

Using the four-factor model discussed in Chapter 28 of Mathletics, the Houston Rockets had the worst record in the NBA in the 2020-2021 season, finishing with only 17 wins. The four-factor model consists of metrics to determine how well a team is performing on the offensive and defensive sides of the ball. The data was collected by copying and pasting advanced team stats from [basketball-reference.com](https://www.basketball-reference.com) for the 2021 and 2022 season. These data sets were then filtered for the offensive and defensive four-factors, and each team was ranked in each category. It was evident after examining the datasets that the teams with the most wins were in the top half for most of the categories and excelled in a couple of them. We were 24th in effective field goal percentage, 22nd in turnover percentage, 28th in offensive rebound percentage, 18th in free throws per field goal attempt, 27th in opponent effective field goal percentage, 22nd in defensive rebound percentage, and 23rd in opponent free throws per field goal attempt. Surprisingly we were 10th in opponent turnover percentage. We were in the bottom half of all of the offensive factors, meaning that we weren't taking good shots, taking care of the ball, or coming down with offensive rebounds. On the defensive side, even though we were above average in forcing turnovers, we still gave up too many good looks, rebounds, and fouls. This year we are still tied in last place, with only 15 wins so far. We are 16th in effective field goal percentage, 30th in turnover percentage, 21st in offensive rebound percentage, 30th in opponent effective field goal percentage, 14th in opponent turnover percentage, 25th in defensive rebound percentage, and 28th in opponent free throws per field goal attempt. We were however 3rd in free throws per field goal attempt. We improved in effective field goal percentage, offensive rebound percentage, and opponent turnover percentage but we are still average/below average in those factors compared to the rest of the league. We also improved dramatically in free throws per field goal attempt, as we are currently one of the top teams in that category. However, we declined in all of the factors, including finishing last in opponent effective field goal percentage and turnover percentage. This shows we still need a lot of help in all areas on the offensive and defensive sides of the ball.

## NBA 2020-2021

Team	W	OeFG%_Rank	OTOV%_Rank	ORB%_Rank	OFT/FGA_Rank	DeFG%_Rank	DTOV%_Rank	DRB%_Rank	DFT/FGA_Rank
Utah Jazz*	52	5	19.5	3	10.5	1	30	5	2
Los Angeles Clippers*	47	3.5	14.5	12.5	19.5	8	18.5	6	10.5
Phoenix Suns*	51	3.5	4.5	24.5	26.5	10.5	16	9	14
Milwaukee Bucks*	46	2	11.5	9	26.5	13	24.5	3.5	1
Philadelphia 76ers*	49	14	21	10	2	3	2.5	11.5	21.5
Denver Nuggets*	47	6	13	2	28.5	21	10.5	7	21.5
Brooklyn Nets*	48	1	14.5	19	5	8	28.5	20.5	12
Los Angeles Lakers*	42	18	27.5	14	7	5	4	3.5	9
Dallas Mavericks*	42	8	3	23	16	10.5	24.5	16.5	19
New York Knicks*	41	23	8.5	16.5	15	2	21	8	17
Atlanta Hawks*	41	16	8.5	4	1	6	28.5	16.5	13
Portland Trail Blazers*	42	15	1	11	10.5	22	27	19	25
Boston Celtics*	36	12	18	5	22.5	15.5	12	13	27
Golden State Warriors	39	7	24.5	30	17	4	7.5	25.5	29
Memphis Grizzlies*	38	22	6	6	25	12	7.5	10	17
Miami Heat*	40	11	23	29	8	19	2.5	14.5	7.5
Indiana Pacers	34	13	8.5	26	24	8	13	30	23.5
New Orleans Pelicans	31	17	19.5	1	4	25.5	18.5	2	10.5
Toronto Raptors	27	21	8.5	24.5	9	20	1	27	30
Chicago Bulls	31	10	27.5	15	30	14	22	1	17
San Antonio Spurs	33	27	2	27	14	17.5	20	20.5	6
Washington Wizards*	34	20	16	20.5	3	15.5	15	18	28
Charlotte Hornets	33	19	24.5	8	22.5	25.5	6	28	3
Sacramento Kings	31	9	11.5	20.5	21	30	17	29	20
Detroit Pistons	20	26	26	16.5	6	17.5	14	23	15
Minnesota Timberwolves	23	25	17	12.5	13	28.5	5	24	26
Houston Rockets	17	24	22	28	18	27	10.5	22	23.5
Cleveland Cavaliers	22	29	29	7	12	28.5	9	25.5	7.5
Orlando Magic	21	30	4.5	18	19.5	23.5	24.5	11.5	5
Oklahoma City Thunder	22	28	30	22	28.5	23.5	24.5	14.5	4

