

Identifying Correlations Between NBA Success and Team Dynamics: A Statistical Analysis

Naithik Pradeep

Received January 30, 2024

Accepted June 19, 2024

Electronic access June 30, 2024

This paper analyzed various correlations between numerous aspects of Team Chemistry as well as Team Diversity and success in the National Basketball Association. NBA stats were collected for all 30 NBA teams for eight years ranging from the 2015-2016 NBA season to the 2022-2023 NBA season. Team Chemistry was measured with an assortment of in-game hustle and tracking NBA team statistics. Team Diversity was measured through the mean and standard deviations of individual characteristics of players within a team. Among these statistics, other team dynamics were tested such as roster continuity and roster composition. For each team, individual characteristics were measured for the top 15 players in minutes played. These statistics were compared to team success statistics and we used linear, quadratic, logarithmic, and other models to identify correlations. This paper found the impact of the various aspects of NBA Team Dynamics. It was found that Team Chemistry ($P = 0.031$), Location Diversity ($P = 0.055$), and Roster Composition ($P = 0.002$) had meaningful conclusions when compared to team success, but Experience Diversity ($P = 0.865$) and Roster Continuity ($P = 0.502$) did not have as meaningful an impact on Team Chemistry.

Keywords: Team Dynamics, Team Chemistry, Team Diversity, Regressions

Introduction

Basketball is a premier global sport that has captivated millions of fans through the electrifying displays of individual talent and teamwork. At the pinnacle of professional basketball stands the NBA, a league in which team dynamics are heavily emphasized when debating team success.

In the NBA, three major components are discussed when looking at an NBA team. One of them is Team Diversity, or the idea of having members of the team ranging from various backgrounds that can all provide something that can lead to success. Another is team chemistry, which is how well players on a team synergize and play together. The last is roster composition or how the roster itself is constructed and how each player got onto the team.

This paper intends to find out how impactful these highly stated ideas are to a team's success. Are factors such as team chemistry, team diversity, and roster composition associated with success in the NBA? I hypothesized that aspects of Team Chemistry and Team Diversity both positively correlated with winning in the NBA. I intend to find out whether this is true and see how strong of an impact they have. I will collect data from the NBA and use regressions to find their strength and impact. I will then create conclusions and ideas about the data.

Similar research has been done in this field by Warren Gonsalves in his paper, "Implications of Diversity, Equity and Inclusion Strategies on Player and Team Performance, and Reten-

tion in Professional Basketball Sport Organizations" (Gonsalves 2022)¹, Stanley Yang's thesis "Predicting Regular Season Results of NBA Teams Based on Regression Analysis of Common Basketball Statistics" (Yang 2015)², Guan-Yuan Wang's article, "The role of diversity in determining team efficiency: an empirical sports team analysis" (Wang 2024)³, and a paper written by Allan Maymin, Philip Maymin, and Eugene Shen called "NBA Chemistry: Positive and Negative Synergies in Basketball" (Maymin et. al 2013)⁴. I also looked into similar research of different sports including Daniel Dasgupta's research, "The Overlooked Element: An Empirical Analysis of Team Chemistry and Winning Percentage in Major League Baseball" (Dasgupta 2017)⁵. This research was sufficient for me to build a solid foundation on this topic. The insights I gathered from them helped motivate me in my journey to share this piece with the world.

Methodology

Data Collection

The various aspects of overall Team Dynamics can be defined with a combination of individual player statistics and overall in-game team statistics. The in-game statistics like assist percentage and screen assists per game were collected from nba.com. Individual player characteristics like experience and salary were taken from websites like Basketball Reference⁶ and RealGM⁷.

Data was collected from the last eight years from the 2015-2016 season to the 2022-2023 season. For individual player characteristics, I only used the top 15 players in minutes played for each team. The data was either manually inputted or webscraped. We know that the process of web scraping and manual data input can be error-prone. For this reason, we checked our final dataset against the true values given on the website. This data validation step was essential to ensure the accuracy of our results.

