NBA Play Style Analysis between Decades

From 1990 to 2019

By Fengsheng Zhou, Jinghong Chen, Yihan Shi, Zhihong Li

Abstract

Basketball is a sport that originated in the United States and is often played indoors on a rectangular court with two teams of five players. Each team attempts to score by putting the ball through the opponent's basket. The National Basketball Association was founded later in the United States (NBA). After 60 years of evolution, we can see that each era of the NBA has its own distinct style of play. With new players appearing and tactical play evolving, today's league is getting more competitive. The league's approach favored physical confrontation and a traditional position game in the 1990s. There are several ways to score including layups, jump shots, free throws, 3-pointers and so on. Today, More teams are a greater focus on three-point shooting and box scores. Since extraordinary heroism does not win a team the championship, today's teams are more focused on the team's overall development. For previous NBA competitions, scores of 100 points were regarded as high in the 1990s to 2000s, and most of the time 80 to 90 points were about the average score. However, The NBA team now focuses on playing fast, and Most teams solve around in ten to fifteen seconds. So, the pace is quick, and teams are more concerned with shooting three points. We can see that the level of physical contact and muscle collision is to identify a good team and a team from the 1990s to the 00s. However, the level of physical contact began to decline after the 2010s, and they paid more attention to improve shooting skill. In other words, it's due to different competition styles.

Introduction

The current game is radically different from the early game style and rhythm, which is also quite different from the game 10 years ago, due to the evolution of basketball and changes in rules, athletes' physical quality, and technical features. For example three-point shooting has gotten a lot of attention lately, and players need to be more balanced in their approach. And for early years players tend to focus more on defense and two-point shooting. In this project we are interested in finding the difference of play style and their trend. We can get our data from the NBA has a comprehensive database that contains the most stats of each game, https://www.nba.com/ API. We collect data for each game by team from 1990 to 2019 and further investigate.

Description of dataset

• Source

- https://stats.nba.com/stats/leaguegamelog
- Attribute Information:

1. SEASON_ID — Game period
2. TEAM_ID — Team number
3. TEAM_ABBREVIATION —- Team appellation
4. TEAM_NAME —- Team name
5. GAME_ID —- Game ID
6. GAME_DATE —- Competition days
7. MATCHUP —- strengths and weaknesses
8. WL —- Win and Lose
9. MIN —- Minutes
10. FGM —- Team's Field Goals Made
11. FGA —- Team's Field Goals Attempted
12. FG_PCT —- Field goal percentage
13. FG3M —- Team's Field Goals Made 3 Pointers
14. FG3A —- Field Goals Attempted 3 points
15. FG3_PCT —- three-pointer percentage
16. FTM —- Team's Free Throws Made
17. FTA —- Team's Free Throws Attempted
18. FT_PCT —- free throw percentage
19. OREB —- Team's Offensive Rebounds
20. DREB —- Team's Defensive Rebounds
21. REB —- Team's Rebounds
22. AST —- Team's Assists
23. STL —- Team's steal
24. BLK —- Team's block
25. TOV —- Team's Turnovers
26. PF —- Personal Fouls
27. PTS —- Points
28. PLUS_MINUS —- Plus-Minus
29. VIDEO_AVAILABLE —- VIDEO AVAILABLE

Through web surfing, we consider that is the most efficient way for the project since we scrape the various information that we need from the NBA official website. **We used API techniques to extract data for each basketball team from 1990 to 2019.** Our dataset contains 29 variables, which are shown as above.

• Preview of the dataset

REVIATION	TEAM_NAME	GAME_ID	GAME_DATE	MATCHUP	WL	MIN	FGM	...	REB	AST	STL	BLK	TOV	PF	PTS	PLUS_MINUS	VIDEO_AVAILABLE
WAS	Washington Bullets	29001100	1991-04-21	WAS vs. MIN	L	240	39	...	39	21	3	1	17	16	87	-2	0
MIN	Minnesota Timberwolves	29001100	1991-04-21	MIN @ WAS	W	240	37	...	37	22	8	3	9	15	89	2	0
SAC	Sacramento Kings	29001106	1991-04-21	SAC vs. LAC	W	240	39	...	41	27	11	9	22	20	105	4	0
LAC	Los Angeles Clippers	29001106	1991-04-21	LAC @ SAC	L	240	39	...	46	26	10	8	18	22	101	-4	0
CLE	Cleveland Cavaliers	29001097	1991-04-21	CLE vs. PHL	W	240	47	...	48	36	9	7	13	20	123	13	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
BOS	Boston Celtics	21900008	2019-10-23	BOS @ PHI	L	240	33	...	41	18	4	2	11	29	93	-14	1
NOP	New Orleans Pelicans	21900001	2019-10-22	NOP @ TOR	L	265	43	...	53	30	4	9	19	34	122	-8	1
TOR	Toronto Raptors	21900001	2019-10-22	TOR vs. NOP	W	265	42	...	57	23	7	3	17	24	130	8	1
LAC	LA Clippers	21900002	2019-10-22	LAC vs. LAL	W	240	42	...	45	24	8	5	14	25	112	10	1
LAL	Los Angeles Lakers	21900002	2019-10-22	LAL @ LAC	L	240	37	...	41	20	4	7	15	24	102	-10	1

In the NBA official website, we scrape and collect all the NBA team information and data for thirty years. There are a total of 70082 observations of the game recorded. As mentioned above, 29 variables are recorded in this dataset.