## NBA 2021-2022

Team	W	OeFG%_Rank	OTOV%_Rank	ORB%_Rank	OFT/FGA_Rank	DeFG%_Rank	DTOV%_Rank	DRB%_Rank	DFT/FGA_Rank
Phoenix Suns	49	4.5	6	18.5	23	3	8	14	10
Golden State Warriors	43	2.5	28	16.5	21	2	4	1.5	18.5
Utah Jazz	38	1	21	8	2	7.5	29	9	1
Boston Celtics	36	16.5	17	11	11	1	11.5	13	12
Miami Heat	41	7.5	27	6	5	14	3	8	25.5
Memphis Grizzlies	43	24	4	1	19	11	6	10.5	16
Cleveland Cavaliers	36	9	29	9	15	4	9.5	20	2
Milwaukee Bucks	37	6	10.5	15	9	15	23.5	10.5	6
Dallas Mavericks	36	12	8	24	18	7.5	11.5	6	9
Denver Nuggets	36	2.5	25.5	23	24.5	17	21	3	11
Philadelphia 76ers	37	13.5	6	30	1	6	17.5	17	13
Chicago Bulls	39	4.5	9	27	7	18.5	22	6	16
Toronto Raptors	33	25	3	2	27	22	1	26.5	24
Minnesota Timberwolves	33	16.5	14.5	4	17	18.5	2	30	30
Atlanta Hawks	29	7.5	2	16.5	7	24	28	12	7
San Antonio Spurs	24	16.5	1	10	30	20	17.5	26.5	5
Los Angeles Clippers	32	19.5	23.5	28.5	27	5	16	29	4
Charlotte Hornets	30	10	6	14	27	25.5	7	28	14
Brooklyn Nets	32	13.5	13	12.5	13	11	27	24	18.5
Los Angeles Lakers	27	11	23.5	26	21	16	9.5	21	16
New York Knicks	25	27	14.5	7	7	11	25.5	6	20.5
New Orleans Pelicans	25	26	21	3	3.5	25.5	13	4	22
Indiana Pacers	21	19.5	19	5	21	23	25.5	15	27
Washington Wizards	27	21	12	25	11	9	30	18	25.5
Sacramento Kings	23	22.5	17	20	11	27	23.5	22.5	8
Portland Trail Blazers	25	22.5	17	18.5	15	29	14.5	1.5	23
Oklahoma City Thunder	19	29	10.5	22	29	13	19.5	19	3
Orlando Magic	15	28	25.5	28.5	24.5	21	19.5	16	20.5
Houston Rockets	15	16.5	30	21	3.5	30	14.5	25	28
Detroit Pistons	15	30	21	12.5	15	28	5	22.5	29

A multiple linear regression was run on the data to determine the relative importance of each of the four factors. This was the appropriate method as it explains how much of an influence each factor has on the number of wins. Four factors were created and selected for the model. They were created by taking the differences between the offensive four factors and their defensive counterparts (e.g. effective field goal percentage - opponent field goal percentage). After inputting the factors into the regression model and setting the response variable to wins, the following formula was calculated:

**Games\_won = 49.84 + 340.88(Effective Field Goal Percentage Difference) - 231.98(Turnover Percentage Difference) + 24.86(Rebound Percentage Difference) + 102.63(Free Throws Per Field Goal Attempt).**

Call:

```
lm(formula = Four_Factors_2020_2021$W ~ Shooting_Dev + Turnover_Dev +
    Rebound_Dev + Free_Throw_Dev)
```