Data Features

Experience-Based Features

The first feature is the *Average Years Played*, which describes the mean of every player's experience on a team. I chose this feature because it can identify the difference between an experienced team and an inexperienced team and find the best range of ages for a team to have. Players in the NBA usually have their peak years six to eleven years after they start playing, and this feature can help determine if winning is correlated to these prime years.

The next feature is the *Standard Deviation of Years Played*, which outlines the amount of variance within the tenures of the individual players of each team. The purpose of this feature was that it was a good metric to detail age diversity within a team and find out whether a group of players who were all of similar experience or a group of players with varying levels of experience is better. It can identify the significance of varying experiences.

Another important aspect of winning is the veteran, outlined in this next feature, *Number of Veterans*. A veteran is defined as a player who has played for seven or more years because, by this point, players will have enough experience to play reliably and can be a calming influence to any younger players on the roster. This feature is the total count of such veterans on a team. This feature will be used to measure the value of veterans on winning and whether intangibles other than performance can benefit a team.

This next feature, *Coach's Experience*, measures the number of years a head coach has coached. Along with the experience of a player, the experience of a coach can determine how successful some of the plays and schemes that teams run can be. The difference between an experienced and inexperienced coach is what can be determined through this feature.

The final feature in this list of age-based features is the *Combined Tenure of Teams*. This feature will take both the *Average Years Played* and the *Coach's Experience* data features into account. The formula for this feature is $(Average\ Years\ Played + (Coach's\ Experience / 2)) / 2$. The reasoning for this formula is based on the assumption that the players' experience is twice as valuable as the coach's experience. This feature aims to find the overall impact of the age of each team member on the success they bring.

Chemistry-Based Features

Our first chemistry-based feature is the *Roster Continuity*. It is calculated by finding the percentage of regular season minutes filled by players from the previous season's roster. In other words, it tracks how continuous the roster is from the previous year minutes-wise. This feature is a measure of team chemistry because it can determine the effect of having a familiar roster and players with built-up synergy.

Our next feature is the *Standard Deviation of Minutes Played*. This feature outlines the level of variance within playing time in a team. It can be used to find the impact of equal opportunities and can outline the effect of ego dynamics within a team. This feature is a measure of chemistry because it reflects the level of balance or cohesion in a team and can find the impact of equitable distribution of playing time.

Standard Deviation of Salaries is the next feature and it takes each player's salary on a team and finds the level of variance within these values. Although there isn't any direct correlation between this feature and team chemistry, a higher standard deviation might potentially lead to issues in team chemistry. This feature aims to find out whether this assumption is true and see whether an equitable distribution of salaries might benefit a team's dynamics and overall chemistry.

Our final team chemistry feature is called *Team Chemistry Score*. This feature is what I believe to be an accurate reflection of team chemistry during the game. This feature takes four statistics into account. The statistics are *Number of Passes Made* (The number of passes made per game by a team), *Number of Possessions Off Screens* (The number of possessions in which a screen was used to execute a play in a game), *Assist Percentage* (The percentage of made field goals that were a result of an assist in a game), and *Number of Screen Assists* (The number of field goals that were a result of a screen in a game). The purpose for choosing these metrics was that these statistics are ways to measure the amount and frequency of the plays in which multiple team members were involved. Plays like these require some sort of synergy between players to occur, making these statistics an accurate reflection of team chemistry. The formula for this feature was to scale all the averages of the statistics so that they were all weighted equally and then add them together to create a feature that takes each of these statistics into account equally.

Location-Based Features

Our first location-based feature is the *Number of Unique States*. For this feature, the state that each player on a team was born in was researched. This feature is the total number of unique states on a team. The United States is extremely diverse with each state providing a unique blend of culture, landscaping, and heritage. This feature aims to find out the effect of these differences on team play and see whether diversity within the United States has any impact on the NBA.

Similar to our previous feature is the *Number of Unique Coun-*

tries. This feature outlines the total number of unique birth countries for players on a team and is used as a measure of location diversity. The goal of this feature is to see whether differences between different countries and cultures play a part in the cohesion and synergy of a team.

