

Please tick ✓ or click if using MS WORD

FOUNDATION DIPLOMA DEGREE MASTER

Assignment Coversheet

Please complete all details required clearly. For softcopy submissions, please ensure this cover sheet is included at the start of your document or in the file folder.

Assignment & Course Details:

Subject Code: XBDS2034N (e.g. XCAT1234)	Subject Name (e.g. Fundamentals of Computing): Data Science Toolbox		
Course (e.g. Bachelor in Computing) : Bachelor in Computer Science (Hons)			
Lecturer Name: Ms.Norsyela M			
Assessment Due Date: (dd/mm/yy)	27/11/2023	Assessment Title:	NBA Analysis

I/We declare that:

- This assignment is my/our own original work, except where I/we have appropriately cited the original source.
- This assignment or parts of it has not previously been submitted for assessment in this or any other subject.
- I/We allow the assessor of this assignment to test any work submitted by me/us, using text comparison software for plagiarism.
(For more information, Please read the Academic Integrity Guidelines)

Name : Marco Setiawan Student ID: 0134172 Email : 0134172@student.uow.edu.my Mobile No: 0145197082. Signature: MS Date: 27/11/2023	Name : Wong Yen Chik Student ID: 0133621 Email : wongyenchik@gmail.com Mobile No: 011-0159 5371 Signature: wong Date: 27/11/2023	Name : Ling Yan En Student ID: 0132114 Email : 0132114@student.uow.edu.my Mobile No: 010-299 8033 Signature: Ling Date: 27/11/2023
Name : Student ID: Email : Mobile No: Signature: Date:	Name : Student ID: Email : Mobile No: Signature: Date:	Name : Student ID: Email : Mobile No: Signature: Date:

For office use only – Lecturer comments (if applicable)	Marks Breakdown	

Table of Contents

<i>List of Figures</i>	3
1.0 Introduction	5
1.1 Summarization of Literature Review	5
1.2 Hypothesis	6
1.3 Research Question	6
1.4 Methods to accomplish Aims	6
1.5 Preview of Main Results	7
2.0 Methodology	8
2.1 Data Acquisition	8
2.2 Data Preprocessing	8
2.3 Data Visualisation	9
2.2.1 World Map Plot	9
2.2.1.1 NBA players across different countries	10
2.2.2 Bar Chart	10
2.2.2.1 Number of international players in NBA by year (non-US)	11
2.2.2.2 Total 3PA per year	11
2.2.2.3 Total 3PA and FTA by Year	12
2.2.2.4 Total Points for Each Pick in the First and Second Round Draft	13
2.2.2.5 Total Points Scored in Playoffs and Regular Season for both Draft Rounds	14
2.2.3 Line Chart	14
2.2.3.1 Average Three Point Attempt of the USA vs International Players	15
2.2.4 Scatter Plot	15
2.2.4.1 Players PPG vs FGA	16
2.2.4.2 Players Height vs Their Career PPG	17
2.2.4.3 League's Total Free Throw Attempts (FTA) by Year	17
2.2.4.4 League's Field Goal Attempts by Year	18
2.2.4.5 FGA vs FTA for Each NBA Player	19
2.2.4.6 Player's FGA vs TOV	19
2.2.4.7 Players PPG vs the Predicted Top 3 Players' PPG by Year	20
2.2.4.8 Players FGA vs the Predicted Top 3 Players' FGA by Year	20
2.2.4.9 Players 3PA vs the Predicted Top 3 Players' 3PA by Year	21
2.2.4.10 Players ORB vs the Predicted Top 3 Players' ORB by Year	21
2.2.4.11 Player's FTA vs the Predicted Top 3 Players, FTA by Year	22
2.5 Prediction	22
2.5.1 Algorithm	22
2.5.2 Top 3 NBA Players with the Highest PPG in Regular Season in 2021	23
3.0 Results and Discussion	24
3.1 Visualisation	24
3.1.1 World Map Plot	24
3.1.1.1 NBA players across different countries	24
3.1.2 Bar Chart	25

