



National Basketball Association  
(NBA) 2001-05 & 2016-20, a semi-decade analysis of point guards and centers.

CC7183 Data Analysis and Visualization

WORD COUNT-2678.  
**SAMRAT RAI**  
21025221

### Abstract:

There has always been a comparison and questioning of which decade was better in sports and in the NBA, usually players from the older decades tend to argue that their decade or era was better because of physicality in the league, harsher rules, hands on defence and this argument seems to be had with every decade of players from different era arguing that their decade was better. In this paper I am going to analyse which decade is tougher/better. I am going to be conducting a semi decade analysis of the 2001-05 NBA regular season statistics and 2016-20 NBA regular season statistical data. I am going to be using two positions which is point guard and Center. The statistics I will be using will be using to analyse are points per game, rebounds per game, blocks per game, assists per game, steals per game, turnovers per game, field goal made per game, field goal attempted per game, 3point made per game, 3 points attempted per game, games played, and minutes played per game. My hypothesis is that older decade players tend to play a more traditional basketball style which goes by the rulebook and new decade players tend to play more modern basketball style which is unconventional or unorthodox to basketball standards. My aim is to evaluate the versatility, growth, and shortcomings of the new and old semi-decade players.

### Methods:

I have used Microsoft excel to store my raw data which I have extracted from the official Espn.com website. I have cleaned and calculated the statistical data's I will be using for this paper in Microsoft excel. I have then imported the raw calculated data of Microsoft excel to tableau. I have used tableau for data visualization. The methods of visualization I have used are time series analysis, time series clustering and k-means.

### Keywords:

Point guard, center, average, time series, k-means, NBA, statistical data, sports analytics, analytics.

#### 1. Introduction:

Sports is an area which has numerous amounts of statistical data. Statistical data has become more and more imminent along with the advancement of technological generations

throughout different decades. sports analytics was first started in 1958 in the sport of baseball when henry Chadwick created a metric system known as box score (Kashyap, 2021). Statistical data were not as important and analysed in the past as it is today, most sports started using analytics around 2002 (Schroer, *How sports analytics are used today, by teams and fans* 2022). currently in this decade of 2020 statistics and data plays more a role than ever before and is the general bar of evaluation we grade a sports team, an individual athlete, group of athletes etc. NBA is one of the sports which is heavily embedded in the use of statistical data and uses it as a tool for rating a player and evaluating their overall strengths and weaknesses. my aim in this research journal is to evaluate the overall growth of points guards and centers regardless of if my hypothesis remains true or false, which is older semi-decade players tend to have a traditional/rule book plays than the newer semi-decade players. By traditional I mean to play by the rule which the game has created which in this paper is the rules that NBA has created. In general point guards are supposed to play the role of passing and centers are supposed to play the role of defending. I am going to analyse how the game has changed by analysing these two positions of NBA.

## 2. Models used:

There is a total of three models used for this research journal which are timeseries, timeseries clustering and k-means clustering. These models were selected as it is compatible with the NBA data set and would be an efficient method for data visualization.

### 2.1 Time series and time series clustering:

Time series is a collection of well-defined data gathered throughout a series of time (*Australian Bureau of Statistics Web Site* 2022), most time series data have long time frame consisting of months, years or even decades of data. Time series analysis is typically used for predictions and to see trends. This method of visualization is used as this journal consists of two semi-decades of NBA statistical data with 14 different variables to be presented and analysed which works well with a time series analysis. Time series clustering is a form of clustering with time values included which makes it unique from other forms of clustering methods (Verma, *All you need to know about time-series clustering* 2022). Time series

clustering method is used to separate timelines of the older and newer semi-decade data into two clusters where each cluster represents a timeline.

## 2.2 k-means clustering:

the third models used in this research journal is k-means clustering which separates data points into groups where K is the mean or centroid of each cluster (*Gajawada, K-means clustering for beginners 2019*).