This next feature, *Combined Location Diversity*, will take both the *Number of Unique States* and the *Number of Unique Countries* features into account. This feature aims to find the impact of overall team location diversity and combines the two features above using the following formula: $\text{Number of Unique States} + \text{Number of Unique Countries} * 2$. This formula was created with the assumption that a different country was twice as diverse as a different state. This assumption was made because all the states are in one country and have many similarities that different countries do not.

Roster Composition Features

The first roster composition feature is the *Number of Drafted Players*. This feature looks at every player on a team and is the count of the number of players that got onto the team through the draft. This feature is used to determine the effect of building up a team from scratch with home-grown talent and whether this method can lead to success.

Our next feature is the *Number of Traded Players*. This feature is the total number of players that were traded to the team they are currently on. Trades in the NBA come in two varieties: Trading for the future and trading for the present. Many times, trades with the goal being future success involve draft picks and young players and our traded to rebuilding teams. The other type of trade happens when teams need better players and trade for either stars or valuable role players who can help the team succeed. Taking both of these types of trades into account, this feature aims to find the overall impact of trades and the effect of having traded players on a team. It also aims to find the effect of lowered team chemistry because a new player on a team will have to build up synergy with the existing players on the team.

The last roster composition feature is the *Number of Free Agency Players*. This feature takes the number of players that were signed to a team through free agency and is the total amount of such players on a team. The purpose of this feature is to find out the impact of making moves in free agency in the off-season, the impact of the salary cap, and how the amount of money a team spends to strengthen their roster impacts success.

Success Features

Our first success feature is the *Winning Percentage*. feature takes the number of wins by a team in a season and divides it by the number of games a team plays in a season. This statistic is extremely useful and reliable when looking at how successful a team was in the regular season.

Our final feature is Net Rating, which is the average point differential per game of a team in a season. This feature is another great way of defining success because it looks past

the results of individual games and looks at the level of team performance regardless of victories.

Data Analysis

All the features excluding *Team Chemistry Score*, *Winning Percentage*, and *Net Rating* are the response features and will be compared to *Team Chemistry Score*, *Winning Percentage*, and *Net Rating*. Linear, quadratic, and cubic regressions will be used to find several summary features. For each comparison, I originally fitted a linear regression to the data and looked at its characteristics. For many of the regressions, the linear fit was kept, but for others with different and more identifiable patterns, various other regressions were chosen to match the characteristics of the data. These models are not inferential but are to find and showcase interesting relationships between variables. These will be the R Squared and the Slope with its corresponding P-value. These coefficients will help find both the strength and impact of the regressions. Graphs will be shown as evidence to help support any arguments. After the analysis, conclusions will be formed about the regressions, and the data and its relevance will be discussed.

Results

When looking at the regressions including the experience of players, most regressions were positive and linear with small slopes. Some notable positive slopes include:

1. Average Years Played and Team Chemistry Score linear regression in which for each increasing average year played the Team Chemistry Score increased by 5.85. The standard error of this slope is 6.69. Although this point estimate is large, the variance is large so we cannot say that there is a statistically significant correlation.
2. Average Years Played and Winning Percentage linear regression in which for each increasing average year played, the Winning Percentage increased by 4.51% which is roughly 3.7 more wins in an 82-game season. The standard error of this slope is 0.57%. This correlation also has a confidence interval with a lower bound of 3.4% and an upper bound of 5.6%. This is around 2.8 games to 4.6 games more per additional year.
3. Average Years Played and Net Rating linear regression in which for each increasing average year played, the Net Rating would increase by 1.36. The standard error of this slope is 0.19. The lower bound of the confidence interval is 0.985 and the upper bound is 1.735. Having a higher average years played is associated with a higher net rating.

Some negative slopes include:

1. Number of Veterans and Team Chemistry Score cubic regression in which the regression was negative but still negligible. This regression has a slightly significant R Squared value of 0.578. This data was aggregated with the mean.
2. Number of Veterans and Net Rating cubic regression in which the regression was negative but significant. This regression has a significant R Squared value of 0.902, which is the highest R Squared value out of all the regressions. This data was aggregated with the mean.

