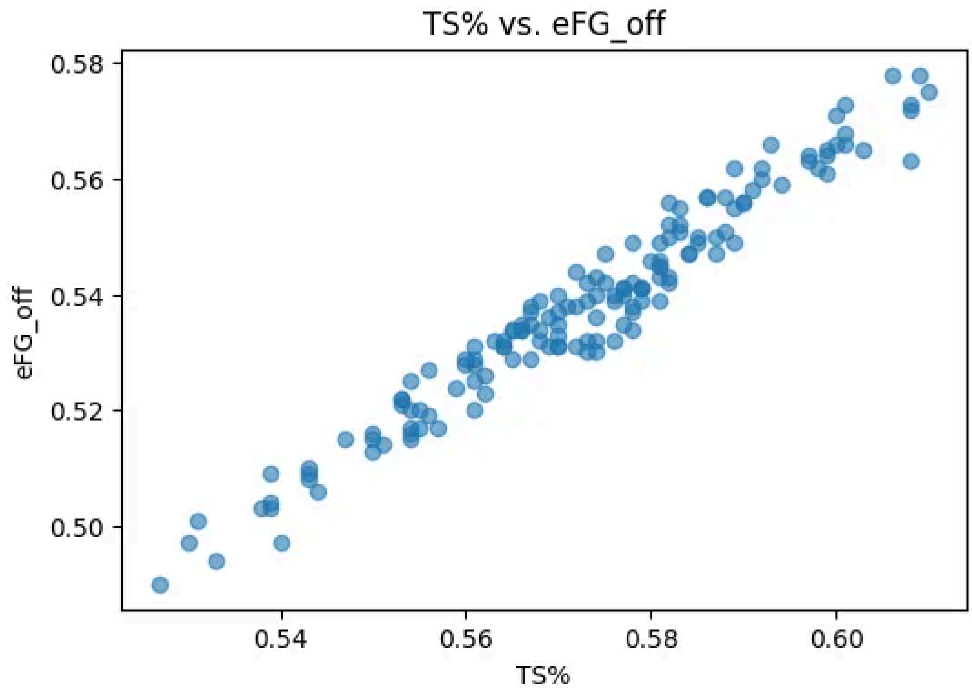
Initial exploratory analysis revealed several important patterns in the data. The correlation heatmap (Figure 2) highlighted significant relationships among several predictors.

Figure 2: *Correlation Heatmap of Offensive/Defensive Metrics and Win Percentage: Note the strong correlations between shooting efficiency metrics (TS% and eFG_off), free throw metrics (FTr and FT_FGA_off), and several significant relationships with win percentage.*



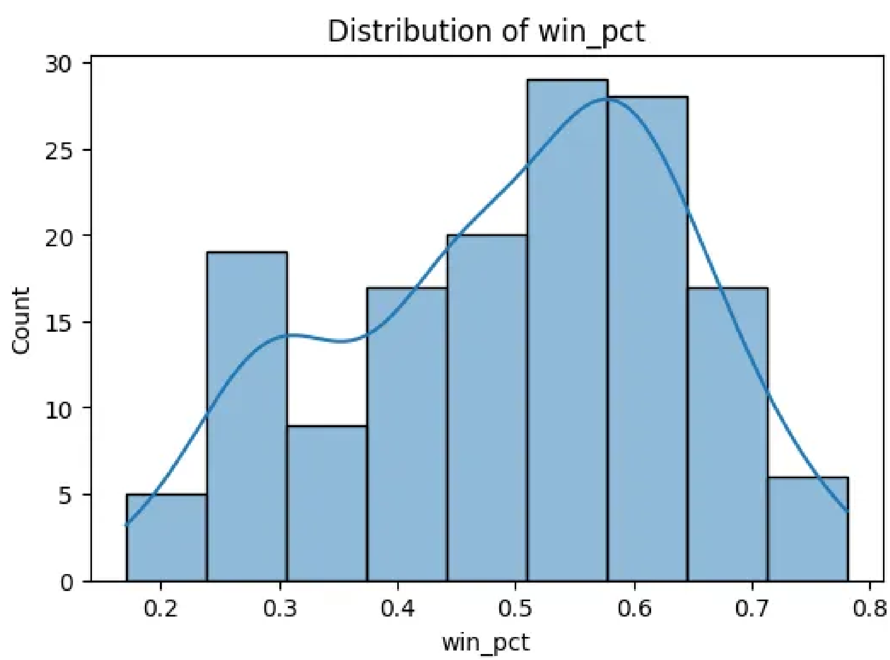
True shooting percentage (TS%) and effective field goal percentage (eFG_off) showed extremely high correlation (r > 0.9), as demonstrated in Figure 3, while free throw rate (FTr) and free throws per field goal attempt (FT_FGA_off) were nearly perfectly correlated. Several offensive metrics displayed moderate to strong correlations with their defensive counterparts.

