**Table of Contents**

1. **Project Description**
2. **Data Preparation and Exploration**

* **2.1 Data information**
* **2.2 Brief Exploration**
* **2.3 Data Transformations**
* **2.4 Bar Chart**
* **2.5 Plot Chart**
* **2.6. Density Plot Chart**

1. **Modeling**

* **3.1. Significant/Insignificant Coefficients**
* **3.2. Predictive Performance**
* **3.3 Accuracy, Recall, Precision**

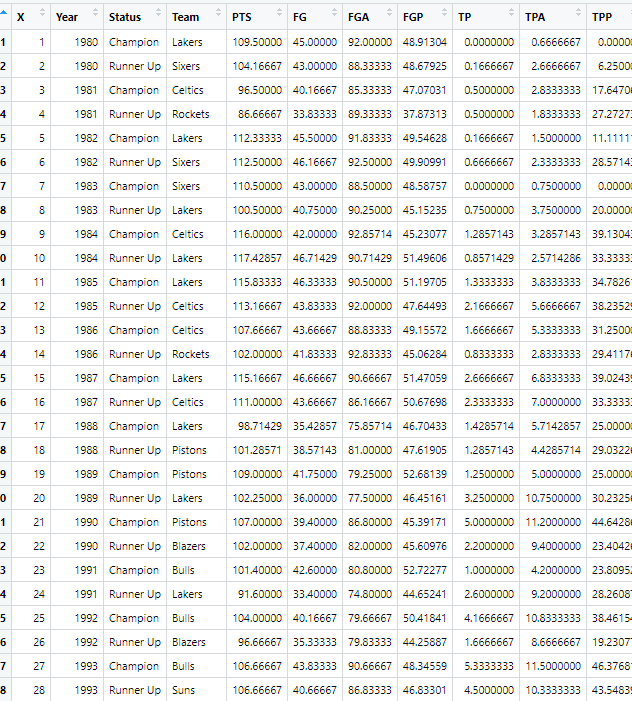
1. **Results and Discussion**
2. **Conclusion**
3. **Project Description**

Conventionally, in the sport of basketball there are defined player positions with standard body profiles. In normal terms, the small players would be able to dribble well and be able to pass. The slightly taller shooting guards would be able to shoot jump shots well (hence their name) like from beyond the 3-point line. Then the incrementally bigger sized small forwards would be stronger and be able to drive to the hole, and the centers would only shoot from the paint (dunk or layup), and hustle for rebounds. However, the style of the game has changed as bigs should now be able to shoot and the value of the 3-point shot is at an all-time high. The lines have been increasingly blurred in what is called a position-less basketball era attributed to the success of the Golden State Warriors in the 2010’s. Coaches have implemented a style called “small ball” that maximizes spacing by positioning smaller and more quicker players beyond the three-point line and maximizing three-point shooting. The business problem is how the 3-point shot affects NBA managers around the association in whether they should be pressured to build rosters with players that maximize 3-point shooting, and whether it correlates to winning championships.

**2. Data Exploration and Preprocessing**

**2.1.** The data was retrieved from [Kaggle](https://www.kaggle.com/daverosenman/nba-finals-team-stats) by the author named Dave Rosenman who made the dataset. He collected the data from [basketball-reference.com](https://www.basketball-reference.com/). The data champsdata.csv contains the total game-by-game team for the final championship team between 1980 and 2017. The data runnerupsdata.csv contains the total game-by-game team for the final runner-up team between 1980 and 2017.