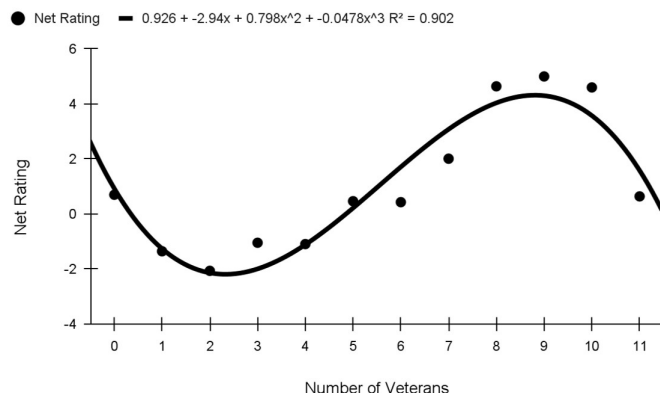


Fig. 1 Number Of Veterans And Net Rating (Aggregated)

When looking at the regressions for the Standard Deviation of Minutes Played, they are all linear but extremely varied:

1. The Standard Deviation of Minutes Played and Team Chemistry Score regression is negative but extremely insignificant.
2. However, compared to Winning Percentage, the regression is positive and significant. For each standard deviation of minutes *per game**, the Winning Percentage increased by approximately 0.046%. This equates to around 0.04 more wins in an 82-game season. The standard error of this slope is 0.006%. The confidence interval is from 0.034% to 0.058%. While this result is statistically significant, it doesn't really have any practical significance.
3. This is also true when compared to Net Rating. For each standard deviation of minutes per game, the Net Rating increased by approximately 0.0144. The standard error of this slope is 0.002. The confidence interval of this correlation is from 0.011 to 0.018.

The Roster Continuity regression turned out to have positive but insignificant correlations. This is also true in the Standard

*This metric was formed by multiplying the standard deviation of total minutes played by average games played.

Deviation of Salaries regressions with the regression when compared with Team Chemistry Score being negative. The Location Diversity feature regressions proved to be more interesting. Some notable regressions include the:

1. The Number of Unique States and Team Chemistry Score linear regression is negative and significant while the Number of Unique States and Net Rating cubic regression is positive and significant. Both success metric regressions had higher R Squared values. This data was aggregated with the mean.
2. The Number of Unique Countries and Net Rating regression is a positive exponential regression. Again, both success metric regressions had higher R Squared values. This data was aggregated with the mean.
3. The Combined Location Diversity regressions were positive except for the regression including Team Chemistry Score. This time, both success metric regressions had slightly significant R Squared values. The regression with Winning Percentage was quadratic.

When comparing Team Chemistry Score to the success metrics, the regressions are positive but extremely insignificant linear regressions. The roster composition features include more significant regressions:

1. The Number of Drafted Players regressions were all positive with the regression with Team Chemistry Score being linear and extremely significant with a slope of 22.1 and a standard error of 13.25 and the regression with Net Rating being quadratic and significant.
2. All the Number of Traded Players regressions were negative. When compared to Team Chemistry Score, the regression was extremely significant with the slope being -11.9. The standard error for this slope was 3.21. The confidence interval is from -18.182 to -5.618.
3. The Number of Free Agency Players and Team Chemistry Score regression was negative and quadratic while the other two regressions compared to success metrics were positive, linear, and slightly significant.

This chart shows that the more players that have been traded to the team, the more games are lost. 2.48% more games will be lost for each traded player. This translates to roughly 2 more games won. Finally, the Team Chemistry Score regressions were insignificant and positive linear regressions.