Data processing

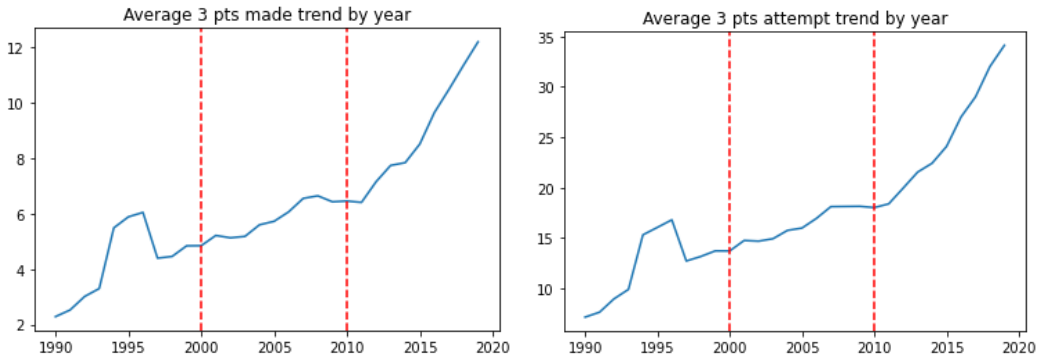
When collecting this data, we intend to filter the data with the usable information. It is essential that the data processing is done correctly so as not to negatively impact the final product or data output. Our main goal is to study different NBA eras and whether the game style has a great impression on the NBA arena. Therefore, the following variables that we will **remove** since it does not relate to our goal, which are 'SEASON_ID','TEAM_ID','TEAM_NAME','GAME_ID','GAME_DATE','MATCHUP','WL','MIN','PLUS_MINUS', 'VIDEO_AVAILABLE'.

The rest of variables of datasets are the ones might directly or indirectly influence our judegment, which are 'TEAM_ABBREVIATION', 'FGM', 'FGA', 'FG_PCT', 'FG3M', 'FG3A', 'FG3_PCT', 'FTM', 'FTA', 'FT_PCT', 'OREB', 'DREB', 'REB', 'AST', 'STL', 'BLK', 'TOV', 'PF', 'PTS', 'era'.

The new dataset is given below:

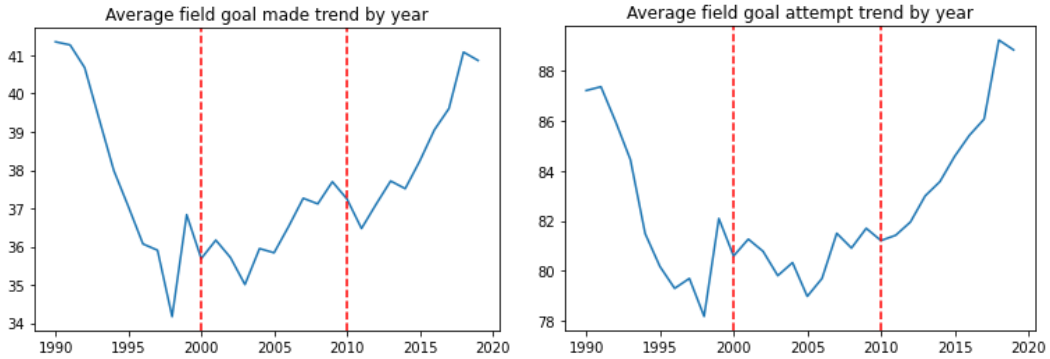
TEAM_ABBREVIATION	FGM	FGA	FG_PCT	FG3M	FG3A	FG3_PCT	FTM	FTA	FT_PCT	OREB	DREB	REB	AST	STL	BLK	TOV	PF	PTS
WAS	39	78	0.500	0	1	0.000	9	14	0.643	7	32	39	21	3	1	17	16	87
MIN	37	90	0.411	3	9	0.333	12	18	0.667	12	25	37	22	8	3	9	15	89
SAC	39	81	0.481	6	9	0.667	21	25	0.840	10	31	41	27	11	9	22	20	105
LAC	39	91	0.429	5	12	0.417	18	24	0.750	15	31	46	26	10	8	18	22	101
CLE	47	89	0.528	3	11	0.273	26	34	0.765	9	39	48	36	9	7	13	20	123

Although we all have a basic thought that the trend of nba teams in the league is to shoot more 3-point shot, we access **the time series point** for the average 3-points trend made by year, average 3-points attempt trend by year, average field goal made trend by year, average offensive rebounds by year, average defensive rebounds trend by year to explore any potential factors or information.

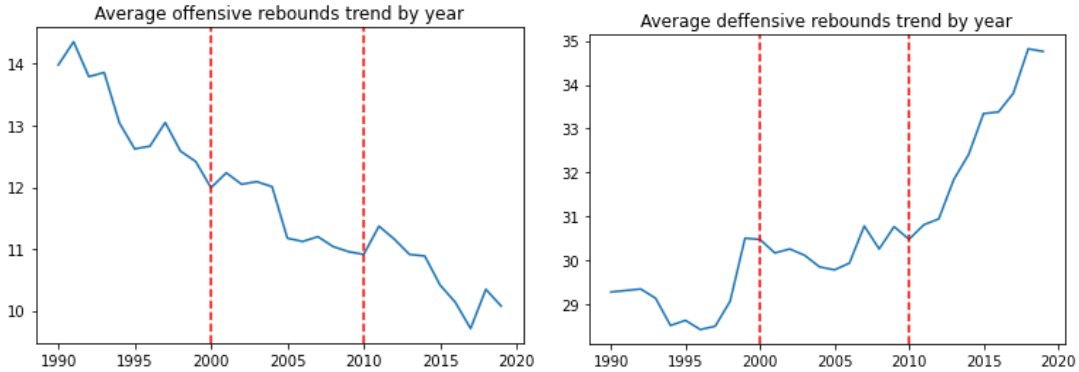


In the first two graphs, apparently, **as the time goes, every NBA team's 3-point made and 3-point attempts have been going up**. During the 1990s, the overall data was lower. There

is a small increase but it is not significant. During the 2000s, the shape of the lines began to have relatively steady movement change. Moreover, since the 2010s, the numbers keep soaring. In a nutshell, the three-point made or the number of three-point attempts per game is growing, which indicates that the style of play in the NBA began to change as the years went by.