**2.2.** Below provides a brief exploration of the data from the ‘champsandrunnerupsseries\_averages.csv' file.



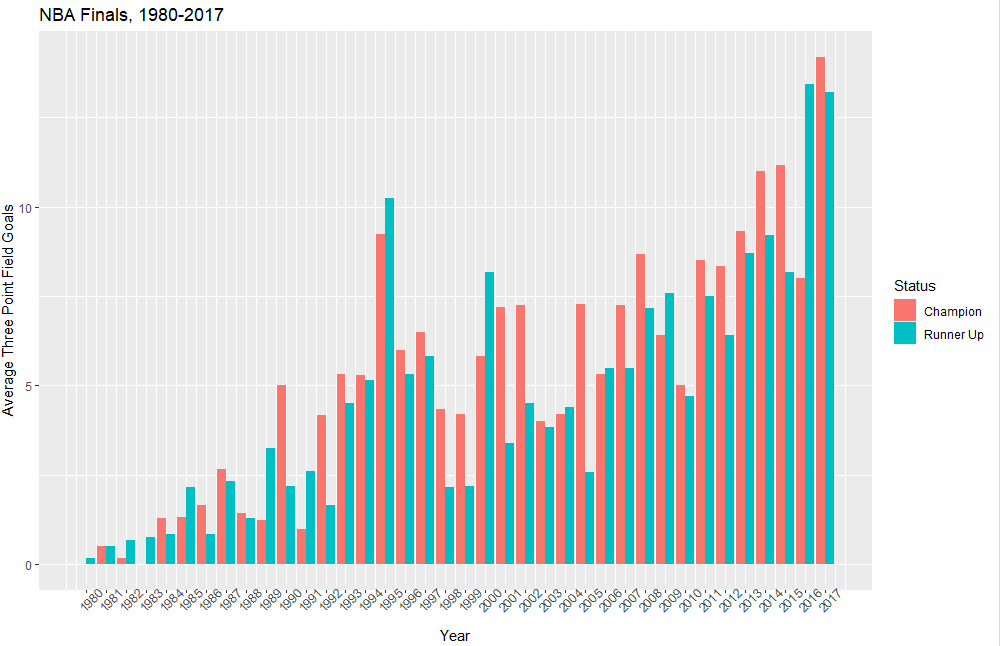
**2.3.** No records appear to be missing. No redundant data either. However, the following transformations were made:

1. Convert Year column to integer type
2. Fixed name issue with ‘Warriorrs’ and 'Heat'. Turn Win and Home into factors. Did the same with the runner’s data.

Combine champs and runners with a row bind, then partitioned 75-25 into a training and test set

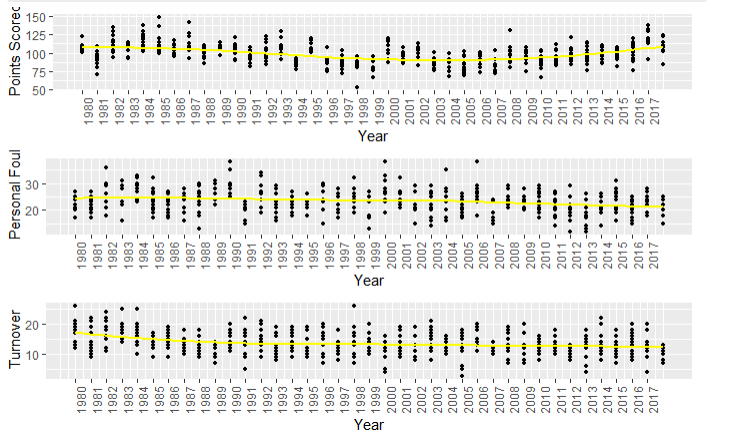
1. Check the team names once more to make sure no duplicates or misspellings.