For each regression and graph, I found multiple summary statistics to display the impact and strength of the regressions. I found simple summary statistics like slope and intercept. I also used the R-squared value and the P-value to explain information

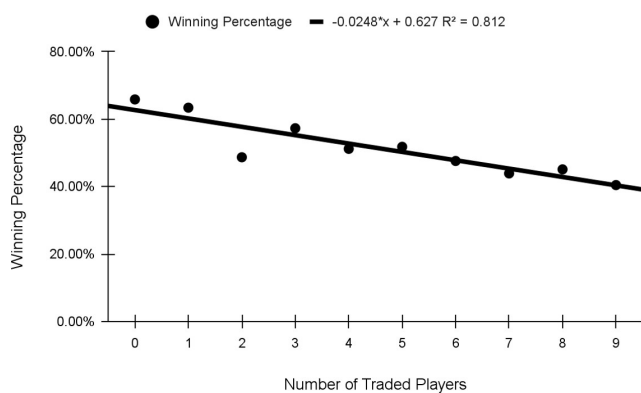


Fig. 2 Number of Traded Players and Winning Percentage (Aggregated)

about the regressions displayed in the Regressions Table in the Index. It's important to note that the Number of Traded Players regressions and Team Chemistry Score regressions have lower P-Values. One of these regressions with the Team Chemistry Score can be seen below.

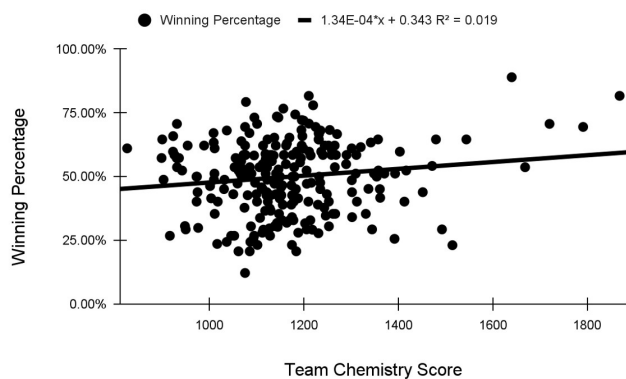


Fig. 3 Team Chemistry Score and Winning Percentage

These various regressions and P-Values can help create conclusions about the data as well as confirm whether these conclusions are by random chance or not.

Discussion

When looking at the features that fall under experience data or experience diversity such as Average Years Played and Standard Deviation of Year Played, there seems to be little to no correlation to success, which suggests that overall team experience and experience diversity are insignificant and can be looked over. The important role of the veteran. The data support that the number of veterans is correlated with success in the NBA which

helps prove this role's significance. The difference between the veterans being significant and other experience metrics being insignificant could be attributed to veterans being a feature that measures the impact of specific players, rather than an overall team interpretation. Roster Continuity, while having positive correlations to success metrics, seemed to have insignificant regressions that could mean that built-up chemistry doesn't affect success in the NBA. While the location, country, and state diversity data have positive correlations with the success metrics, they seemingly do not correlate with Team Chemistry Score. This could mean that location diversity has no impact on the synergy of the players. When looking at the roster composition regressions, the Number of Drafted Players seems to have a positive correlation with success metrics and the Number of Traded Players has a negative correlation with the same success metrics. A possible interpretation is that homegrown talent leads to more winning and trades out of desperation or belief of fit seem to lead to less success. Free Agents joining a team also have a positive correlation with winning which could mean that the willingness to play for a team is significant. Roster composition seems to be very varied and the differences in the correlations can be attributed to the attitudes of the individual players and how they got onto the team originally. Overall, Team Chemistry seemed to not have any correlation with winning whatsoever, which could mean that the synergy of players doesn't matter when looking for team success. This was also seen before in roster continuity. Both the standard deviation of salaries and minutes played had more success with increasing values, but the change was minuscule. This could signify that although diversity in salary and roles is helpful, it isn't really necessary for team success. Overall, it seems like team diversity is somewhat positively correlated with success, while team chemistry isn't correlated to success. When looking at these data features in terms of real life, most of them impact success by either a few Net Rating points or a few games. These games are very valuable when fighting for a higher playoff seed as many teams will be within 2 games of each other at the end of the season. These data features have a much larger impact on the playoff seedings.