3.1.2.1 Number of international players in NBA by year (non-US)	25
3.1.2.2 Total 3PA per year	26
3.1.2.3 Total 3PA and FTA by Year	27
3.1.2.4 Total Points for Each Pick in the First and Second Round Draft	28
3.1.2.5 Total Points Scored in Playoffs and Regular Season for both Draft Rounds	29
3.1.3 Line Chart	30
3.1.3.1 Average 3 Point Attempt of the USA vs International Players	30
3.1.4 Scatter Plot	31
3.1.4.1 Player's PPG vs FGA	31
3.1.4.2 Player's Height vs Player's Career PPG	32
3.1.4.3 League's Total Free Throw Attempts (FTA) by Year	33
3.1.4.4 League's Field Goal Attempts by Year	34
3.1.4.5 FGA vs FTA for Each NBA Player	35
3.1.4.6 Player's FGA vs TOV	36
3.3 Prediction	36
3.3.1 Top 3 NBA Players with the Highest PPG in Regular Season in 2021	37
3.3.1.1 Predicted Results	37
3.3.1.2 Actual Results	37
3.3.2 Justifications of the Prediction	38
3.3.2.1 Players PPG vs the Predicted Top 3 Players' PPG by Each Year	38
3.3.2.2 Players FGA vs the Predicted Top 3 Players' FGA by Each Year	38
3.3.2.3 Players 3PA vs the Predicted Top 3 Players' 3PA by Each Year	39
3.3.2.4 Players ORB vs the Predicted Top 3 Players' ORB by Each Year	40
3.3.2.5 Players FTA vs the Predicted Top 3 Players' FTA by Each Year	40
3.3.3 Study Comparison Analysis	41
4.0 Conclusion	41
5.0 References	42
Appendix	43
Website Link	43
About Us	43
Main Topic	44

List of Figures

Figure 1.0	Prediction of top 3 players
Figure 2.0	NBA players across different countries
Figure 2.1	Hover to a country
Figure 2.2	Number of international players in NBA by year (non-US)
Figure 2.3	Total 3PA per year
Figure 2.4	Total 3PA and FTA by Year
Figure 2.5	Total Points for Each Pick in the First Round Draft
Figure 2.6	Total Points for Each Pick in the Second Round Draft

Figure 2.7	Total Points Scored in Playoffs and Regular Season for both Draft Rounds
Figure 2.8	Average Three Point Attempt of the USA vs International Players
Figure 2.9	Players PPG vs FGA
Figure 2.10	Players Height vs Their Career PPG
Figure 2.11	League's Total Free Throw Attempts (FTA) by Year
Figure 2.12	League's Field Goal Attempts by Year
Figure 2.13	FGA vs FTA for Each NBA Player
Figure 2.14	Player's FGA vs TOV
Figure 2.15	Players PPG vs the Predicted Top 3 Players' PPG by Year
Figure 2.16	Players FGA vs the Predicted Top 3 Players' FGA by Year
Figure 2.17	Players 3PA vs the Predicted Top 3 Players' 3PA by Year
Figure 2.18	Players ORB vs the Predicted Top 3 Players' ORB by Year
Figure 2.19	Player's FTA vs the Predicted Top 3 Players, FTA by Year
Figure 3.0	NBA players across different countries
Figure 3.1	Number of international players in NBA by year (non-US)
Figure 3.2	Total 3PA per year
Figure 3.3	Total 3PA and FTA by Year
Figure 3.4	Total Points for Each Pick in the First Round Draft
Figure 3.5	Total Points for Each Pick in the Second Round Draft
Figure 3.6	Total Points Scored in Playoffs and Regular Season for both Draft Rounds
Figure 3.7	Average Three Point Attempt of the USA vs International Players
Figure 3.8	Player's PPG vs FGA
Figure 3.9	Player's Height vs Player's Career PPG
Figure 3.10	League's Total Free Throw Attempts (FTA) by Year
Figure 3.11	League's Field Goal Attempts by Year
Figure 3.12	FGA vs FTA for Each NBA Player
Figure 3.13	Player's FGA vs TOV
Figure 3.14	Prediction of Top 3 players
Figure 3.15	Actual Top 3 players
Figure 3.16	Players PPG vs the Predicted Top 3 Players' PPG by Each Year
Figure 3.17	Players FGA vs the Predicted Top 3 Players' FGA by Each Year
Figure 3.18	Players 3PA vs the Predicted Top 3 Players' 3PA by Each Year
Figure 3.19	Players ORB vs the Predicted Top 3 Players' ORB by Each Year
Figure 3.20	Player's FTA vs the Predicted Top 3 Players, FTA by Year