**2.4.** The chart below illustrates the three-point field goals for NBA finals teams from 1980 onwards.



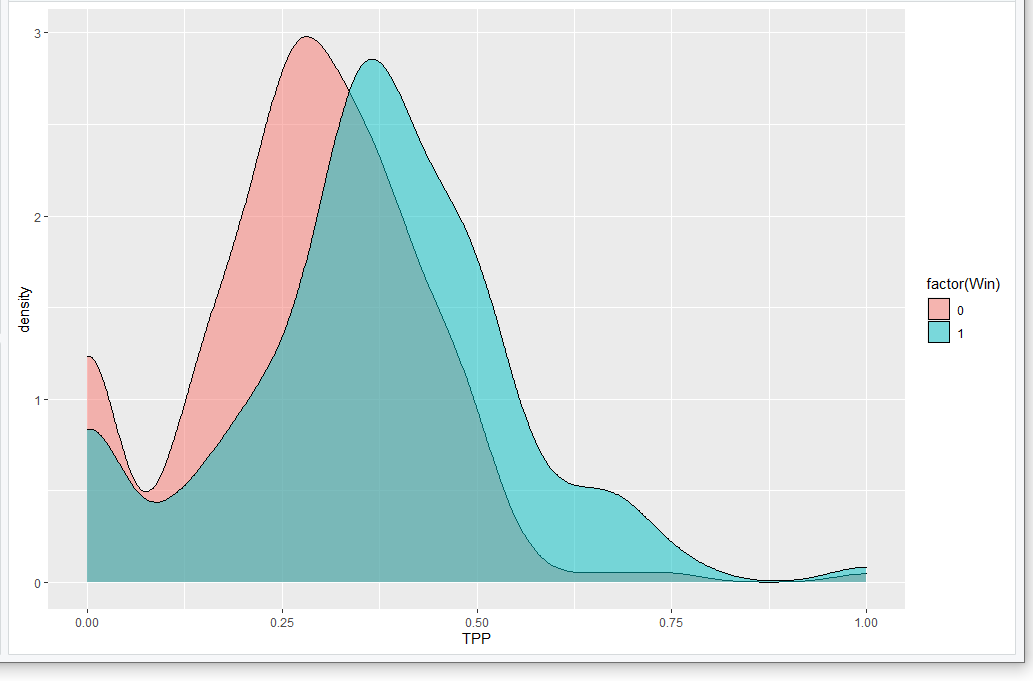
**2.5.**

From the plots of field points below, we can see the results shows that the percentage of field goal points of the total points are decreasing.



**2.6. Win vs. TPP**

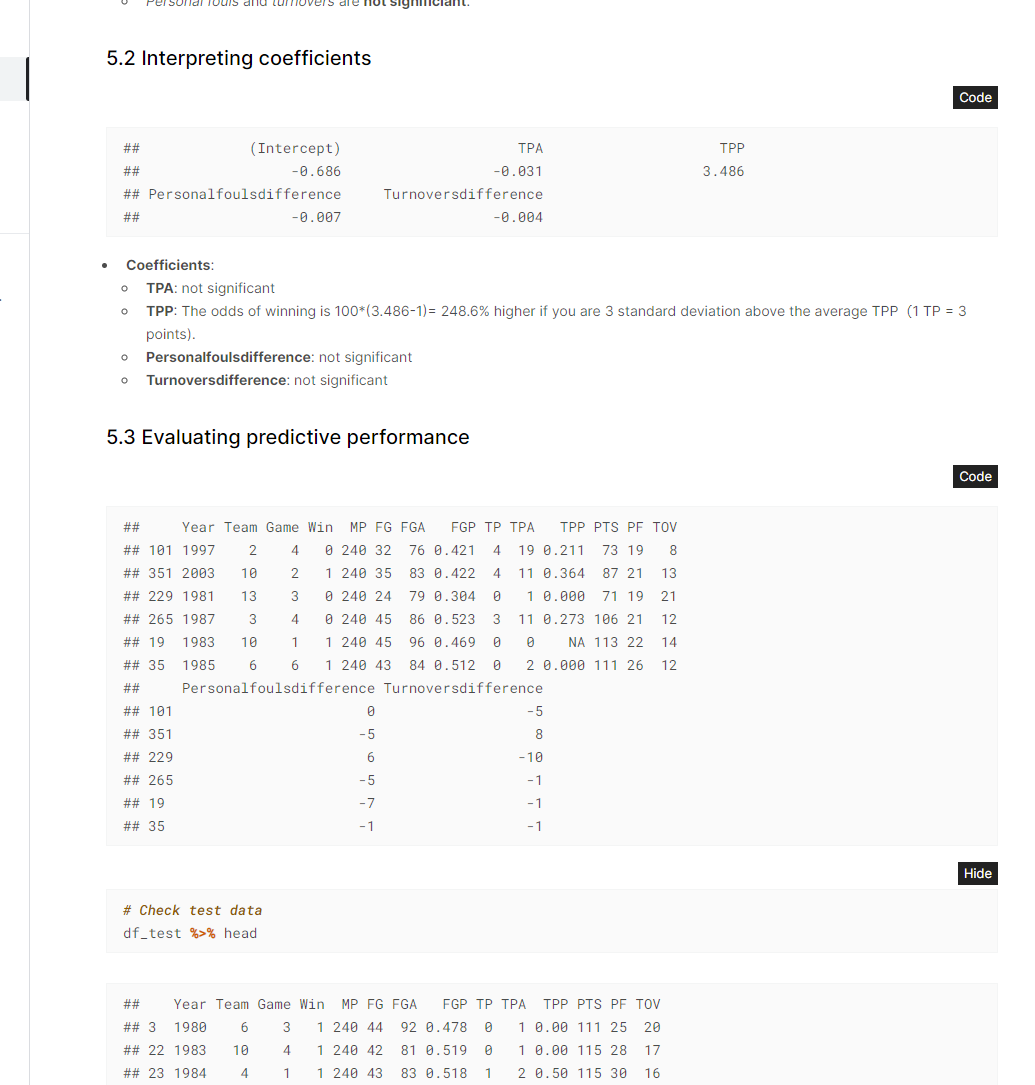
In this next chart I will look at the density plot of Winning versus TPP (Three Point Percentage). Since TPP are continuous, density plot was the choice.