## 3. Literature review:

The NBA first started using statistical analysis in 1955 where the mid 50's is known to be as the dark age era of NBA analytics where there were some complete game logs, stats of line-up and scores, a basic box score analysis consisting of assists, points, rebounds, field goals, free throws, and a lack of clarity when it came to the team efficiency estimates (*Analytics in the NBA: Definitive guide 2022*). After a stagnant period in the 1960's, the 70's brought analytics into the frame again when in 1971 the league started incorporating rebound percentages (*Analytics in the NBA: Definitive guide 2022*). on the same decade, box plus/minus (BPM) was introduced in 1974(*Myers, Box plus-minus (BPM) explained 2020*), BPM is a box score metric that estimates a player's contribution to a team when playing (*McGettigan, The box plus/ minus (PBM) metric to estimate a player's contribution to the team 2020*), it helps to evaluate the level of impact a player brings which ultimately separates the good, bad, and great players. In terms of BPM the league average is 0.0 which denotes an average player, + 2 denotes a good player, -2.0 denotes a bench player and +5.0 denotes a superstar level player (*Play-by-play 2022*). Bpm is a good regular season statistics which denotes how players are and it plays a role in determining the potential salary of a player. In 1983 complete game logs for every team were made available, then in 1994 seasonal plus minus was created and in 1997 play by play was created where an announcer or commentators would give the details of live in game action play by play (*Basketball 2022*). In 2005 synergy tracking was installed where live in game action is monitored by real time statistical video indexing camera which analyses players playing patterns and styles

(*Basketball team performance- optical tracking 2022*). And the latest stat tech to be released is optical tracking in 2014, which is the more advanced version of player tracking where sport vu camera is installed in the basketball court to analyse player movement, record their statistics, and helps immensely when scouting for potential recruits as it summarizes their performance (*Basketball player positions 2022*). In this research journal I have used simple box score statistics which analyses the overall impact players have over a span of years, which for this journal is a span of five years. Box score metric system is used in almost every sport and is the main way to determine a players resume which further allows them to get promoted from a bench player to starter and vice versa. Box score plays a vital role in determining the overall accolades a player has achieved throughout their career and seals the chance of becoming one of the greatest in the game, a chance of being remembered in NBA history and a chance to secure a spot in the NBA hall of fame. Box score metric also serves as a solid avenue to later progression in life for players after they retire where they could work as a coach, side-line personnel, sports commentator, sports journalist etc which many successful basketball players tend to do after they retire. I have extracted box score data sets from the official NBA page located in ESPN's website (*2022-23 NBA stat leaders 2022*). These box scores will be used to compare analyse which era of point guards and centres play by their original positions and follow the rulebook. Years of box score analysis is a solid way of findings how the league has changed in terms of core positions like point guard and centre.

#### 4. Methods:

##### 4.1 Data sources:

The main source of data I have used is the official ESPN website (*ESPN: Serving sports fans. anytime. anywhere. 2022*), which has numerous collections of sports ranging from fight sports like boxing and mixed martial arts, Olympic sports like running and swimming, luxury sports like golf, a variety of x-games sports, winter sports, racing sports like MotoGP, college sports like NCAA etc.it also it resourceful in term of news and analytics. I have used the box score statistics of regular NBA season of two semi decades 2001-05 & 2016-20 from ESPN (*2022-23 NBA stat leaders 2022*), a box score is a summary of averages arranged in a tabular form for teams and players to check their impact on the game (*Basketball box score 2022*).For this research, in terms of point guards' analysis the box scores used are of

14 statistics which are rebounds, points, steals, turnovers, games played, minutes, assists, blocks, three points made, three points attempted, free throw made, free throw attempted, field goal made, and field goal attempted. In terms of centre analysis, I have again used a total of 14 box score statistics which are identical to the stats I have used for point guard analysis. I have used the game played statistics which doesn't fall on the typical daily box score, but it falls into the regular seasons box score stats that is used to analyse a player's durability and commitment. The statistics I have used is to analyse a semi-decade performance which in total is 5 regular seasons where one regular season equals to a year. I have used this box score of semi-decade statistics to analyse my hypothesis.

