

## First Draft

# STAT333 Final Project Proposal

Casey Gay, Tom Holley

October 2025

## 1 Background Information

Basketball and the way it is played has transformed and evolved over time. We are interested in statistical performance factors that are key to success in today's NBA. We seek to identify which factors best explain a team's success, as determined by winning percentage. This study will highlight the most important factors for success, thus providing areas of focus for successful teams.

## 2 Literature Review

Several other studies over the years have done similar research and explored similar hypotheses.

Chatterjee, Campbell, and Wiseman (1944) made a regression model explaining an NBA team's winning percentage. They found a few key statistics, which were field goals, free throws, turnovers, and rebounds, to be significant in predicting a team's chance to win games. The field goals differential specifically was found to explain the largest proportion of the variability of the data. Assists and steals were never found to be statistically significant, and 3-point percentage was largely insignificant with the exception of a high-leverage point. It is important to note that this study examined the differentials between competing teams rather than the team's statistics themselves, whereas we will be using the team's statistics themselves in our study. The foundation of this study is similar to our topic in question - using performance statistics to predict team success. Our study is different, however, as it is focused on more recent years. This is largely relevant because of the development of game strategy - namely the widespread implementation of the 3-point shot. [Chatterjee et al., 1994]

Cabarkapa, Deane, Fry, Jones, Cabarkapa, and Philipp (2022) studied more recent years (2016-2019). They found the key statistics to be shooting percentage and defensive rebounding. Overall shooting efficiency (free-throws, two-point shots, and 3-point shots combined) was also found to explain considerable portions of the variance. The study noted differences in the pace of games and how pace may impact the key statistics in predicting the winner of the game.

One point of focus within this study was the changing of the significance of the performance statistics between the regular season and the post-season, due to different game strategies and the importance of each game. This difference is not something that we plan to examine in our study. We plan to restrict our study to the regular season, since all 32 teams will be equally represented in this format. This study cites the Basketball Reference website as its primary data source, which is the same source that we plan to utilize for our study. [Cabarkapa et al., 2022]

D'Urso, De Giovanni, and Vitale (2022) used a multivariate copula-based approach of the Bayesian networks to identify the main determinants affecting the player winning percentage taking into account the most important indicators suggested in other literature. This study uses a notably different approach than our own, and it also looks to answer a slightly different question. It is still of great value, however, since it analyzes the relationship between performance statistics and basketball success in the NBA, which is an underlying theme of our own study. Offensive and defensive rating were found to be the most important player statistics for player success. Hence, players do better when they contribute meaningfully to their team, which in turn boosts team success as well. While our study does not explicitly examine the effect of individual players, this study suggests that individual players' success has an impact on the success of their teams. [D'Urso et al., 2022]

### 3 Problem Statement

The goal of this project is to determine which measured performance statistics have the strongest linear relationships with team winning percentage. We aim to predict a team's winning percentage based on the statistics deemed significant. Our study utilizes data from the last 5 relevant NBA seasons\*. We will be able to provide NBA teams with data-driven insights into factors that are most important to their chances of winning a game. The results of this study will be able to be utilized for purposes such as team strategic focuses, team record forecasting, and other related actions.

\*2019-2020 excluded for "Bubble" COVID season

### 4 Questions

- i) What statistics are most important in influencing a team's winning percentage?
- ii) How accurately can our selected model predict a team's winning percentage?

## 5 Hypotheses

Null Hypothesis:

$$\beta_1 = \beta_2 = \dots = \beta_n = 0$$

Team winning percentage does not have a relationship with a team's performance statistics.

Alternative Hypothesis:

$$\text{At least one } \beta_j \text{ does not } = 0$$

There are some performance statistics that have a relationship with team winning percentage.

## 6 Data

We will be using data from basketball reference. This is public data. Our variables will be FGA, FGWe have data from the 2018-2019 season, and data from the 2020-2021 season through the 2023-2024 season. Note that the 2019-2020 season is excluded, since this season had extreme circumstances (COVID "Bubble") that greatly differentiate it from other seasons. We have deemed this time frame most relevant to our study.

## References

- [Cabarkapa et al., 2022] Cabarkapa, D., Deane, M. A., Fry, A. C., Jones, G. T., Cabarkapa, D. V., Philipp, N. M., and Yu, D. (2022). Game statistics that discriminate winning and losing at the nba level of basketball competition. *PLOS ONE*, 17(8).
- [Chatterjee et al., 1994] Chatterjee, S., Campbell, M. R., and Wiseman, F. (1994). Take that jam! an analysis of winning percentage for nba teams. *Managerial and Decision Economics*, 15(5):521–535.
- [D'Urso et al., 2022] D'Urso, P., De Giovanni, L., and Vitale, V. (2022). A bayesian network to analyse basketball players' performances: a multivariate copula-based approach. *Annals of Operations Research*, 325:419–440.

Name: Brock Pham

Author's Name: Carson Gay

Author's manuscript title: Significant statistical predictors  
in Basketball

1.) The research question is:

What statistics are most important in influencing a team's winning %  
How accurately can our selected model predict a team's winning %.

2.) Identify the three sources that the author clearly analyzed thoroughly.

Chatterjee, Campbell, and Wiseman (1994)

Cabarkapa, Denye, Foy, Jones, Cabarkapa, Philion, and Yu (2022)

D'Urso, De Giovanni, and Vitale (2022)

3.) How does their research question nest with the existing research on this topic?

The research does indicate key statistics. However, so some sources appear to be irrelevant (the 1994) or aren't detailed enough (the D'Urso source). Otherwise the research question dues nest pretty with existing research.

4.) Where is the author getting their data from? Is it a trustworthy source?