1.0 Introduction

Basketball is a competitive sport where players and organisations strive to be the best and what they've been doing. Fans, however, sometimes wonder what the future holds for the team

and most of them are curious about what these teams will provide now and years to come. In order for these organisations to improve themselves, they must comprehend what are the current trends in the league so that they can keep up with their competitors, and analysing the league using data science will help boost their insights and allow them to patch up what they are lacking. Basketball analytics is created by utilising the tools in data science to conduct assessment on the performance of NBA teams, where the scope includes analysing the players' information, such as points per game, rebounds per game, three points made, and so forth to observe if there are any significant findings from the information given. Therefore, the aim in this research is to understand more about how data science can generate more detailed information so that people can understand better on these players, and how important these data are in terms of improving one's basketball skills, thus readers can develop their skills themselves in accordance with the information given.

1.1 Summarization of Literature Review

One of the most popular games in the world is basketball, and the companies that surround it have greatly benefited from this sport. The use of artificial intelligence (AI) technology in basketball has gathered a lot of attention in recent years. Currently, research related to basketball focuses on artificial intelligence (AI) coaching systems, intelligent training machines and arenas, prediction of competition outcomes, analysis and prediction of shooting, basketball teams and individual performance, and sports injury prevention. The majority of research has demonstrated that artificial intelligence (AI) can raise basketball players' training standards, assist coaches in creating effective game plans, reduce the risk of sports-related injuries, and increase fan enjoyment (Li & Xu, 2021).

Owing to its widespread popularity and significant financial worth, the National Basketball Association (NBA) has also made significant investments in enhancing training techniques, augmenting athletic abilities, and forecasting sports results. This leads to the extensive utilisation of AI algorithms in the NBA, including as Neural Networks (NNs), naive Bayes, decision trees, support vector machines, and random forests, and others. Moreover, NNs have been used to forecast basketball game results by analysing sports data from 890 games. It has been discovered that defensive rebounds and two-point shots are two crucial components for winning games among a number of factors such as two-point shots, three-point shots, and defensive rebounds that affect sports performance. Four types of neural networks (NNs) such as feed-forward, probabilistic, radial basis, and generalised regression were trained using statistics from 620 NBA games, and these algorithms' predictions of winning teams have outperformed those of basketball experts (Wang et al., 2022).

Therefore, as what researchers had learned through these research papers, researchers understand that AI applications in basketball and NBA are still in their early stages. Thus, organisations such as the NBA and others ought to invest in this domain, nurturing developments

in basketball and enhancing the overall quality of the game. Thus, enhancing the excitement of the game further as its global popularity continues to grow (Li & Xu, 2021).

1.2 Hypothesis

Hypothesis is a probable clarification for something that occurs or that you observe and believe to be valid. It can also assist in confirming the relationship among multiple variables that you believe may be connected to each other. Hence, these are the hypothesis of this research:

- Number of International players will (Increase / Decrease) the 3PA in NBA
- Height in NBA players (Do / Don't) matter in NBA
- It is (Easier / Harder) to get fouled these days in NBA
- Draft picks (Do / Don't) matter in players' success in NBA
- NBA players (Improve / Worsen) their PPG by making (More / Less) FGA.