Residuals:

Min	1Q	Median	3Q	Max
-7.8082	-1.7138	0.3538	1.2505	7.1540

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	49.84	23.08	2.160	0.04060	*
Shooting_Dev	340.88	25.83	13.200	9.08e-13	***
Turnover_Dev	-231.98	63.92	-3.629	0.00127	**
Rebound_Dev	24.86	41.49	0.599	0.55444	
Free_Throw_Dev	102.63	33.61	3.054	0.00531	**

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.684 on 25 degrees of freedom

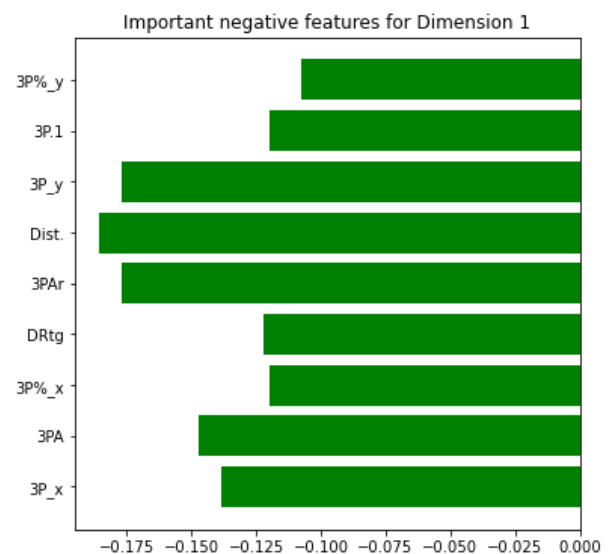
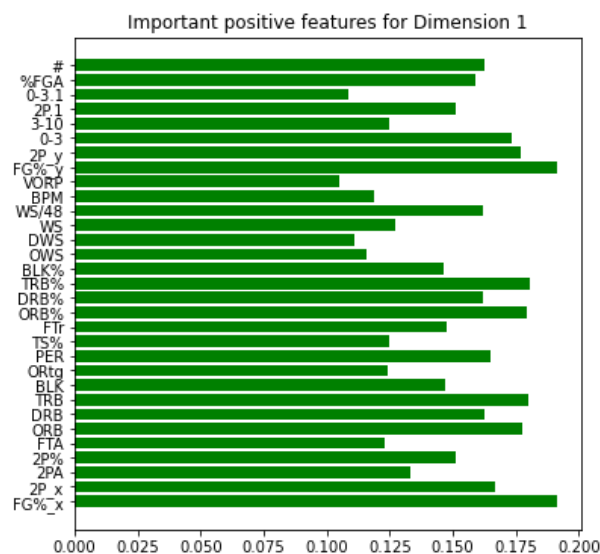
Multiple R-squared: 0.8833, Adjusted R-squared: 0.8646

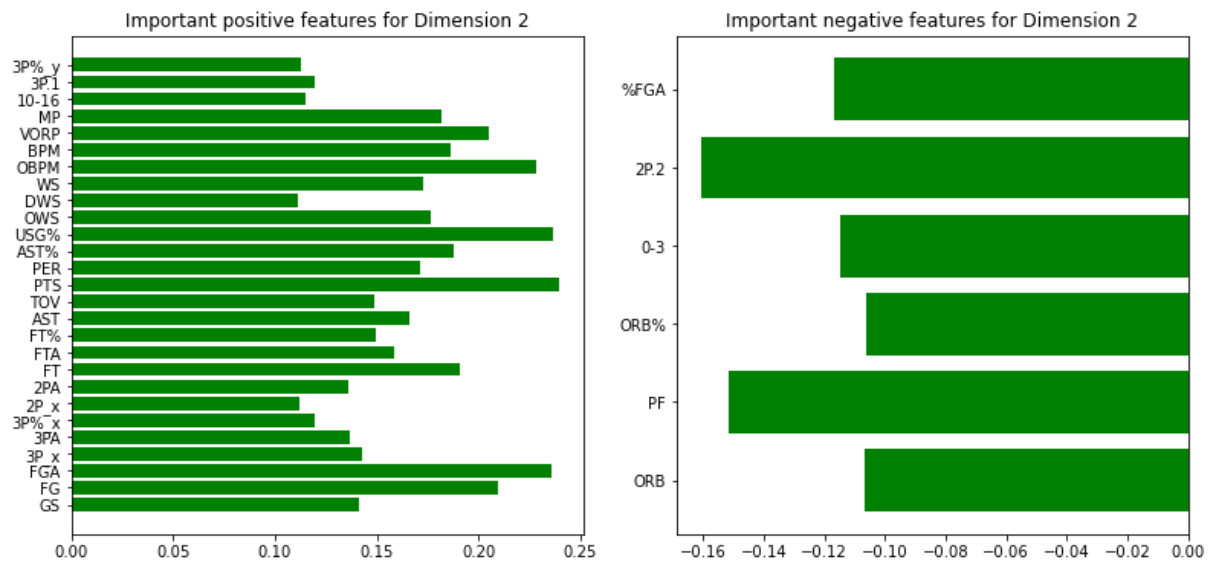
F-statistic: 47.31 on 4 and 25 DF, p-value: 2.622e-11