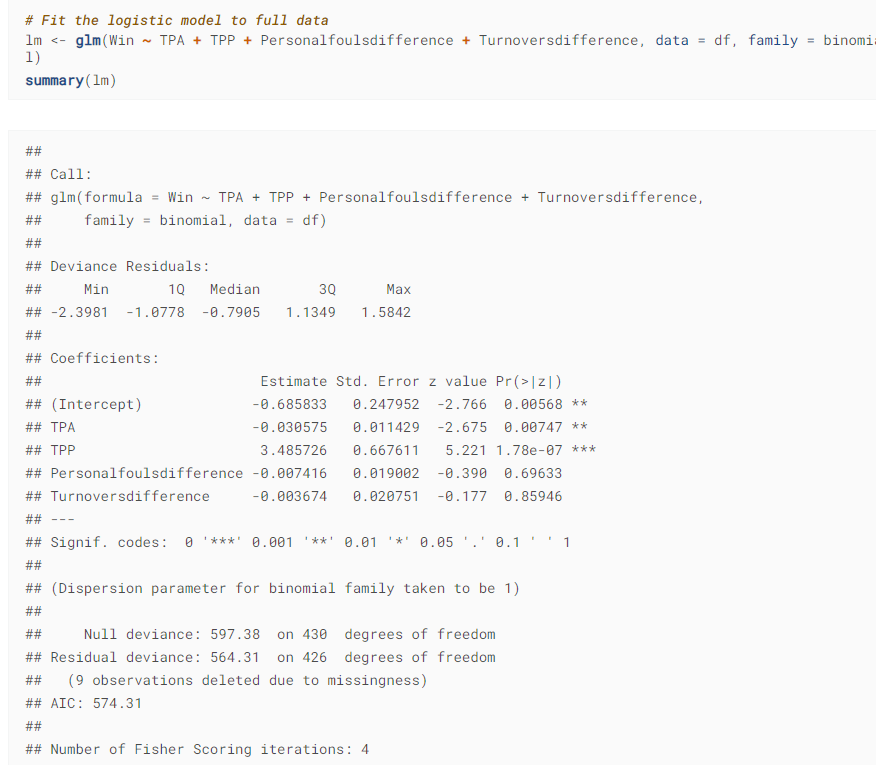
1. **Modeling**

For the modeling part I will use logistic regression to predict how impactful three pointers are to win. In my starting point logistic model, I began with three predictors TPA, and TPP which I thought would be of importance from the start.