From basketballreference. Appears to have stats from every season and is easily accessible to the public.

5.) Identify at least 3 errors in formatting, punctuation, spelling, wrong-word use, passive voice, or general errors.

or the document

I understand Latex was used in generating this, but figure out how to reduce margin size down such that the project proposal can fit in two pages.

Name: Thao Le

Author's Name:

Author's manuscript title: Ba - Missing title

1.) The research question is:

- What statistic are most important in influencing a team's winny percentage?
- How accurately can our selected mode predict a team's winny %?

2.) Identify the three sources that the author clearly analyzed thoroughly.

- 1) Game statistic that discriminate winning and losing at the nba level of basket ball competition
- 2) Take that game! an analysis of winny % for nba teams
- 3) A bayesian network to analyse basket ball players' performance; a multivariate copula-based approach.

3.) How does their research question nest with the existing research on this topic?

Sport - basketball ~~can~~ data research provide overall view of what factor can be helpful to predict the chance of winny game  
→ support their 2 questions

4.) Where is the author getting their data from? Is it a trustworthy source?

- Data from basketball reference - public data.
- It is unclear to know if the data is trustworthy or not with the description & how the data was formed.

5.) Identify at least 3 errors in formatting, punctuation, spelling, wrong-word use, passive voice, or general errors.

- Citation paper name, some part need to re-check about capital letter
- Should have the same margin for all paragraph (3 & 2 section align)
- " " are not quite right (3-in "Bubble" covid section)
- Input title

Name: Tom Holley

Author's manuscript title:

Misgng

Author's Name: Tom / casey

1.) The research question is:

what factors best explain a team's success, as determined by winning percentage?

2.) Identify the three sources that the author clearly analyzed thoroughly.

[cabarkapa et al., 2022]

[Chatterjee et al., 1994]

[D'Urso et al., 2022]

3.) How does their research question nest with the existing research on this topic?

other research has been done on winning percentage for each game, and winning percentage for each player. Their study will look at team winning percentage in current years.

4.) Where is the author getting their data from? Is it a trustworthy source?

The seem trustworthy, but I don't know where they came from. Includes article title but not the journal or where found. DATA from Basketball ref. <sup>yes</sup> rivalry.

5.) Identify at least 3 errors in formatting, punctuation, spelling, wrong-word use, passive voice, or general errors.

- I want to know more about where these sources were found.

- The alternative Hypothesis can use f instead of writing at does not equal.

Name: Paige Lai Bolt

Author's Name: Casey, Tom

Author's manuscript title:

1.) The research question is:

- What statistics are most important in fluencing a team's winning percentage?
- How accurately can our selected model predict a team's winning Percentage?

2.) Identify the three sources that the author clearly analyzed thoroughly.

- "Game stats that discriminate win/loss at NBA level"
- "Winning Percentages for NBA Teams
- "A bayesian network ... in a multivariate copular-based approach."

3.) How does their research question nest with the existing research on this topic?

A study found that assists and steals were not found statistically significant but with comparing teams. The research will be with the team themselves.

4.) Where is the author getting their data from? Is it a trustworthy source?

They are getting their data from "basketball reference" and all it says is public data so not sure if trustworthy.

5.) Identify at least 3 errors in formatting, punctuation, spelling, wrong-word use, passive voice, or general errors.

- on page 2, first paragraph, "since all 32..." change to "so all 32..."
- In section Data, first two sentences could be combined.
- During the Literature Review the word "however" could be deleted in many spots.

Name: will uroening

Author's Name: Casey Bear,  
Tom Holler

Author's manuscript title:

(N/A) - Add

1.) The research question is:

what are the key statistical performance factors to success in the NBA. ( which performance stat have the strongest linear relationships with team winning percentage )

could contradict?

\* How would you define "today's" NBA.

2.) Identify the three sources that the author clearly analyzed thoroughly.

→ Chatterjee, Campbell, and Wiseman (1944) ⇒ might be typo

Cabarchaga, Deane, ... , Philipp (2022)

D'Urso, ... Vitalz (2022)

3.) How does their research question nest with the existing research on this topic?

The goal is to understand the performance metrics, and data presented from Sanders to understand team strategic focuses, team record for success and other related actions.

4.) Where is the author getting their data from? Is it a trustworthy source?

- Basketball Reference, Yes.

\* Clark may want a few more years of data via reference ⇒ that's what he suggested from my project using Basketball Reference

5.) Identify at least 3 errors in formatting, punctuation, spelling, wrong-word use, passive voice, or general errors.

- Wiseman is it published 1944 or 1994 via lit review

- might want to indent or enter down

\* 2019-2020 "Bubble"

- Spacing error in Data I circled.

FGWe

Name: Casey Gay  
Author's manuscript title:

Author's Name: Casey Gay  
Tom Holley

N/A (fix this)

1.) The research question is:

(i) What statistics are most important in influencing a team's winning percentage?

(ii) How accurately can the selected model predict a team's

2.) Identify the three sources that the author clearly analyzed thoroughly.

(i) Game statistics that discriminate...

(ii) Take that jam! An analysis of winning...

(iii) A bayesian network to analyze basketball...

3.) How does their research question nest with the existing research on this topic?

Different studies have considered very similar questions with slightly different applications, finding mixed results. The second question was not explicitly addressed.

4.) Where is the author getting their data from? Is it a trustworthy source?

Basketball Reference.com  
Public data, cited in other studies, strong reputation

5.) Identify at least 3 errors in formatting, punctuation, spelling, wrong-word use, passive voice, or general errors.

I do not notice any errors. The citation style may not be completely correct, so we will look into this further.