#### 4.2 Software used:

The software used were tableau for the overall data visualization and Microsoft excel for the raw data cleaning and storing the modified raw data.

#### 5. Findings:

In terms of findings, I have created couple of data visualization diagrams to better understand and for further assessment.

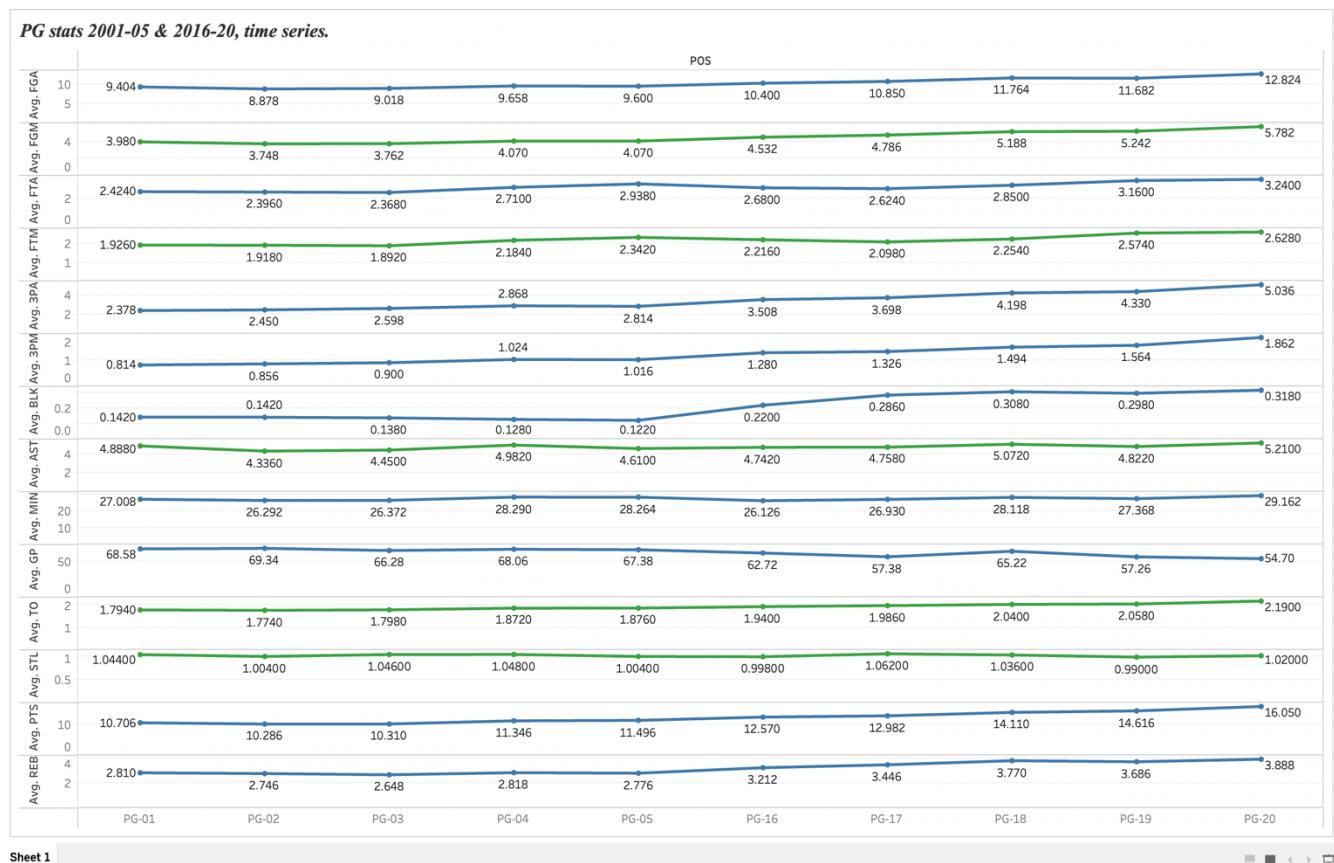


Fig 1- time series analysis of point guard's (PG) stats.