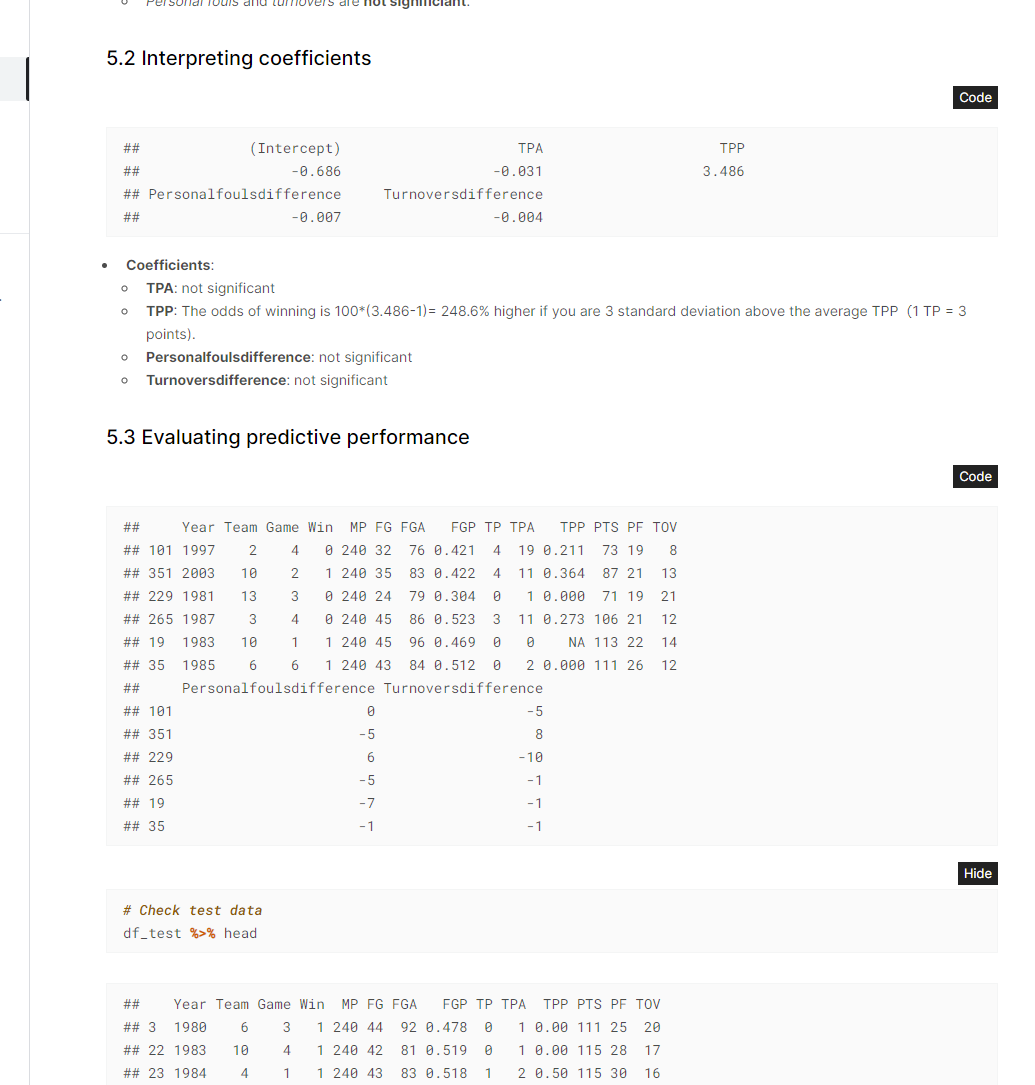
**3.1**



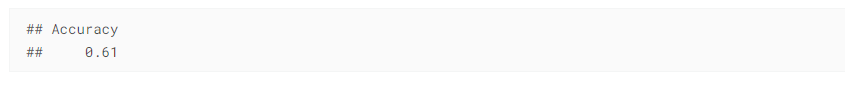
**3.2**



**3.3**



**3.4**







* 61% of the accuracy of the winning forecasts on the test set is accurate.
* The model properly identified 60% of the winning teams in the test set.
* When the model forecasts winning, it is right 62% of the time.

1. **Results and Discussion**

Based on the charts and modeling above, we can determine that in the NBA finals, teams are changing their strategy of scoring from field points to three points because the trend of three-point field goals is increasing (2.4). However, from the plots of PF (personal fouls), TOV (turnovers), we can see the trend is going down, which means the fouls and turnover are less than before (2.5). With the number of fouls and turnovers decreasing, it could be that the NBA is gravitating towards a less physical playstyle. The three-point line gives teams the ability to score more efficiently other than free-throws and two-point field goals for scoring without having to be as physical. Based on my linear regression model, the odds of winning highly correlates with TPP (Three-point percentage) is significantly higher (100\*(3.486-1)= 248.6% higher) (3.1). This statistic is significant showing that the three-point shot is of very high value. Also, based on the density plot, the two histograms comparing three-point percentage and winning seem to be correlated.

1. **Conclusion**

The trend of more three-point percentage of the total points is increasing, so NBA teams are changing their championship strategy to build teams with better three-point shooters and increase the volume of three-point shots versus two-point field goals. Also, personal fouls and turnovers are not significant to winning the NBA final series. Lastly, three points percentage is a significant variable to winning the NBA final series, so to answer the business question on whether teams should invest in three-point shooters? Yes, they should.