The results of the regression model indicate that all of the factors are considered significant in predicting win totals except for rebound percentage difference. Effective field goal percentage difference has the greatest relative influence on number of wins followed by, turnover percentage difference, free throws per field goal attempt, and rebound percentage difference. The Houston Rockets' data for the 2020-2021

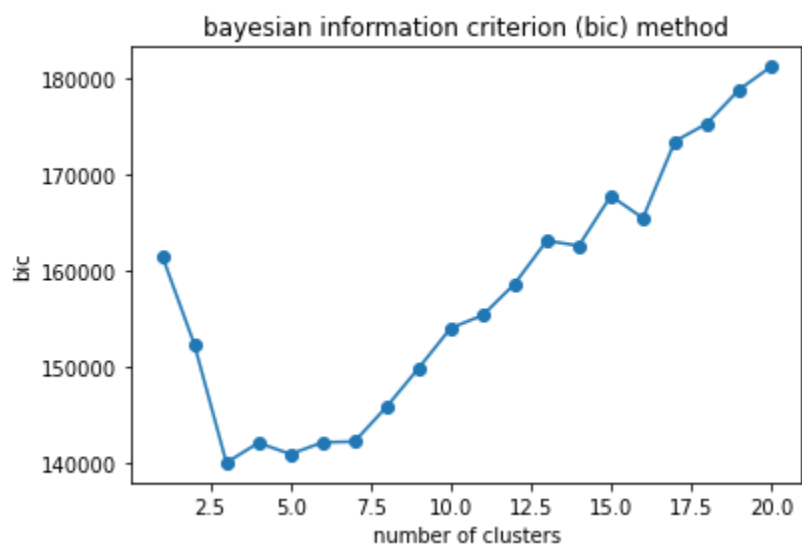
season was input into the regression model and it was predicted that we should have won 22 games last year. This is slightly more than our actual win total last season, indicating that our regression model is fairly accurate.