In fig 1, I have used 14 different statistical variables to analyse the two different eras of point guards. PG-01 to P-05 signifies the 2001-05 semi-decade PG and PG-16 to PG-20 signifies the 2016-20 semi-decade PG. The green lines denote the traditional or conventional role a PG plays in basketball standard according to the rule book where PG is supposed to play the role of pass-first, play maker, primary ball handler, play setter, assister, and the driver of the team (*Basketball player positions 2022*). There is a total of 5 green lines in fig 1, which are statistics a PG is accounted for and is responsible to make the team function. the green lines are average assists, steals, turnovers, field goal made, and free throw made. PG's are the floor general of a team where they set plays and are known for passing which accounts as assists, known to be small and agile which accounts for steals, known for taking quality shots which accounts for field goal made which is the overall successful number of points scored, free throw made as they are known be good free throw percentage and lastly turnovers which they means losing the ball and making bad plays which they don't do often as they are skilled ball handlers. To make my hypothesis clearer I have used time series clustering which is Fig 2.



Fig 2, semi-decade PG 2001-05 & 2016-20 time series clustering.

This is a clustering version of time series which I have used in Fig 1, it is simpler to understand, and the average has been calculated for better analysis. There are two clusters where clusters 1 with blue lines indicate PG-01 to 05 & cluster 2 with orange lines indicate PG-16 to 20. The average points of have risen to 10.8 to 14.0, the average three point attempted risen from 2.6 to 4.1, the average field goal attempted risen from 9.1 to 11.5, average free throw attempted from 2.5 To 2.9, where the lower averages pertain to 2001-05 PG'S and the higher to 2016-20 PG'S. it can be clearly seen through the values and time series graph that older semi-decade PGs tend to be more traditional and have stuck to their role whereas the new semi-decade PG'S tend to score more, take more long range and mid-range shots which typically only scoring positions like shooting guard, small forward and power forward tend to do. They have more free throw attempts as they tend to shoot rather than pass the ball which PGs are not supposed to. When it comes to the traditional stats like average assists has risen from 4.6 to 4.9, average field goal made from 3.9 to 5.1, average turnovers risen from 1.8 to 2.04 and average steals fell from 1.0292 to 1.0212. average turnovers have risen as a result of shooting mire and taking fewer quality shots and making fewer quality plays, average steals have decreased slightly so it doesn't make much of a difference. While modern era PG has changed their style of play which has decreased their

some of their traditional statistics but they have become an overall better players because of their increase in stats like points, assists, field goal made, free throws attempted and made with durability being one of the biggest downside with average games played decreased from 67.8 to 59.4 the higher value being the older semi-decade PG and the lower being the newer semi-decade PG. this has occurred as 2016-20 PG's tend to exhaust themselves more by making more offensive plays and highlights which is unconventional and not a typical PG role.