Conclusion

Are the factors of team diversity and chemistry associated with winning in the NBA? In short, yes. While many aspects of these concepts do have an insignificant impact on winning, other aspects do positively correlate with success. Looking through the regressions, it seems like Roster Continuity and Team Chemistry are factors that are not particularly impactful to winning. The regressions that tell the best story all involve Roster Composition which means that the way a team is constructed is very impactful to success. Finally, many diversity factors seemed to have little to no impact while others like location diversity,

minutes played diversity, and salary diversity have a much bigger impact on success. There are a few limitations to this study. There is a limited temporal scope for the data as it only covers eight years of the vast history of the NBA. There can also be a large variance between the data of these eight years and the current NBA. This study doesn't account for how the NBA changes in playstyle and may not be relevant to the current state of the league. This study was also very data-driven and more exploration of relationships between variables rather than an inferential study. My key takeaway is that when constructing a team to solve a problem, the factors of team dynamics should be considered, but not completely depended upon. I believe that further research should be conducted on ways to measure Team Chemistry as I feel that mine has a few flaws. I also believe that Racial Diversity should be researched. I would also study how the data features included in this paper change with time to try and identify playstyle and team dynamic shifts in the league. Qualitative research can also be taken into account, including player and coach interviews and team assessments. Finally, I would suggest researching the impact of these data features on a larger scale with a cross-cultural analysis, identifying universal principles of team dynamics and highlighting unique aspects of the NBA.

Acknowledgements

Thank you for the guidance of Spencer Matthews from UC Irvine in the development of this research paper.

References

- 1 W. Gonsalves, *Implications of Diversity, Equity and Inclusion Strategies on Player and Team Performance, and Retention in Professional Basketball Sport Organizations*, p. 1–46.
- 2 S. Yang, *Predicting Regular Season Results of NBA Teams Based on Regression Analysis of Common Basketball Statistics*, p. 1–31.
- 3 G. Wang, *J. of Data, Inf. and Manag*, **6**, 85–98.
- 4 A. Maymin, P. Maymin and E. Shen, *International Journal of Computer Science in Sport*.
- 5 D. Dasgupta, *Economics Student Theses and Capstone Projects*, **45**, year.
- 6 B. Reference, *Roster Continuity*.
- 7 B. RealGM, *NBA Team Roster Compositions By Team*.

X-Axis	Y-Axis	Regression	R-Squared	P-Value	Slope	SE of Slope	Low Bound	Up Bound
# Veterans	Net Rating	Cubic	0.902	N/A	N/A	N/A	N/A	N/A
# Veterans	Team Chem	Cubic	0.578	N/A	N/A	N/A	N/A	N/A
# Countries	Net Rating	Exponential	0.679	N/A	N/A	N/A	N/A	N/A
# States	Winning %	Quadratic	0.804	N/A	N/A	N/A	N/A	N/A
Location Div.	Net Rating	Linear	0.432	0.0551	0.294	0.1525	-0.0049	0.5929
Roster Cont.	Net Rating	Linear	0.559	0.000	0.143	0.0212	0.1014	0.1846
# Drafted	Net Rating	Quadratic	0.453	N/A	N/A	N/A	N/A	N/A
# Drafted	Team Chem	Linear	0.451	0.0967	22.1	13.2536	-3.8771	48.0771
# Traded	Winning %	Linear	0.812	0.0048	-0.025	0.0087	-0.0419	-0.0077
Free Agency	Winning %	Linear	0.517	0.0056	0.009	0.0033	0.0027	0.0156
Free Agency	Team Chem	Quadratic	0.827	N/A	N/A	N/A	N/A	N/A
Team Chem	Winning %	Linear	0.019	0.0308	0.0001	0.0001	0.0000	0.0003

Table 1 Regression Analysis Results

Statistic	Min	1Q	Med	3Q	Max	IQR
# Veterans	0	3	4	6	11	3
Coach's Exp.	0	2	5	10	26	8
Roster Cont.	0.1	0.5	0.7	0.8	1	0.24
# Unique States	4	8	9	10	12	2
# Unique Ctry	1	3	4	5	8	2
Location Div.	12	16	18	19	24	3
Team Chem	826	1077	1148	1232	1869	155
# Drafted	0	4	5	6	10	2
# Traded	0	3	4	5	9	2
# Free Agency	1	4	5	7	12	3
Winning %	12	40	51	60	89	19.7
Net Rating	-11	-3	0.2	3	12	6

Table 2 Descriptive Statistics