Figure 3: *Correlation Between True Shooting Percentage and Effective Field Goal Percentage*



Examination of the distribution of win percentages across the dataset showed an approximately normal distribution centered around 0.5, as expected given the zero-sum nature of NBA competition. As shown in Figure 4, the distribution of team win percentages follows this expected pattern.

Figure 4: *Distribution of Team Win Percentages*



Scatter plots between individual predictors and win percentage revealed promising linear relationships for several variables, particularly defensive rebounding percentage (DRB_def) and offensive turnover percentage (TOV_off), suggesting their potential importance in predicting team success.

2.5 Feature Selection Considerations

Based on the exploratory analysis and domain knowledge, we made initial feature selections to address correlation issues before modeling. We removed TS% due to high correlation with eFG_off and removed FT_FGA_off due to redundancy with FTr. Despite weak correlation with win percentage, we kept Pace to test its impact when controlling for other factors. We also kept Age to explore whether team experience independently affects success. These preliminary selections were later refined through the elastic net modeling process, which incorporates automated feature selection through regularization.

3. Methodology

3.1 The Challenge of Correlated Variables in Basketball Analytics

A fundamental challenge in basketball analytics is the high degree of correlation between different performance metrics. This creates significant problems for traditional regression approaches. When predictors are highly correlated, coefficient estimates become highly sensitive to small changes in the data. This correlation also increases the variance of coefficient estimates, making it hard to identify which predictors are truly significant. With correlated predictors, it's challenging to isolate the independent effect of any single variable. In the NBA context, this is particularly problematic. Teams that shoot well from three-point range often have good overall field goal percentages. Teams that play at a faster pace tend to have different defensive profiles than slower teams. Offensive and defensive rebounding percentages are inherently related due to their zero-sum nature. These connections make it difficult to answer questions like "How much does increasing three-point attempt rate independently contribute to winning?" using standard statistical approaches.

3.2 Traditional Regression Approaches and Their Limitations

Ordinary Least Squares (OLS) regression has long been a standard approach for predicting outcomes and identifying relationships in sports analytics. OLS minimizes the sum of squared residuals between observed and predicted values:

$$\min_\beta \|y - X\beta\|^2 \quad (1)$$

Where:
- y is the vector of win percentages
- X is the matrix of predictor variables
- β is the vector of coefficients to be estimated

While OLS provides unbiased estimates under certain conditions, it struggles when predictors are highly correlated. When predictors are highly correlated, OLS can produce coefficient estimates with incorrect signs (opposite of what theory or logic would suggest), coefficients with implausibly large magnitudes, and models that fit training data well but perform poorly on new data (overfitting). These limitations require more sophisticated approaches to modeling NBA team success.

3.3 Elastic Net: A Solution for Correlated Predictors

Elastic net regression combines the strengths of both Ridge and Lasso regularization methods by incorporating both L1 and L2 penalties:

$$\min_\beta\{\|y - X\beta\|^2 + \lambda[(1-\alpha)\|\beta\|_2^2/2 + \alpha\|\beta\|_1]\} \quad (2)$$

Where:

- λ controls the overall strength of regularization
- α determines the mix between L1 and L2 penalties ($0 \le \alpha \le 1$)

This hybrid approach offers several advantages that make it ideal for our NBA analysis. First, when variables are highly correlated, elastic net tends to include or exclude them as a group, unlike Lasso which might arbitrarily select one. Like Lasso, elastic net can reduce coefficients to zero, creating more interpretable models. Similar to Ridge regression, elastic net produces more stable coefficient estimates for correlated predictors. Finally, by tuning α, we can balance between the feature selection properties of Lasso and the group handling capabilities of Ridge. Given these advantages, we selected elastic net regression as our primary modeling approach to address the relationships between basketball metrics while maintaining interpretability.

3.4 Hyperparameter Selection via Cross-Validation

Elastic net requires tuning two hyperparameters: Alpha (λ), which controls overall regularization strength, and L1 ratio (α), which controls the balance between L1 and L2 penalties. We employed k-fold cross-validation (with k=5) to select optimal values for these hyperparameters. The process involved testing a range of alpha values spanning several orders of magnitude (10^{-6} to 10^{6}) and testing a range of l1_ratio values from 0.01 (mostly Ridge) to 0.99 (mostly Lasso), then selecting the combination that minimized prediction error on held-out data. Through this process, we identified optimal values of Alpha (λ) = 0.123285, representing a moderate regularization strength that was strong enough to address correlation concerns but not so strong as to overly constrain the model. The optimal L1 ratio (α) = 0.01 indicates a strong preference for Ridge regression (99%) over Lasso (1%), suggesting that for NBA team success prediction, addressing the variance of coefficient estimates is more important than aggressive feature selection.