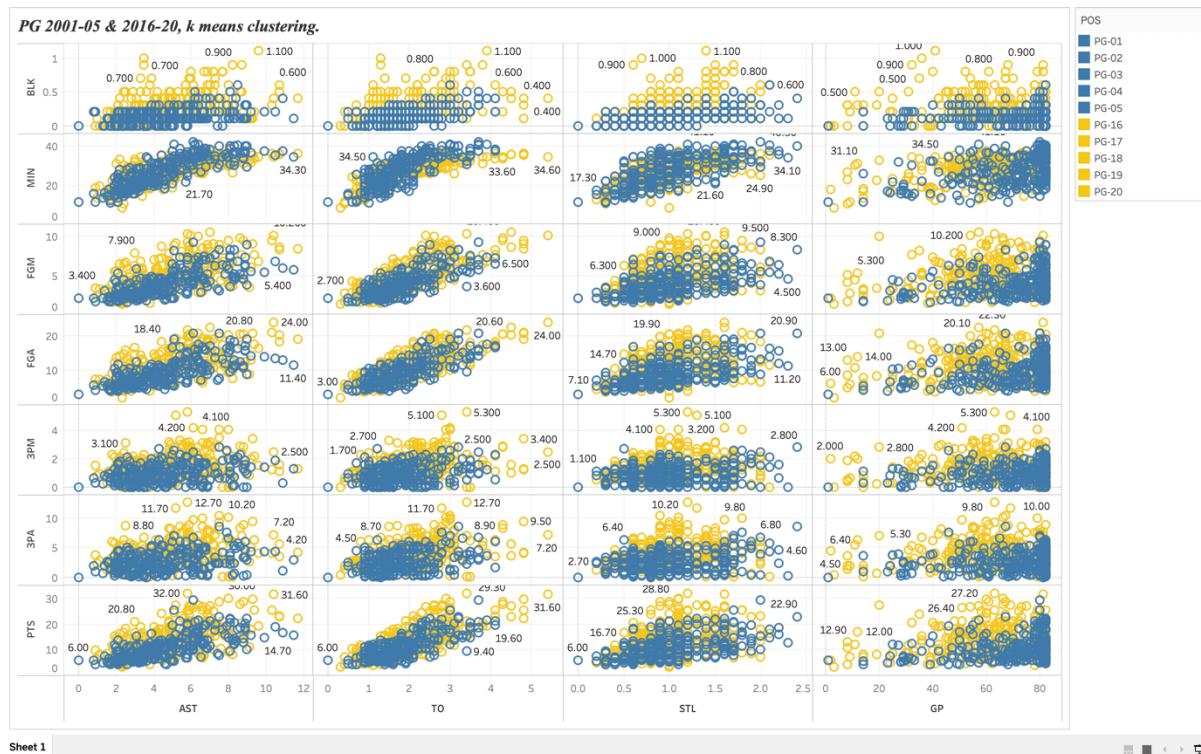


Fig 3, k-means clustering of PG 2001-05 & 2016-20.

The above figure is k-mean clustering where the blue dots represent PG-01-05 & yellow dots represent PG-16 to 20. The x-axis shows the traditional rules book statistical variables while the y-axis shows the unconventional statistical variables which goes out of the rule book. As you can see the in the assists axis the blue dots tend to stick to assists more whereas the yellow dots tend to spread out more towards points, 3points attempts, 3 points made, field goal made and blocks while also making the same amount of assist or more. The only field where the blue dots and yellow spread equally is the assists and mins axis. The same goes true for all the other traditional variables the yellow dots tend to be more spread towards the y-axis while the blue dots tend to be more intact to the x-axis which proves that older semi-

decade PG'S of 2001-05 era tend to play more traditional style of basketball than the newer 2016-20 semi-decade players as the x-axis represents the traditional statistics.



Fig 4, time series analysis of centre.

The green line represents the traditional roles, or the roles typically played by centre positions while the blue line represents unconventional roles usually played by positions other than centre. Huge spikes in blue lines indicate that 2016-20 centres tend to play an unconventional role than the 2001-05 centres.