From the plots of FGM and FGA, we can see that the trend of average field goal made and average field goal attempt during the 1990s is overall negative. This may tell us that the team's style at that time is more defensive-oriented. During the 2000s, we can observe that the average field goal made and average field goal attempt are facing a fluctuation, which indicates a more stable play style of defensive-oriented. However, after the 2010s, the team's trend of average field goal made and average field goal attempt increased suddenly. This also gives us a hint that **the NBA teams' style of play changed from defensive-oriented to offensive-oriented.**

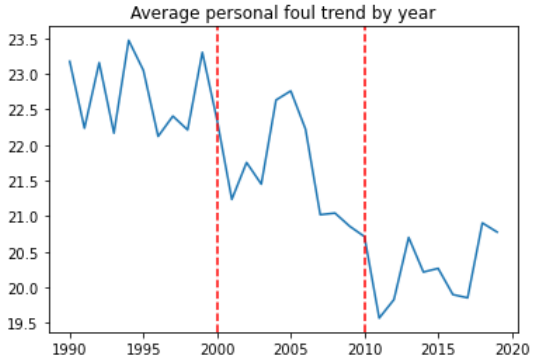


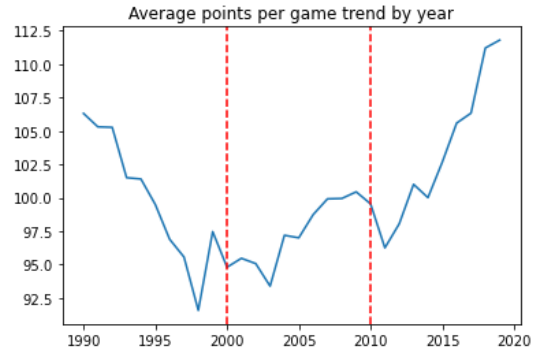
From the two plots, we surprisingly found that the average offensive rebounds trend by year is negative, while the average defensive rebounds trend by year is negative. We speculate that the reason teams are no longer looking for offensive rebounds is that first Big players nowadays like to attempt more outside shots. Unlike in the past, big men are more likely to attack from the key area. They rarely shoot from beyond the arc, but many big guys in today's basketball are opting for more three-point attempts. Because of their placement, defensive players are better positioned to grab rebounds, which has a big

impact on offensive rebounding. If the offensive team's big players attack from the perimeter, the odds of grabbing offensive rebounds are significantly fewer when the interior is vacant.

Second, in situations where a team can't get an offensive rebound, which brings the team to the speed of defensive-to-offensive transition. Higher speed of play also means quicker shots and more converted fast breaks, and long rebounds, which occur more often on long shots, are the best assistants to restoring fast breaks. Defensive teams can start running the fast break at the opponent's free throw line or even the 3-point line after receiving a long rebound, which increases the threat several notches compared to running from the baseline. In addition, the long rebound is unpredictable, and it is not easy for the offense to rush or rebound, so it is not a good choice to invest too many resources to fight for the unpredictable long rebound with the disadvantage of offensive positioning. Therefore, many teams will give up the opportunity to fight for rebounds after the offense, especially after the three-pointers, and choose to go straight back to defense to avoid being scored by the opponent's direct fast break. This is also one of the reasons for the decline of offensive rebounding.

Based on the above two points, the decline in the number of offensive rebounds in the league and the increase in three-point shots might be related, and the increase in three-point shots has become the trend of the league. **Three-pointers have all literally changed the ball game, so teams are focusing more on defensive rebounding than offensive rebounding.**