These hyperparameter values reflect the nature of basketball analytics where many factors contribute to winning, and their effects are often subtle and interrelated rather than sparse and isolated.

3.5 Model Evaluation Metrics

To assess model performance, we utilized several complementary metrics. R-squared (R^2) measures the proportion of variance in win percentage explained by the model, with higher values indicating better explanatory power. Root Mean Squared Error (RMSE) measures the average magnitude of prediction errors in win percentage units, with lower values indicating better predictive accuracy. We also used train-test performance comparison, comparing metrics on training vs. test data to identify potential overfitting. Residual analysis was also used to examine patterns in prediction errors to check the model assumptions and find areas for improvement. We also validated the model by applying it to predict outcomes for the most recent (2024-2025) NBA season, which allows us to see the real-world utility of the model.

4. Predictor Selection

4.1 Initial Feature Set

Our initial analysis considered a comprehensive set of team statistics that could potentially influence win percentage. These metrics represent different aspects of basketball strategy and performance, including team composition metrics, game pace and style indicators, offensive efficiency metrics, and defensive effectiveness metrics. Team composition is represented by Age, the average age of players weighted by minutes played, reflecting team experience and development stage. Some individual player impact is measured by star_ppg which tracks each team's leading points per game player to serve as a proxy to measure "star power". Game pace and style indicators include Pace (estimated possessions per 48 minutes, indicating how fast a team plays), FTr (free throw rate, reflecting how often a team draws fouls), and 3PAr (three-point attempt rate, indicating shooting strategy). Offensive efficiency metrics encompass TS% (true shooting percentage, a comprehensive shooting efficiency metric), eFG_off (effective field goal percentage, accounting for the added value of three-pointers), TOV_off (turnover percentage, measuring ball security), ORB_off (offensive rebound percentage, showing second-chance opportunity creation), and FT_FGA_off (free throws per field goal attempt, another measure of foul-drawing ability). Defensive effectiveness metrics include eFG_def (opponent effective field goal percentage, measuring shooting defense), TOV_def (opponent turnover percentage, indicating defensive disruption), DRB_def (defensive rebound percentage, showing ability to end opponent possessions), FT_FGA_def (opponent free throws per field goal attempt, reflecting defensive discipline), and DRtg (defensive rating, an estimate of points allowed per 100 possessions). Each metric represents a different strategic element or performance outcome that coaches and players can potentially influence through training, tactics, and roster decisions.

4.2 Correlation Analysis and Feature Assessment

Our correlation analysis revealed significant relationships among several metrics. Shooting efficiency metrics (TS% and eFG_off) showed extremely high correlation ($r > 0.9$), indicating they measure nearly the same concept. Free throw metrics (FTr and FT_FGA_off) were almost perfectly correlated, making their simultaneous inclusion redundant. Some metrics showed moderate negative correlations between their offensive and defensive versions, reflecting the inherent trade-offs in basketball strategy. DRtg showed strong correlations with several component defensive skills, suggesting it aggregates their effects. These relationships highlighted the need for a modeling approach that could handle correlation appropriately, reinforcing our choice of elastic net regression.

4.3 Feature Engineering and Selection Rationale

Based on our correlation analysis and basketball domain knowledge, we made several feature engineering decisions. We removed TS% in favor of eFG_off, as both measure shooting efficiency, but TS% incorporates free throw shooting which is already partially captured by FTr. We kept FTr and removed FT_FGA_off since these metrics provide similar information, but FTr is more commonly used in basketball analytics. Despite moderate correlations, we kept Age to evaluate whether team age can reflect intentional strategic choices between youth/athleticism and experience/skill. We also kept Pace despite weak correlation with win percentage, thinking that pace might show significant effects when controlling for other factors. We also decided to exclude some metrics from our initial modeling, specifically overall efficiency metrics like Offensive Rating (ORtg) and Defensive Rating (DRtg), as they aggregate the effects of more specific skills, and advanced composite metrics that would be difficult for coaches to directly influence. These decisions aligned with our goal of creating an interpretable model focused on actionable strategic factors.