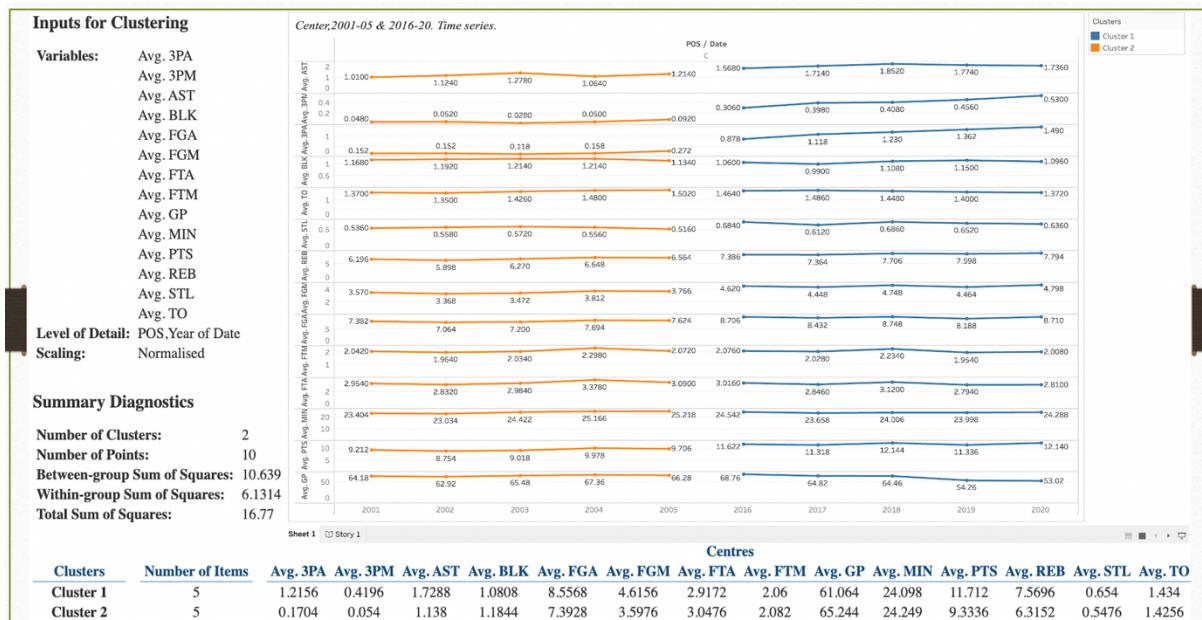


Fig 5, time series clustering of centre.

In the above figure cluster 1 represents 2016-20 centre & cluster 2 represents 2001-05 centre. The traditional roles expected from centres are blocks, rebounds and free throw made (FTM). Average blocks have decreased from 1.18 to 1.08 & average rebounds have risen from 6.3 to 7.5 and FTM has decreased from 2.08 to 2.06. the slight decrease in rebounds and FTM indicates 2001-05 centres being a little more efficient but there is a huge increase in average blocks which indicates the 2016-20 being better overall statistically. Also, the rise in unconventional attributes like average points, 3 points attempted and made, field goals attempted and made, free throw attempted as shown in the graph and mathematical comparison of averages between cluster 1 and cluster 2 highlights clearly that 2001-05 centre play a traditional role while 2016-20 play an unconventional role.