1.3 Research Question

"A question that a research project sets out to answer" is what is known as a research question. In both qualitative and quantitative research, selecting a research question is crucial. The investigation will need to gather and analyse data, and there will be a wide range of approaches used in this process. Good research questions are typically focused and specialized, with the goal of advancing understanding on a significant subject. Thus, these are the research questions of this research:

- Will the number of international players influence the Three Point Attempts (3PA)?
- Does height matter in the NBA?
- Is it easier to get fouled nowadays in the NBA?
- Do draft picks matter in players' success in the NBA?
- Can NBA players improve their Points Per Game (PPG) by making more Field Goal Attempts (FGA)?

1.4 Methods to accomplish Aims

As we know, the aims in this research are to provide useful information for the general public and organisations to improve their basketball skills individually and as a whole. Therefore, the way to accomplish these aims is to conduct experiments by utilising NBA dataset and generate various visualisations to tackle the questions above, such as plotting height with PPG, Total 3PA each year, and many more, along with creating a deployment that can be accessed in the appendix below. Moreover, we also provide one prediction using the ARIMA model to predict who will be the top 3 best players in the future, and also visualise all the information related to these players such as Offensive rebound (ORB), Points Per Game (PPG), Three Point Attempt (3PA), Field Goal Attempt (FGA), and so forth. These data are very crucial for teams that want to improve their players or even individuals that desire to hone their skills to go to the next stage.

1.5 Preview of Main Results

	Player	Predicted_PPG
131	James Harden	34.234559
223	Bradley Beal	33.209660
246	Giannis Antetokounmpo	29.442734

Figure 1.0 - Prediction of top 3 players

As you can see this is the predictions that we discovered during the prediction process and it shows three players with their predicted PPG, which are the highest three from all other players in the same season. Thus, from the predictions generated, we will also provide information on why these players become the best on that season by providing visualisations on their other stats so that these data can be contributed to the general public that wish to be able to play or perform like these top players, or organisations that want to train their own players to become like them. Therefore, more information about the prediction will be discussed further below.

2.0 Methodology

2.1 Data Acquisition

The dataset used in this research is from Kaggle where the data inside consist of various players from the year 2000 to 2020 with distinct information such as height, high school, draft picks, player name, 3PA, Team, Stage, and many more. Furthermore, it does not only have the NBA as the League, other leagues such as the Australian league, the Euro league are also inside the data. Hence, the reason we choose this dataset is because this is the most complete dataset with diverse info that is hard to find in other datasets.

2.2 Data Preprocessing

Data preprocessing refers to any kind of processing done on raw data in order to get it ready for another data processing step. Historically, it has been a crucial first stage in the data mining process. In recent times, methods for preparing data have been modified to train AI and machine learning models and to make conclusions from them. Through data preprocessing, information is changed into a format that can be handled more quickly and efficiently for tasks including machine learning, data mining, and other data science techniques (Lawton, 2022). The pre-processing steps we have taken in our dataset are:

- Added ‘FG’ which represents the field goal percentage by dividing the field goals made by the field goal attempts.
- Added ‘3P%’ which represents the three point percentage by dividing the three points made by the three point attempts.
- Added ‘Age’ which represents the player’s age by deducting the player’s date of birth from the season’s year.
- Added ‘PPG’ which represents the player’s points per game by dividing the player’s total points by the player’s total games played.
- Added ‘RPG’ which represents the player’s rebounds per game by dividing the player’s total rebounds by the player’s total games played.
- Added ‘SPG’ which represents the player’s steals per game by dividing the player’s total steals by the player’s total games played.
- Added ‘BPG’ which represents the player’s blocks per game by dividing the player’s total blocks by the player’s total games played.
- Added ‘APG’ which represents the player’s assists per game by dividing the player’s total assists by the player’s total games played.
- Dropped the rows with null values, which consists of 11 rows from ‘Team’ and 318 rows from ‘birth_year’ and ‘birth_month’.
- Created a separate dataset called ‘nba_data’, which consists of NBA player’s data only.