4.4 Final Predictor Set

After initial feature engineering, we identified the following set of ten predictors for our elastic net model:

- Age: Average team age, reflecting strategy choices between youth and experience
- Pace: An estimate of possessions per 48 minutes, indicating offensive tempo
- FTr: Free throw rate, showing how often teams draw fouls per FG attempt
- 3PAr: Three-point attempt rate, representing proportion of shots from 3-pt range
- TOV_off: Offensive turnover percentage, measuring ball security
- ORB_off: Offensive rebound percentage, indicating second-chance opportunities
- TOV_def: Defensive turnover percentage forced, showing defensive disruption
- DRB_def: Defensive rebound percentage, reflecting ability to end opponent possessions
- FT_FGA_def: Opponent free throws per FG attempt, measuring defensive discipline
- star_ppg : Measures star player power by highest scoring player’s points per game

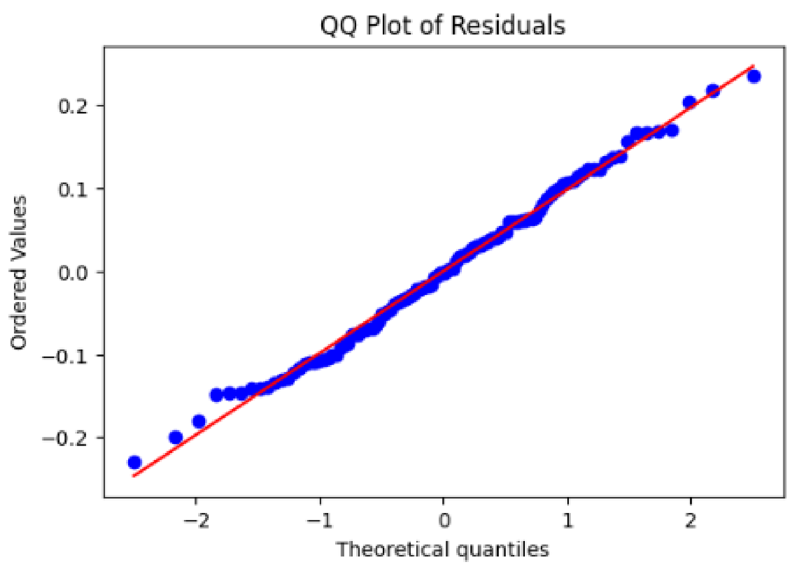
This set balances offensive and defensive factors while focusing on elements that teams can potentially influence through strategic decisions. The elastic net regularization process would further refine this set by adjusting coefficient magnitudes based on their predictive importance.

5. Results and Analysis

5.1 Model Performance

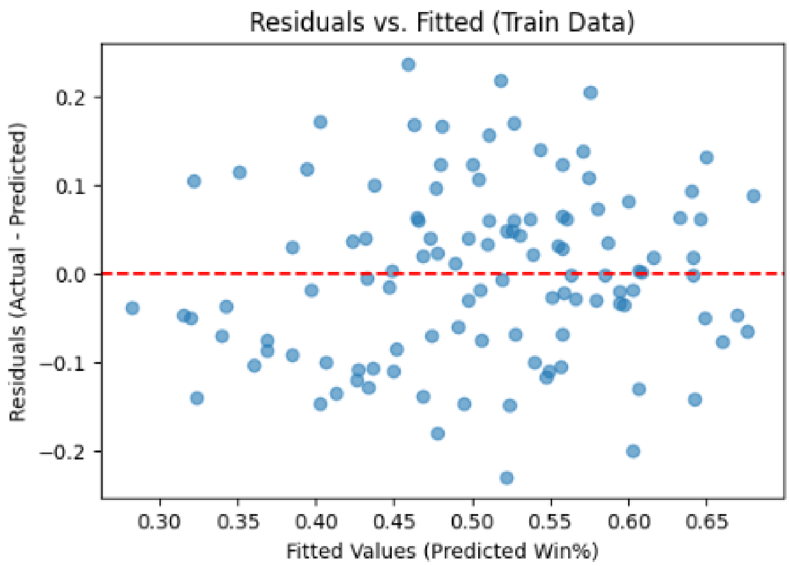
Our elastic net regression model demonstrated strong predictive power for NBA team success. On the training data, the model achieved an R^2 of 0.499 and an RMSE of 0.099. Test data performance was similarly strong with an R^2 of 0.495 and an RMSE of 0.103. When validated on the 2024-2025 season data, the model showed even better performance with an R^2 of 0.544 and an RMSE of 0.108. The similarity between training and test performance metrics suggests the model generalizes well without overfitting. The R^2 value of approximately 0.5 indicates that our model explains about half the variance in NBA team win percentage using just these strategic and stylistic variables, which is impressive considering the model doesn't include specific player talent metrics or injury information. The RMSE of approximately 0.1 means that, on average, the model's win percentage predictions are within ±0.1 of actual results. In an 82-game NBA season, this translates to predictions within about ±8 wins of a team's actual record. To ensure our model was performing properly we checked model assumptions. In Figure 5, we examine the normality of model residuals, which is an important assumption for valid statistical inference. As shown in the QQ plot, the residuals closely follow the theoretical normal distribution line, indicating that our model satisfies the normality assumption.