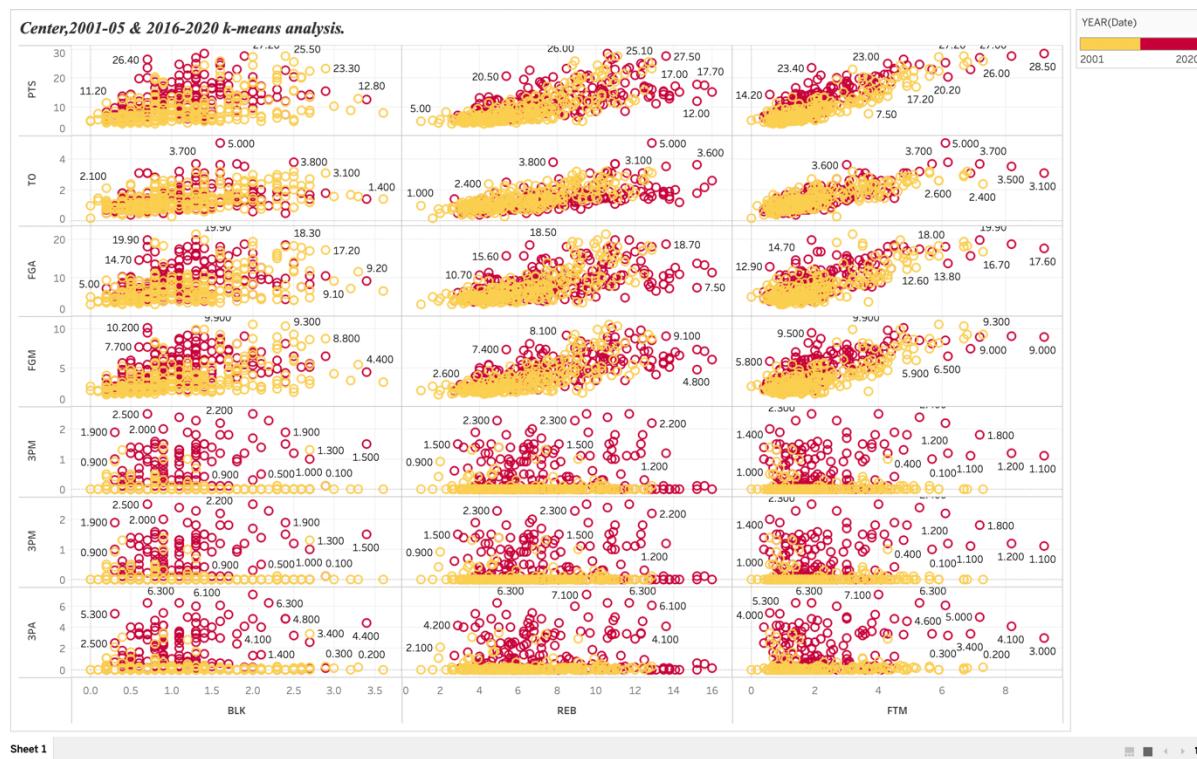


Fig 6, k-means clustering of centres 2001-05 & 2016-20.

In the figure above the yellow dots represent 2001-05 Centres & red dots represent 2016-20 centres. The x-axis is the traditional statistical variable axis & y-axis is the unconventional statistical variable axis. In terms of blocks the yellow dots lean more towards the x-axis whereas in the red dots spread out more towards points, 3 points made & attempted, field goal made & attempted. The only instance both dots are evenly spread out is in average turnovers. K-means clustering points out yellow dots being more intact with the x-axis while red dots are for most of the part spread out to the y and x-axis which pertains to 2001-05 centres being more traditional and 2016-20 centres being more unconventional when it comes to basketball playstyle.

## 6. Critical reflection:

For this research journal I have only used two NBA positions which are point guard and centre. Although there are five positions in total, point guards and centres are the one who have a sturdy role that cannot be changed while other roles like small forward, shooting guard and power forward can switch positions with each other and play according to their teams needs. This is one of the main reasons I have selected Centres and PG'S for this journal. Having said that not having the complete 5 positions makes this journal a incomplete but

would also require more intensive study, research and would require a lot of length would exceed the word limit required for this journal. my hypothesis of 2001-05 semi-decade players(pg/c) being more traditional in-terms of basketball position and 2016-20 players(pg/c) being more unconventional proves to be true as from the data sets, I have acquired, and the visualizations created. While my hypothesis tests to be true and ways of playing have changed which may seem like a downside but It has increased the overall scoring, playing making, rebounding, shooting abilities of both of the modern PG & centres of 2016-20 when comparing to the 2001-05 PG & centres. The only seen downside is that 2016-20 semi-decade players having less durability because of fewer games played.

## References:

H. Kashyap, “A Primer on Sports Analytics: A new dimension of sports,” *Analytics India Magazine*, 05-Aug-2021. [Online]. Available: <https://analyticsindiamag.com/a-primer-on-sports-analytics-a-new-dimension-of-sports/>. [Accessed: 16-Dec-2022].

(Kashyap, 2021)

A. Schroer, “How sports analytics are used today, by teams and fans,” *How Sports Analytics Are Used Today, by Teams and Fans*, 26-Aug-2022. [Online]. Available: <https://builtin.com/big-data/big-data-companies-sports>. [Accessed: 16-Dec-2022].