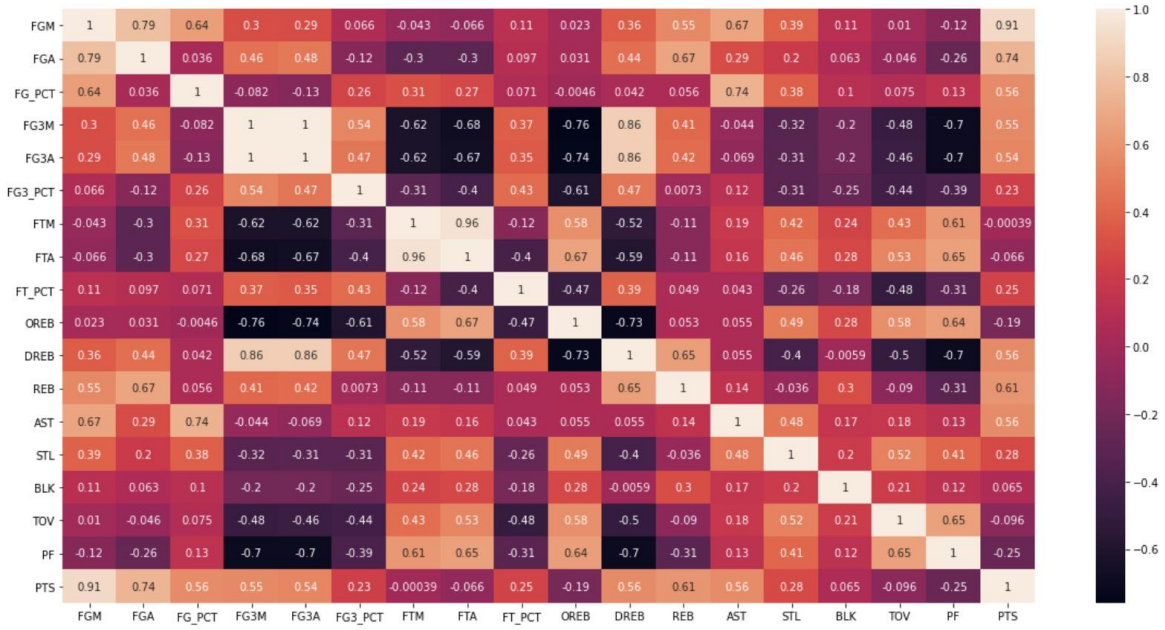


Players' fouls per game have been declining over time. But the area of the average points per game can be divided into three parts. In the first part during the 90s, the average points per game is decreasing, which is caused by so-called defensive-orientedness. Starting from 2000 to 2010, the average points per game is gradually increasing. But since 2010, the average points per game has increased significantly. So, **we believe that the different era may have its representative play style.**

REVIATION	TEAM_NAME	GAME_ID	GAME_DATE	MATCHUP	WL	MIN	FGM	...	REB	AST	STL	BLK	TOV	PF	PTS	PLUS_MINUS	VIDEO_AVAILABLE	era
WAS	Washington Bullets	29001100	1991-04-21	WAS vs. MIN	L	240	39	...	39	21	3	1	17	16	87	-2	0	90s
MIN	Minnesota Timberwolves	29001100	1991-04-21	MIN @ WAS	W	240	37	...	37	22	8	3	9	15	89	2	0	90s
SAC	Sacramento Kings	29001106	1991-04-21	SAC vs. LAC	W	240	39	...	41	27	11	9	22	20	105	4	0	90s
LAC	Los Angeles Clippers	29001106	1991-04-21	LAC @ SAC	L	240	39	...	46	26	10	8	18	22	101	-4	0	90s
CLE	Cleveland Cavaliers	29001097	1991-04-21	CLE vs. PHL	W	240	47	...	48	36	9	7	13	20	123	13	0	90s
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
BOS	Boston Celtics	21900008	2019-10-23	BOS @ PHI	L	240	33	...	41	18	4	2	11	29	93	-14	1	10s
NOP	New Orleans Pelicans	21900001	2019-10-22	NOP @ TOR	L	265	43	...	53	30	4	9	19	34	122	-8	1	10s
TOR	Toronto Raptors	21900001	2019-10-22	TOR vs. NOP	W	265	42	...	57	23	7	3	17	24	130	8	1	10s
LAC	LA Clippers	21900002	2019-10-22	LAC vs. LAL	W	240	42	...	45	24	8	5	14	25	112	10	1	10s
LAL	Los Angeles Lakers	21900002	2019-10-22	LAL @ LAC	L	240	37	...	41	20	4	7	15	24	102	-10	1	10s

From the research, **we decided to separate the data into three periods (1990-2000, 2000-2010, and 2010-2020), which we dubbed "90s," "00s," and "10s."** We would like to find some relationship between the different eras. And we want to distinguish the eras by analyzing the variables of the dataset.

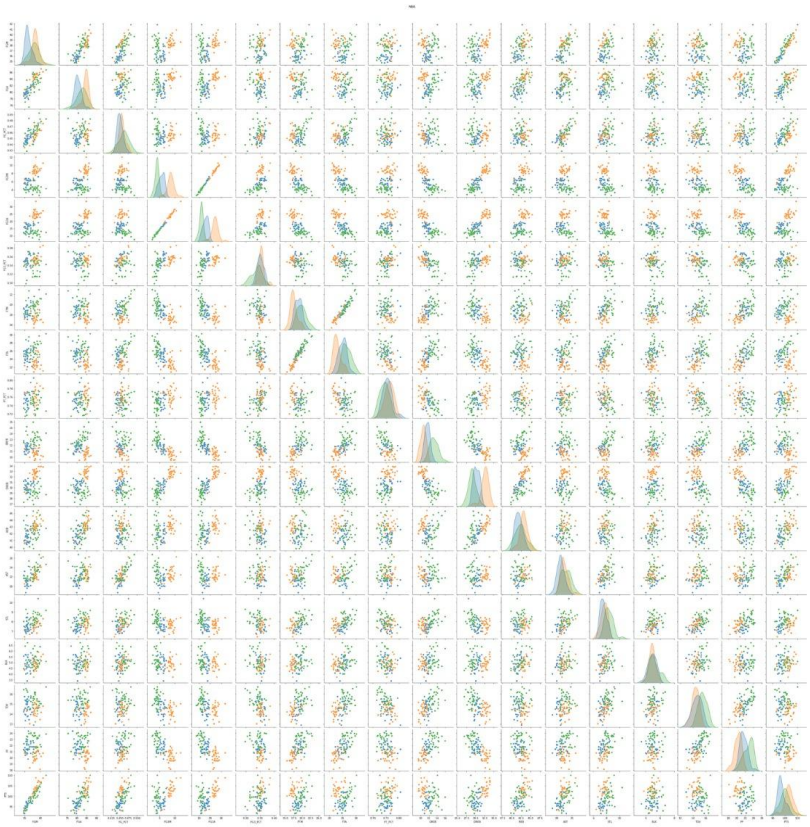
Then we draw the correlation heatmap for these parameters and find the correlation between them:



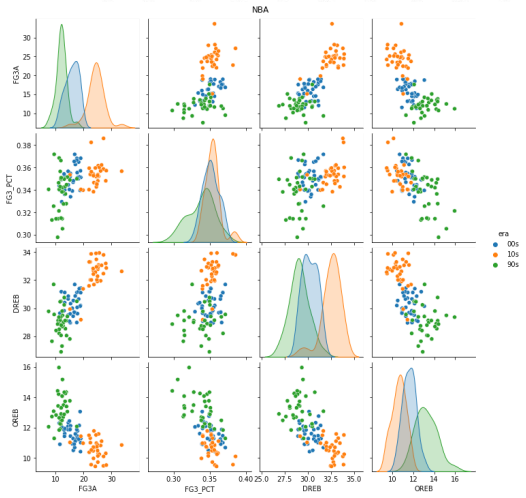
From the relationship plot we can observe the number of **attempts of different types of scoring (field goal, 3 point shot, free throw) are highly correlated with the number of made and percentage of each type of scoring.**

We can observe that the three-point shots and three-point attempt have a correlation of 1, which means that they carry exactly the same information. Meanwhile, free throws made and free throw attempts have a correlation of 0.96, which indicates that they contain highly similar information. So **we are going to use three-point shots and the free throw made and percentage so that we avoid collinearity issues.** Also, the correlation between DREB and FG3M and FG3A is both 0.86, which is higher than other measurements. This means that when a team increases their defensive rebounding, the team will get more offensive possessions. From there, their three-point attempts and three-point shooting percentage will improve greatly.

We will **group the data by era and team and find the mean stat of each team by era, and see if we can find any interesting difference between each era**. Because there are so many measurements in our data, we firstly draw the pairplot between different measurements to find which parameters can be observed clearly.



Blue part: “90s”
Green part: “00s”
Orange part: “10s”
From the plot above, we can see that the histograms of FG3A, FG3_PCT, DREB, and OREB are clear to distinguish these three periods. Then we drop other columns and zoom in to have a closer look at the significant columns.



For the first histogram, basketball players attempted the fewest 3-pointers in the 90s period, followed by the 00s and the 10s attempted most. For the second histogram, when calculating the 3-point percentage, basketball players in the 10s

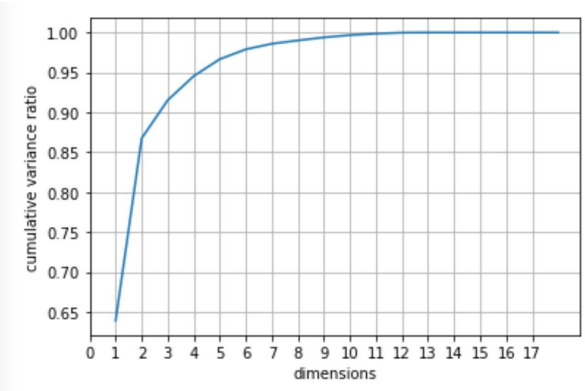
made the highest probability of making a 3-pointer, and the lowest in the 90s.
Combining the third and final histogram, we found that as time goes by, NBA teams tend to have more and more defensive rebounds and have less and less offensive rebounds.

> Thus, we could trust that **NBA teams' style of play changed from defensive-oriented to offensive-oriented and they would like to make more 3-point shots in the competition since the 90s.**

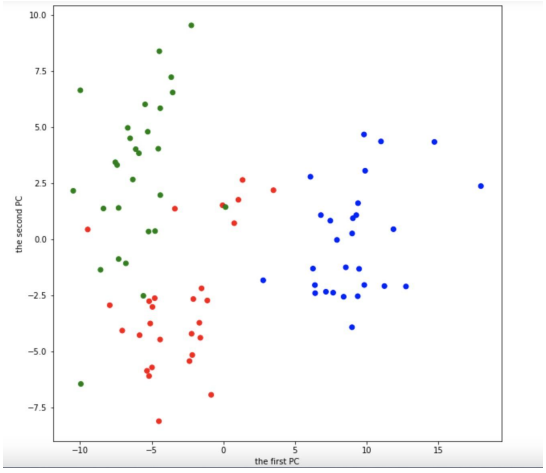
Visualization

> Because there are too many features and it is hard for us to look at everything at once. We will use the **PCA method** to reduce dimension and get a clear look. Then we can check if there are potential clusters. First we need to find out how many components are needed to explain most of the information in our data.