Figure 5: *QQ Plot of Model Residuals showing the residuals follow a normal distribution*



We plotted the residuals against fitted values to help assess the homoscedasticity assumption, and as we can see in Figure 6, the residuals are fairly evenly distributed across the range of fitted values without any obvious patterns or funneling, suggesting that the model's assumption of constant variance is reasonably satisfied.

Figure 6: *Residuals vs. Fitted Values*



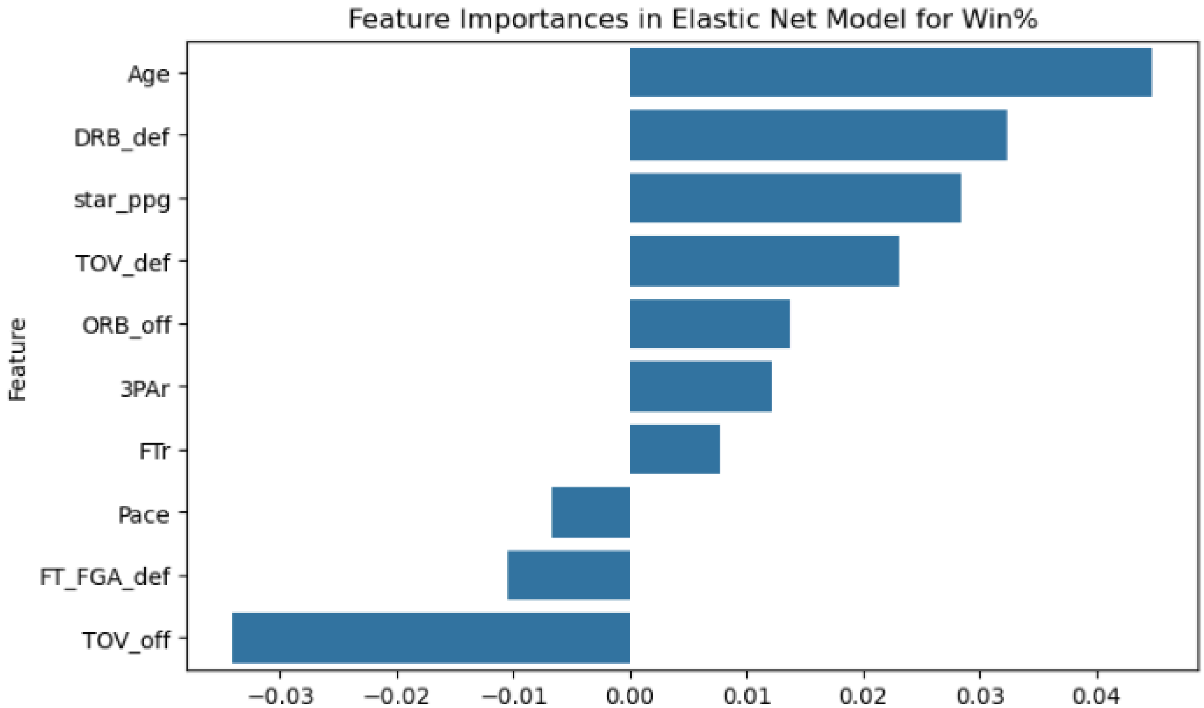
5.2 Coefficient Interpretation

The elastic net model identified nine non-zero coefficients, providing insights into the relative importance of different factors for NBA success. As shown in Table 1 and Figure 7, these coefficients reveal which factors have the greatest impact on team winning percentage.

Table 1*: Elastic Net Model Coefficients and Their Impact on NBA Win Percentage*

Feature	Coefficient
Age	0.044810
DRB_def	0.032363
star_ppg	0.028372
TOV_def	0.023049
ORB_off	0.013730
3PAr	0.012149
FTr	0.007786
Pace	-0.006738
FT_FGA_def	-0.010461
TOV_off	-0.034052
Intercept	0.507

***Figure 7**: Bar chart showing the relative importance of each feature in the elastic net model. Positive coefficients (blue bars extending right) indicate factors associated with increased winning, while negative coefficients (blue bars extending left) indicate factors associated with decreased winning.*



These standardized coefficients reveal several important patterns. Ball security is paramount, with offensive turnovers (TOV_off) having the strongest negative impact, while forcing defensive turnovers (TOV_def) has a substantial positive effect. Rebounding matters significantly, as both defensive rebounding (DRB_def) and offensive rebounding (ORB_off) show positive associations with winning. Experience correlates with success, as team age has the strongest positive coefficient, suggesting veteran teams tend to win more games. Star power matters considerably, with a team's highest scorer's points per game (star_ppg) having the third strongest positive coefficient, indicating the importance of having elite scoring talent. Three-point shooting strategy is important, with higher three-point attempt rates (3PAr) associated with more winning, reflecting the modern NBA's emphasis on perimeter shooting. Free throw dynamics matter, as drawing more free throws (FTr) positively impacts winning, while allowing opponents more free throws (FT_FGA_def) negatively impacts success. Finally, pace has minimal impact, with the small negative coefficient suggesting that, controlling for other factors, the speed of play has little independent effect on team success.