(Schroer, *How sports analytics are used today, by teams and fans* 2022)

“Analytics in the NBA: Definitive guide,” *NBAstuffer*, 04-Jul-2022. [Online]. Available: <https://www.nbastuffer.com/analytics101/nba-analytics-movement/>. [Accessed: 16-Dec-2022].

(*Analytics in the NBA: Definitive guide* 2022)

“ESPN: Serving sports fans. anytime. anywhere.,” *ESPN.com*. [Online]. Available: <https://www.espn.co.uk/>. [Accessed: 16-Dec-2022].

(*ESPN: Serving sports fans. anytime. anywhere.* 2022)

“2022-23 NBA stat leaders,” *ESPN*. [Online]. Available: <https://www.espn.co.uk/nba/stats>. [Accessed: 16-Dec-2022].

(*2022-23 NBA stat leaders 2022*)

“Basketball box score,” *rookieroad.com*, 2022. [Online]. Available: <https://www.rookieroad.com/basketball/stats/box-score/>. [Accessed: 16-Dec-2022].

(*Basketball box score 2022*)

“Australian Bureau of Statistics Web Site,” *Australian Bureau of Statistics*, 2022. [Online]. Available: <https://www.abs.gov.au/websitedbs/d3310114.nsf/home/time+series+analysis:+the+basics>. [Accessed: 16-Dec-2022].

(*Australian Bureau of Statistics Web Site 2022*)

Y. Verma, “All you need to know about time-series clustering,” *Analytics India Magazine*, 22-Apr-2022. [Online]. Available: <https://analyticsindiamag.com/all-you-need-to-know-about-time-series-clustering/>. [Accessed: 16-Dec-2022].

(Verma, *All you need to know about time-series clustering 2022*)

S. K. Gajawada, “K-means clustering for beginners,” *Medium*, 24-Oct-2019. [Online]. Available: <https://towardsdatascience.com/k-means-clustering-for-beginners-2dc7b2994a4>. [Accessed: 16-Dec-2022].

(Gajawada, *K-means clustering for beginners 2019*)

D. Myers, “About box plus/minus (BPM),” *About Box Plus/Minus (BPM)*, Feb-2020. [Online]. Available: <https://www.basketball-reference.com/about/bpm2.html>. [Accessed: 16-Dec-2022].

(Myers, *About box plus/minus (BPM) 2020*)

D. Myers, “Box plus-minus (BPM) explained,” *NBAstuffer*, 27-Sep-2020. [Online]. Available: <https://www.nbastuffer.com/analytics101/box-plus-minus/>. [Accessed: 16-Dec-2022].

(Myers, *Box plus-minus (BPM) explained 2020*)

S. McGettigan, “The box plus/ minus (PBM) metric to estimate a player's contribution to the team,” *Futsal Focus*, 08-Jul-2020. [Online]. Available: <https://www.futsalfocus.net/box-plus-minus-metric-basketball-futsal/>. [Accessed: 16-Dec-2022].

(McGettigan, *The box plus/ minus (PBM) metric to estimate a player's contribution to the team* 2020)

“Play-by-play,” *Cambridge Dictionary*, 2022. [Online]. Available: <https://dictionary.cambridge.org/dictionary/english/play-by-play>. [Accessed: 16-Dec-2022].

(*Play-by-play* 2022)

“Basketball,” *Synergy Sports*, 14-Jan-2022. [Online]. Available: <https://synergysports.com/sport/basketball/>. [Accessed: 16-Dec-2022].

(*Basketball* 2022)

“Basketball team performance- optical tracking,” *Stats Perform*, 23-Jun-2022. [Online]. Available: <https://www.statsperform.com/team-performance/basketball/>. [Accessed: 16-Dec-2022].

(*Basketball team performance- optical tracking* 2022)

“Basketball player positions,” *rookieroad.com*, 2022. [Online]. Available: <https://www.rookieroad.com/basketball/player-positions/>. [Accessed: 16-Dec-2022].

(*Basketball player positions* 2022)