In the plot, we can see the curve of the cumulative variance ratio. When we have one component, we can explain about 65% of the variance. When we have two components we can explain about 87% of the variance, and when we have three components we can explain over 90% of the variance. **Therefore, we only need 2 or 3 components to explain most of the information in our data.**

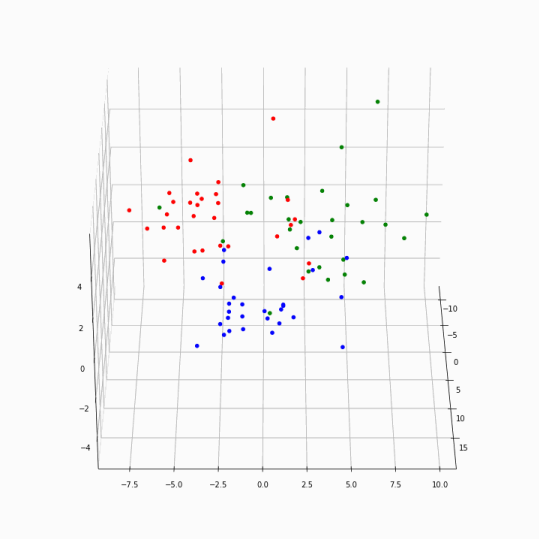


In the plot below, we reduce the dimension into two components, which contain **around 88% of the data.**



Green part: 90s
Red part: 00s
Blue part: 10s

We can find that the blue part (10s) is super distinguishable except there is only one point in the graph so that we conclude that the **style of NBA competition is very different from other eras considering every aspect**. Thus we consider that over 88% is a really good result.



In the 3d plot, we reduce the dimension into three components, which contain around **92% of the data**.

We created a 3D GIF and, in three dimensions, **almost all of the 10s data is separated from data of the other two eras**, which also proved the conclusion above that the style of NBA competition is very different from other eras considering every aspect.

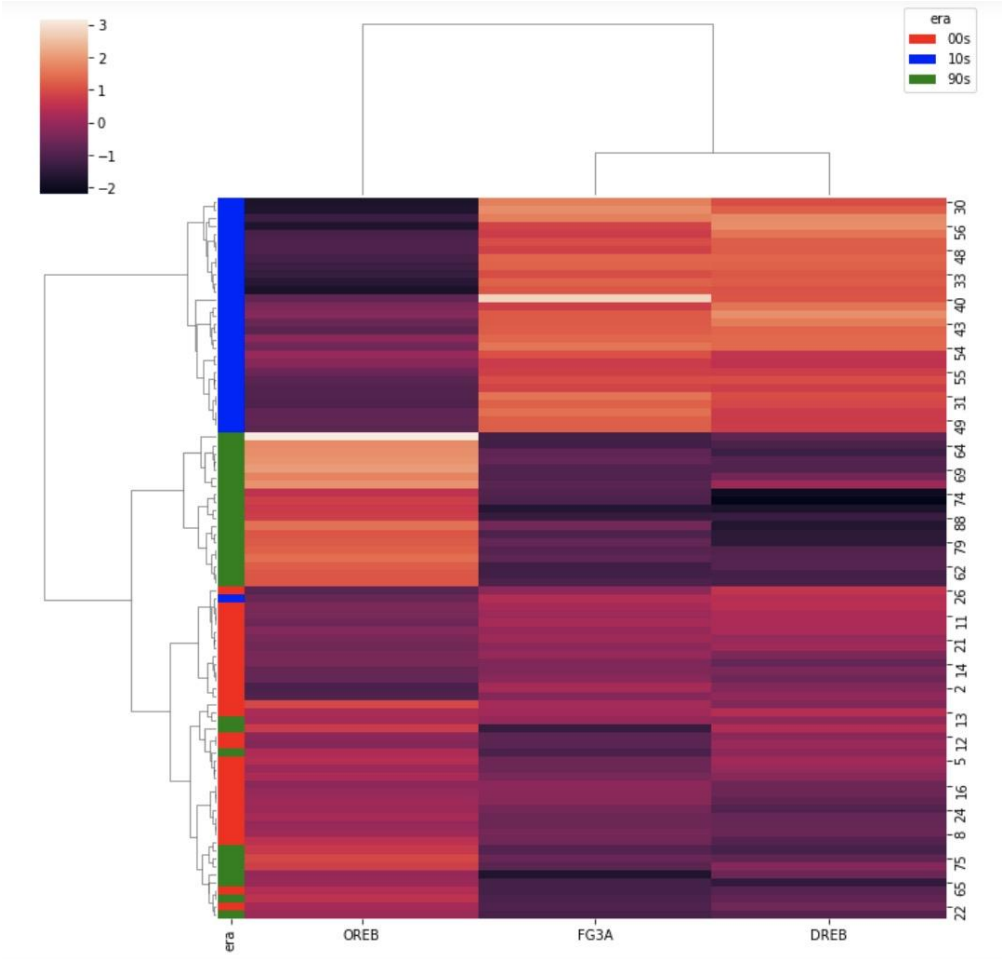
Next, we wanted to find out what caused the differences in the three eras. Therefore, we **calculated the feature importance** and ranked them from most important to least important below. We will pick the top three important features to build our HC tree. Since FG3A and FG3M are highly correlated we will only need one of them. So we will choose **FG3A, DREB and OREB**.

	4	3	10	9	1	0	17	16	7	12	6	13	15	11	5	2	14	8
feature	FG3A	FG3M	DREB	OREB	FGA	FGM	PTS	PF	FTA	AST	FTM	STL	TOV	REB	FG3_PCT	FG_PCT	BLK	FT_PCT
importanc	0.148613	0.141346	0.115942	0.079477	0.077522	0.073607	0.05055	0.049808	0.046371	0.035115	0.033648	0.03136	0.024949	0.024324	0.023323	0.022964	0.012065	0.009017

Then, we used these features to plot the HC tree and Heatmap.