5.3 Impact Quantification: Wins Added per Standard Deviation

To make these coefficients more interpretable in basketball terms, we converted them to represent wins added per standard deviation change in each predictor. Table 2 presents these impacts in terms of additional wins per season.

***Table 2**: Impact of Each Factor in Terms of Wins Added per Standard Deviation*

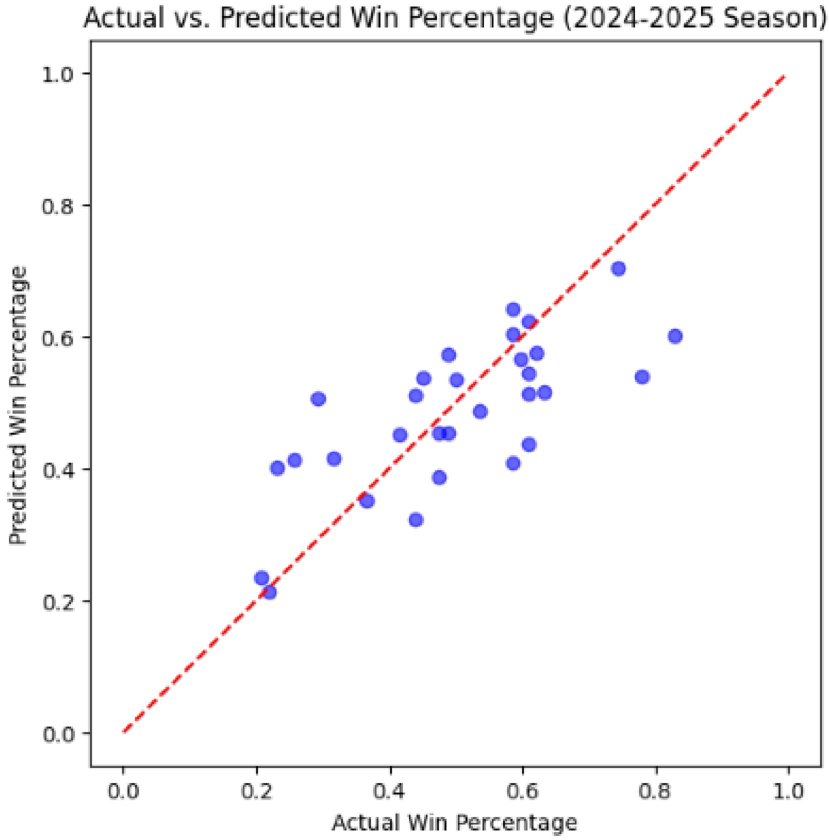
Feature	Coefficient	Standard Deviation	Wins Impact per SD
Age	0.045	1.76	3.7 wins per 1.76 change in Age
DRB_def	0.032	1.66	2.7 wins per 1.66 change in DRB_def
star_ppg	0.028	3.78	2.3 wins per 3.78 change in star_ppg
TOV_def	0.023	1.03	1.9 wins per 1.03 change in TOV_def
ORB_off	0.014	2.29	1.1 wins per 2.29 change in ORB_off
3PAr	0.012	0.04	1.0 wins per 0.04 change in 3PAr
FTr	0.008	0.02	0.6 wins per 0.02 change in FTr
Pace	-0.007	1.99	-0.6 wins per 1.99 change in Pace
FT_FGA_def	-0.010	0.02	-0.9 wins per 0.02 change in FT_FGA_def
TOV_off	-0.034	0.90	-2.8 wins per 0.90 change in TOV_off

This analysis provides actionable intelligence for NBA teams. A team that is 1.76 years older (one standard deviation) tends to win about 3.7 more games per season, suggesting experience is valuable. Improving defensive rebounding percentage by 1.66 percentage points (one standard deviation) is associated with about 2.7 additional wins, while having a leading scorer who averages 3.78 more points per game correlates with approximately 2.3 more wins. Likely the most actionable takeaways come from the fact that reducing offensive turnovers by 0.90 percentage points can lead to about 2.8 more wins per season on average, and increasing three-point attempt rate by just 0.04 (4 percentage points) is associated with 1.0 additional wins (So shoot more threes!). These estimates provide teams with a framework for evaluating the potential impact of strategic changes or player acquisitions on their win totals.

5.4 Team-by-Team Analysis

The model's predictions for the 2024-2025 season revealed interesting patterns of over- and underperformance. As shown in Figure 8, our model performs strongly in predicting team success for the out-of-sample validation season.