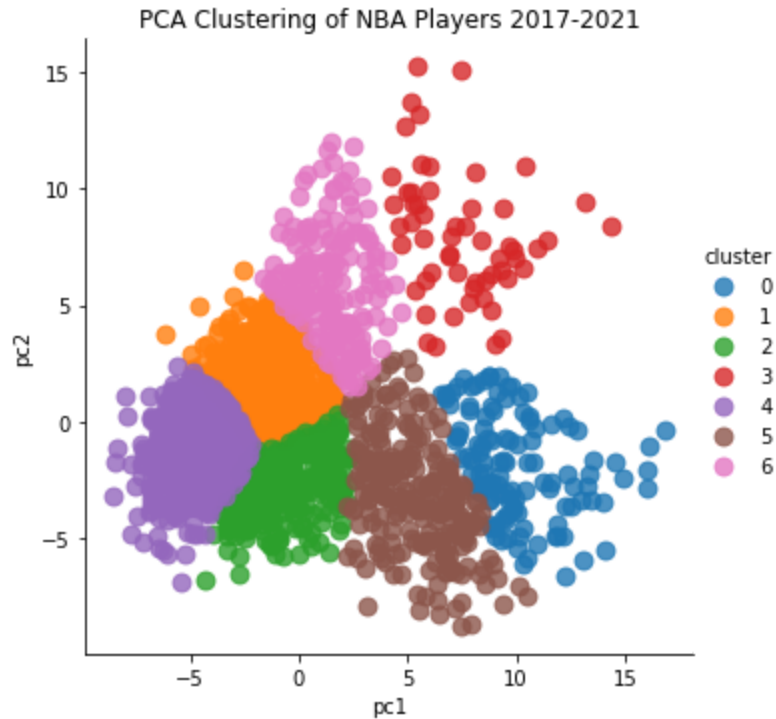
Using historical data, a new classification framework was created for identifying player type. A script was written to scrape data from basketball-reference.com from the past 5 seasons (not including current season). The player data that was obtained was player per 100 poss stats, advanced stats, and shooting stats, and they were all merged into one dataset. All of the duplicate columns and columns filled with 'NA' were dropped from the dataset as well. The data was standardized and filtered to select for players that played in at least 40 games to get an accurate representation. The methodology for the model was based off of Alex Chang's model in the article, "Using Machine Learning to Find the 8 Types of Players in the NBA". Principal Component Analysis (PCA) was first performed on the cleaned data set to reduce the dimensions down to 2. The important positive and negative features for each dimension were then obtained.





Next, a Gaussian Mixture model was developed and fit to the principal components to identify clusters for normally distributed subpopulations within the dataset. Using the Bayesian Information Criterion method, it was determined that 7 clusters would be represented on the dataset. After the principal components were fit to the final model, the NBA players in the dataset were categorized based on the cluster they were in.





#### **Cluster 0 - Efficient Traditional Center**

This cluster consists of players that had the highest average for rebounds, points near the basket, field goal percentage, and blocks. These players are known for scoring easy buckets, protecting the rim, and cleaning up on the boards. Notable examples are Rudy Gobert, Clint Capela, and Jarret Allen.

#### **Cluster 1 - Sharpshooters**

The players in this cluster have the second highest average for 3 point attempts and free throw percentage, and the highest average for 3 point percentage. They are good shooters that mainly catch and shoot to space out the floor. Notable examples are Joe Ingles, Buddy Hield, and Joe Harris.

#### **Cluster 2 - Ball Dominant Scorers**

Ball dominant scorers have the highest average for field goal attempts, free throw percentage, assists, and usage percentage. These players are able to score from anywhere on the court and are able to draw a lot of fouls to get to the free throw line. With their high usage, they score a lot of points, dish out a lot of assists, and have a high Value Over Replacement Player(VORP). The negative of having a high usage, is

that they have a higher turnover rate as well. Notable examples are Trae Young, Stephen Curry, and Luka Doncic.

### **Cluster 3 - Low Usage Role Players**

This cluster of players have the lowest average for usage percentage, field goal percentage, and VORP.

These are mainly role players that don't touch the ball a lot and mainly shoot 3 pointers when they do.

They're not as good shooters as the sharpshooters cluster, but they make up for it with defense, which is not as easily quantified. Notable examples are Luguentz Dort, P.J. Tucker, and Patrick Beverley.

### **Cluster 4 - Stretch Bigs**

Players in this cluster don't excel at any particular statistic but they compose of players that average an above average number of rebounds and blocks, and an average number of 3 point attempts and 3 point percentage. They are usually big men that can stretch the floor and shoot. They score a majority of their points near the rim and behind the 3 point line, and over 94% of their 3 pointers are assisted. Notable examples are Frank Kaminsky, Kristaps Porzingis, and Bobby Portis.

### **Cluster 5 - Low Efficiency Rim Protectors**

These players have the second highest average for rebounds and blocks. They score a majority of their points near the rim but their field goal percentage isn't as high as the efficient traditional centers. These players are normally big men that are less efficient than starters so they are usually back ups. Notable examples are Tristan Thompson, Daniel Theis, and Taj Gibson.