- HC Tree with method ward:

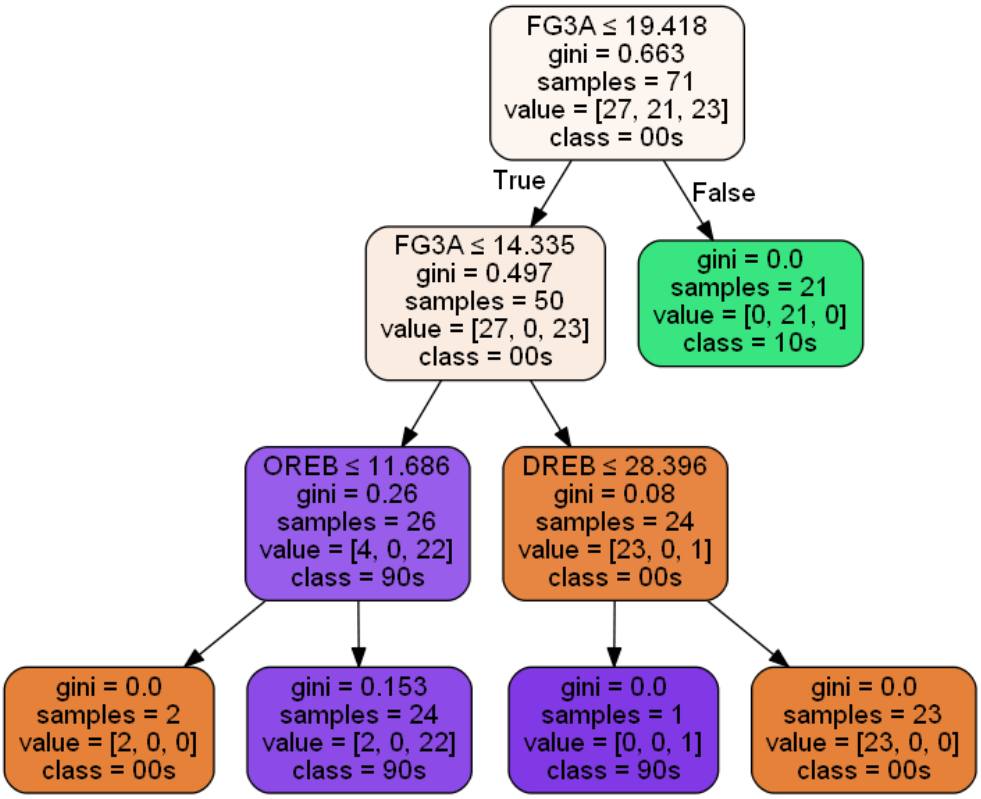
Using function: $D_{kL} = B_{KL} = \frac{||\overline{x_k} - \overline{x_L}||}{\frac{1}{N_K} + \frac{1}{N_L}}$



In this plot, we can see it has three clusters, blue part(10s) on the top ,green part (90s) behind that, and red part (00s) under green clusters.
On the first column of the heatmap we can see on the top, the color is relatively dark, which indicates 10s tend to have less offensive rebounds. And the lighter color chuck following it is correlated to 90s, that indicates 90s tend to have more offensive rebounds. And 00s offensive rebounds tend to be less than 90s and more than 10s.
On the second column, we can see on the top, compared to other eras, 10s have way more three-point attempts. Also, 90s have less three-point attempts in general while 00s are in the middle.
On the last column, we can see on the top, compared to other eras, 10s have more defensive rebounds. And 90s have less defensive rebounds while 00s are in the middle.

To more specifically distinguish which dataset belongs to which eras, we made the **decision tree by features FG3A, OREB and DREB.** Because in the correlation plot, we found that the

goal made and attempted have a high correlation, we ignore the goal made only consider the goal attempted.



On the above, we build a decision tree with 2 depths.

On the first layer, if FG3A is larger than 19.418, we will consider the observation is in the 10s era. Then we continue to investigate, on the second layer. If the observation's FG3A is less than 14.335 and OREB is less than 11.686, it is considered 00s. If the observation's FG3A is less than 14.3355 and OREB is more than 11.686, it is considered 90s. If the observation's FG3A is more than 14.335 and DREB is less than 28.396, it is considered 90s. If the observation's FG3A is more than 11.686 and DREB is more than 28.396, it is considered 00s.

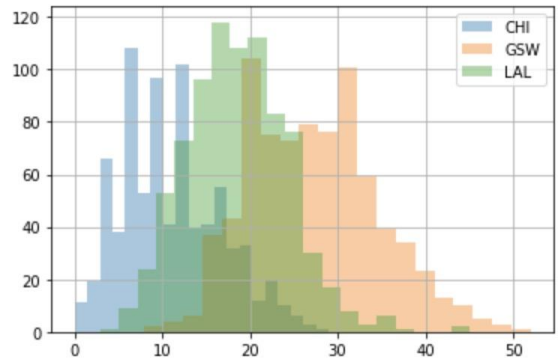
By using cross-validation with 20% testing sample size, we achieve **94.444% accuracy**. Thus, we are confident in our model.

```
clf = DecisionTreeClassifier(max_depth=3)

clf = clf.fit(X_train,y_train)

y_pred = clf.predict(X_test)

print("Accuracy:",metrics.accuracy_score(y_test, y_pred))
```

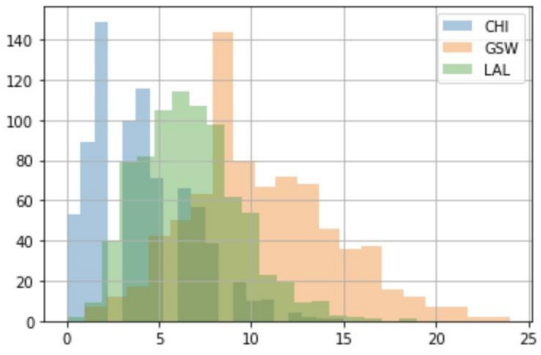
```
Accuracy: 0.9444444444444444
```

Next, we chose the **best NBA teams in different eras (The team with the most championships in the decade)—they are CHI(Chicago Bulls) in the 90s, LAL(Los Angeles Lakers) in 00s, and GSW(Golden State Warriors) in the 10s** and drew the histograms of FG3A and FG3M of them separately.