Figure 8: *Actual vs. Predicted Win Percentage: Scatter plot comparing actual win percentages to model predictions for the 2024-2025 NBA season.*



The model demonstrates strong predictive performance on this out-of-sample validation set. It can be seen in Table 3 that notable overperformers included the Cleveland Cavaliers (+0.241 win percentage, ~20 more wins than predicted), and the Oklahoma City Thunder (+0.229 win percentage, ~19 more wins than predicted). Notable underperformers included the Philadelphia 76ers (-0.213 win percentage, ~17 fewer wins than predicted) and the Charlotte Hornets (-0.171 win percentage, ~14 fewer wins than predicted). These discrepancies highlight factors outside our model's scope that significantly influence NBA outcomes such as star player impact, injuries, coaching effectiveness, and team chemistry. Teams that overperformed might be executing strategies particularly well or have exceptional talent that elevates their performance beyond what basic metrics would predict.

Table 3: Team Performance vs Predictions

Team	Predicted Win%	Actual Win%	Difference
Boston Celtics	0.705013	0.743902	0.038889
Golden State Warriors	0.642912	0.585366	-0.057546
Los Angeles Clippers	0.623210	0.609756	-0.013454
Milwaukee Bucks	0.603312	0.585366	-0.017946
Oklahoma City Thunder	0.600677	0.829268	0.228591
New York Knicks	0.575982	0.621951	0.045969
Sacramento Kings	0.574266	0.487805	-0.086461
Minnesota Timberwolves	0.567079	0.597561	0.030482
Los Angeles Lakers	0.543736	0.609756	0.066020
Cleveland Cavaliers	0.539127	0.780488	0.241361
Miami Heat	0.536789	0.451220	-0.085570
Orlando Magic	0.535760	0.500000	-0.035760
Houston Rockets	0.516976	0.634146	0.117171
Denver Nuggets	0.514182	0.609756	0.095574
Phoenix Suns	0.510366	0.439024	-0.071342
Philadelphia 76ers	0.505384	0.292683	-0.212701
Detroit Pistons	0.486740	0.536585	0.049846
Atlanta Hawks	0.454891	0.487805	0.032914
Dallas Mavericks	0.454007	0.475610	0.021603
San Antonio Spurs	0.451621	0.414634	-0.036986
Indiana Pacers	0.437774	0.609756	0.171982
Brooklyn Nets	0.415650	0.317073	-0.098577
New Orleans Pelicans	0.412547	0.256098	-0.156450
Memphis Grizzlies	0.408349	0.585366	0.177017
Charlotte Hornets	0.402365	0.231707	-0.170658
Chicago Bulls	0.387966	0.475610	0.087644
Toronto Raptors	0.350721	0.365854	0.015133
Portland Trail Blazers	0.322164	0.439024	0.116860
Utah Jazz	0.236025	0.207317	-0.028708
Washington Wizards	0.212718	0.219512	0.006794

6. Discussion

6.1 Key Insights

Our analysis yields several important insights for NBA team strategy and management. Ball security is critical, with the strong negative impact of offensive turnovers making this perhaps the most actionable finding. Teams should prioritize players who limit turnovers and implement systems that reduce risky plays. Defense creates offense, as forcing turnovers and securing defensive rebounds are among the most impactful factors. This supports the basketball adage that "defense wins championships." Experience matters significantly, with the substantial positive effect of team age suggesting that veteran leadership and experience translate to tangible success. Teams rebuilding with young players should recognize this potential short-term cost. Star power is crucial, with the significant positive impact of having a high-scoring leading player demonstrating that elite offensive talent remains a key driver of success in the NBA. The positive association between three-point attempt rate and winning validates the league's shift toward perimeter shooting, even when controlling for other factors. Pace is largely neutral, with the minimal impact of pace on winning suggesting teams can play at tempos that suit their personnel without significantly affecting their win total, provided they execute well in other areas. Both drawing fouls and avoiding defensive fouls independently contribute to success. While significant, offensive rebounding shows less impact than defensive rebounding, suggesting teams might be justified in prioritizing transition defense over offensive boards in some situations. These insights provide a data-driven foundation for strategic decision-making by NBA front offices and coaching staffs.