### **Cluster 6 - High Usage Versatile Bigs**

High Usage Versatile Bigs are similar to ball dominant scorers in that they have one of the highest averages for usage percentages, points, field goal attempts, free throw attempts, VORP, and turnovers. In addition, they average a high number of rebounds and blocks, as well as an average number of assists. They are capable of scoring anywhere on the court but they shoot a majority of the shots in the paint. Notable examples are Giannis Antetokounmpo, Anthony Davis, and Joel Embiid.

The players that currently make up our starting 5 are:

Dennis Schroder – Sharpshooter

Kevin Porter Jr. – Low Usage Role Player

Jae'Sean Tate – Stretch Bigs

Christian Wood – Stretch Bigs

Jalen Green – N/A(Rookie)

The top 5 lineups with the highest plus/minus and that have played in at least 20 games this season are:

**Boston Celtics**

Jayson Tatum – Ball Dominant Scorer

Jaylon Brown – Ball Dominant Scorer

Marcus Smart – Sharpshooter

Al Horford – Stretch Big

Robert Williams – Efficient Traditional Center

**Memphis Grizzlies**

Steven Adams – Low Efficiency Rim Protector

Kyle Anderson – Stretch Big

De'Anthony Melton – Sharpshooter

Ja Morant – Ball Dominant Scorer

Zhaire Williams – N/A(Rookie)

**Utah Jazz**

Rudy Gay – Sharpshooter

Mike Conley – Ball Dominant Scorer

Rudy Gobert – Efficient Traditional Center



Jordan Clarkson – Sharpshooter

Joe Ingles – Sharpshooter

### **Milwaukee Bucks**

Jrue Holiday - Ball Dominant Scorer

Khris Middleton - Ball Dominant Scorer

Giannis Antetokounmpo - High Usage Versatile Big

Bobby Portis - Stretch Big

Grayson Allen – Sharpshooter

### **Minnesota Timberwolves**

Patrick Beverley - Low Usage Role Player

D'Angelo Russell - Sharpshooter

Karl-Anthony Towns - High Usage Versatile Big

Jarred Vanderbilt - Low Efficiency Rim Protector

Anthony Edwards - Sharpshooter

After analyzing the most efficient lineups, they all contain at least one sharpshooter and a ball dominant scorer or high usage versatile big. These appear to be the common denominator in all of these teams.

Whenever a ball dominant scorer is in the lineup, they are usually paired with one of the big men that can grab a lot of rebounds(Efficient Traditional Center, Low Efficiency Rim Protector, High Usage Versatile Big). The Rockets need help in every area but to start it would be beneficial to add a ball dominant scorer as we are lacking one. This will help our effective field goal percentage and opponent turnover percentage. The Rockets hope Jalen Green develops into this type of player. In addition, the Rockets could look to add an efficient traditional center to help with rebounding and effective field goal percentage. A limitation of this classification framework is that it doesn't take into consideration defense

as much as offense. Offense is easily quantified in basketball because there are numerous stats to analyze but defense is more difficult since there aren't as many notable stats outside of blocks and steals. Future research would involve obtaining defensive stats that aren't as readily available such as shot contest rate, points scored on, and turnovers forced. Obtaining these stats should improve the classification framework and uncover clusters that didn't exist before.

## References

*Basketball statistics and history.* Basketball. (n.d.). Retrieved March 11, 2022, from <https://www.basketball-reference.com/>

Chang, A. (2017, March 2). Using Machine Learning to Find the 8 Types of Players in the NBA. Retrieved March 11, 2022, from <https://medium.com/fastbreak-data/classifying-the-modern-nba-player-with-machine-learning-539da03bb824>.

*Official NBA stats.* NBA Stats. (n.d.). Retrieved March 11, 2022, from <https://www.nba.com/stats/>

Winston, Wayne L. 2012. *Mathletics: How Gamblers, Managers, and Sports Enthusiasts Use Mathematics in Baseball, Basketball, and Football*. Princeton, NJ: Princeton University Press.