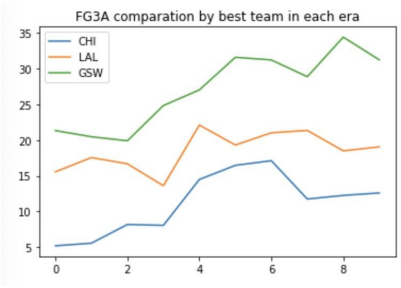
```
team
CHI     AxesSubplot(0.125,0.125;0.775x0.755)
GSW     AxesSubplot(0.125,0.125;0.775x0.755)
LAL     AxesSubplot(0.125,0.125;0.775x0.755)
Name: FG3A, dtype: object
```

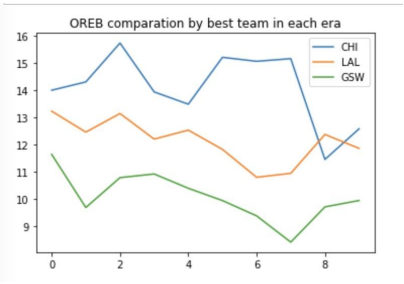


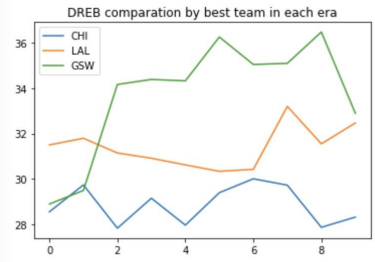

```
team
CHI     AxesSubplot(0.125,0.125;0.775x0.755)
GSW     AxesSubplot(0.125,0.125;0.775x0.755)
LAL     AxesSubplot(0.125,0.125;0.775x0.755)
Name: FG3M, dtype: object
```



Both histograms apparently shows that the **best team GSW in 10s attempted and made the most 3-point shots than LAL in 00s and CHI in 90s**, which proved our previous conclusions that as time goes by, NBA teams would like to make more 3-point shots to win the competition.







Besides, we also plotted the FG3A, OREB, and DREB lines for the best teams in their respective glory era. The result is that most trends of DREB is similar to that of FG3A, meaning that **the number of three point made and defensive rebounds in GSW (10s) is the most and CHI(90s) is the least.** Meanwhile, combining two plots, we found that the OREB is negatively correlated with the DREB. Thus, **the number of offensive rebounds in GSW(10s) is the least and CHI(90s) is the most.**
These examples of the best team in each era proved all basic results we got in the previous study.

Conclusion

Basketball is a very popular sport in America, and millions of people watch the NBA competition every year. However, some think that the style of NBA competition has changed. Thus, we are interested in exploring whether there are differences in the playing styles of basketball in different eras or not. First of all, we found nearly 30 years' data on an NBA website, and extracted the data we wanted by using the API method. Then we removed the unnecessary columns to build up a new dataset.

We firstly drew the time series separately for the average 3-points trend made by year, average 3-points attempt trend by year, average field goal made trend by year, average offensive rebounds by year, average defensive rebounds trend by year to explore the changes these years.We got the conclusion in the following:

1.The three-point made or the number of three-point attempts per game is growing as the year went by.

2.Decreasing field goal percentage and hitting percentage means that NBA teams' style of play changed from defensive to offensive as time went by.

3. NBA teams focus more on defensive rebounding than offensive rebounding as time goes by.

4.NBA teams prefer offensive rather than defensive competition according to declining foul points and increasing team points.

Due to these histograms, we decided to group the data by teams and eras— 90's(1990-1999), 00's (2000-2009) and 10'(2010-2019) and find the mean of each team.Then we drew the correlation heatmap and pairlot to find the relationship between so many measuring methods.

From heatmap, we found that the attempts of field goal, 3 point shot, free throw is positively correlated with the number of made and percentage. Thus, we want to only keep the attempted number to consider in the following exploration. Also, there is a positive correlation between DREB and FG3M or FG3A.

From pairplot, we think that NBA teams' style of play changed from defensive-oriented to offensive-oriented and they would like to make more 3-point shots in the competition since the 90s.

By exploiting the PCA methods, we calculated how much data could be explained in different dimensions. With 88% data in 2-D, and 92% data in 3-D plot, we confirm that the data of the 10s are significantly different from those of the other two eras.

Then we want to seek which parameters are important to cause the difference. Because we ignore the made feature, we only consider “FG3A” “DREB” and “OREB” by calculating the

feature importance and use it to plot HC tree and Heatmap. We get the more clear conclusion from Heatmap and know how to distinguish the era when we only have the data from the Decision Tree with 94.44% accuracy.

Finally, we chose the team which got the most winners in different eras – CHI in the 90s, LAL in the 00s, and GSW in the 10s, and then plotted the histograms and lines of FG3M, FG3A, DREB and OREB. The result we got confirms our conclusion. Therefore, with the high accuracy and the example proved, we are confident about our whole conclusions that the play style of 10s has changed from defensive-orientedness to offensive-orientedness.