6.2 Practical Applications

Our findings can guide NBA teams in several concrete ways. For player acquisition, teams could use these coefficients to identify undervalued players who excel in high-impact areas like turnover prevention and defensive rebounding. Player development programs could emphasize skills that contribute most to winning, such as ball security and three-point shooting. Coaches could allocate practice time and tactical focus proportionally to the factors that most influence winning. The strong positive effect of team age suggests balancing youth with veteran experience, particularly when building contending teams. The significant impact of star scoring confirms the value of acquiring elite offensive talent, though interestingly this impact is less than some might expect, suggesting team-oriented approaches remain competitive. Teams can confidently choose pace and style based on their personnel strengths rather than league trends, knowing these factors have less independent impact on winning. Front offices could allocate resources (salary, coaching staff, analytics) toward addressing weaknesses in the highest-impact areas. For example, a team could estimate that reducing their turnover rate by one percentage point while improving defensive rebounding by two percentage points might add approximately 5-6 wins to their season total on average which is valuable intelligence to have when making roster decisions.

6.3 Model Limitations

Despite its strong performance, our model has several important limitations. The model excludes critical factors like player talent level, injuries, strength of schedule, and coaching quality that significantly impact outcomes. While the model identifies associations between factors and winning, it cannot definitively establish causal relationships. Teams with better players likely win more games and have better statistics. The model treats each factor independently, potentially missing important interaction effects (e.g., how pace might differently affect teams with varying age profiles). Using season-level data obscures within-season adjustments and game-to-game strategic variations that teams employ. The five-season sample may not capture longer-term trends or strategic cycles in the league. The model doesn't explicitly account for evolving league trends, treating all seasons as equivalent. Different teams may have different optimal strategies based on their unique personnel, which this general model cannot capture.

6.4 Potential Improvements

Future research could address these limitations through several enhancements. We could incorporate player-level data to help separate strategy effects from talent effects, and include more sophisticated metrics like RAPM (Regularized Adjusted Plus-Minus) to better capture player impact. Using game-level data instead of season-level aggregates would likely reveal more nuanced strategic insights about how teams adapt throughout a season. We could also explicitly model year-to-year changes to better understand evolving league trends, while accounting for team-specific random effects might help identify which organizations consistently over- or underperform statistical expectations. Adding injury data, strength of schedule, and travel burden would improve predictive accuracy by capturing important contextual factors. Another promising direction would be exploring tree-based models like Random Forest or XGBoost with these additional predictors, which could take this research to the next level by capturing non-linear relationships and interactions. While these enhancements would require more complexity and additional data sources, they could yield even more actionable insights for NBA teams looking to gain competitive advantages.

7. Conclusion

7.1 Summary and Implications

This research used elastic net regression to identify and quantify the impact of various strategic factors on NBA team success. Our model explains approximately 50% of the variance in team win percentage, showing that a substantial portion of basketball success can be predicted from team statistical profiles. For NBA teams, these results suggest several strategic priorities. First, given the substantial impact of turnovers it is important to invest in ball security, teams should prioritize guards and ball-handlers who protect possessions and implement intelligent systems that minimize risky plays. Second, value defensive glass cleaners, as players who excel at defensive rebounding contribute significantly to winning in ways that may not always be fully appreciated. Third, balance youth with experience, as while developing young talent remains important for long-term success, veteran presence appears to significantly impact winning in the short term. Teams should invest in star offensive talent as having a leading scorer who can consistently produce points makes a significant contribution to team success. Teams should embrace the three-point revolution. The positive impact of three-point attempts supports modern shooting strategies, but finding a balance and creating layups does remain important. Finally, teams should choose tempo based on personnel. With pace showing minimal independent effect on winning, teams can confidently play at speeds that maximize their particular strengths.

Teams that align their strategies with these empirically-derived insights may gain competitive advantages, particularly if they identify areas where traditional basketball thinking undervalues certain skills or approaches.

7.2 Future Research Directions

This work opens several promising avenues for future research. Extending this approach to player-level data could help teams identify which players contribute most to these key statistical indicators. Investigating even further how these factors interact with each other (e.g., how pace affects turnover rates) could reveal more nuanced strategic insights. Tracking how the importance of different factors evolves over time could help teams anticipate future trends. Analyzing how these factors differ in importance across player positions could inform more targeted player development strategies. Examining how the impact of these factors varies by game context (close games, blowouts, clutch time) could reveal situational priorities. By continuing to refine and extend this analytical approach, NBA teams can gain increasingly sophisticated insights into the complex relationships between basketball strategies and winning outcomes, ultimately creating competitive advantages in a league where margins of victory are often razor-thin.

8. References

Basketball-Reference.com. (n.d.). *Basketball statistics and history*. Sports Reference LLC. https://www.basketball-reference.com/

Member Contributions:

As a group we met multiple times and planned out the process and did check in with each other on progress. We split up the slides evenly with each of us taking a section which we presented. Drew wrote an initial code and paper rough draft and then Matt and Aaron made big changes and added features and different exploration to the code, and then went through the paper and made large edits and revisions to update it. At the end, we all went through together and checked each other’s writing to make sure we had the desired quality. To summarize, all three of us contributed about the same and it was a great experience learning from each other’s strengths.