Additionally, for the prediction, researchers had created a new data frame from the nba_data dataset with several rules in mind. Thus, the process to create the new data frame are as follows:

- Filter out the data that is not in the regular season stage, “new_nba_data = nba_data[nba_data['Stage'] == 'Regular Season'][['Season_Year','Stage','Player', 'PPG']]”
- Remove all the columns except ‘Season_Year’, ‘Player’, and ‘PPG’
- Create a list of players that are still playing in the league by choosing players that still have data within the year 2015 to 2020. Thus, players that only have data from 2015 and below will not be included into the data frame to reduce the risk of retired players added into the data frame.
- Replace the null values for players who have missed a season or more (thus not having a PPG for that particular season) with their mean career PPG.

Hence, these preprocessing conducted in the datasets will be used later in the visualisation and prediction part in the study.

2.3 Data Visualisation

Data visualisation is the process of displaying data using standard images, including plots, charts, infographics, and even animations. These informational visual displays convey data-driven insights and intricate data linkages in an understandable manner. It is noteworthy that data visualisation is not limited to the use of data teams and can be employed for a multitude of reasons. In addition, management uses it to communicate hierarchy and organisational structure, while data scientists and analysts use it to find and interpret patterns and trends (*What is data visualisation?*, n.d). In our visualisations, we will be visualising the data from the year 2000 to the year 2020.

2.2.1 World Map Plot

World map plot is a technique used to plot or visualise world map and countries inside it, we can also provide more data onto each country, such as the population, size, and so forth. Thus, the visualisation below used a world map to visualise the results.

2.2.1.1 NBA players across different countries

NBA Players Nationality

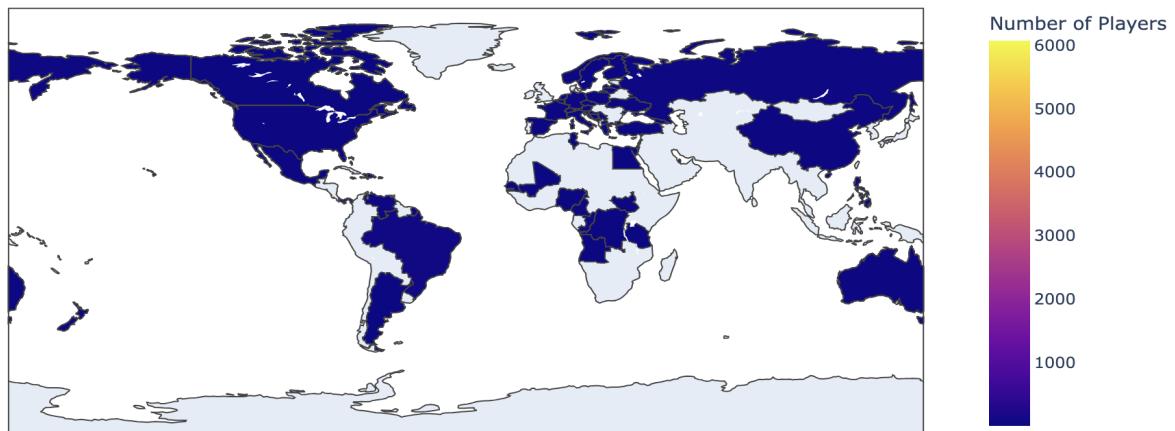


Figure 2.0 – NBA players across different countries

The reason researchers plot this geological map is because as what we see in the modern NBA, there are a lot of international players had already played in NBA and even dominating in the league where most members are Americans, thus we provide this graph for a representation on the sheer difference in comparison between American players to international players. Moreover, if we hover in the map above, we can see the number of players in each country for extra information.



As you can see, if we hover over Russia, it can show the number of players, and this applies to other countries as well.

2.2.2 Bar Chart

A bar chart, sometimes called a graph, uses bars of varying lengths to arrange data into a visual representation. The size of the information these bars represent is reflected in their proportional length. Thus, a long bar chart means there is a huge amount of data in that category, and the shorter ones contain less information on other categories. Thus, the visualizations below used bar charts as the method to visualise the information.

2.2.2.1 Number of international players in NBA by year (non-US)

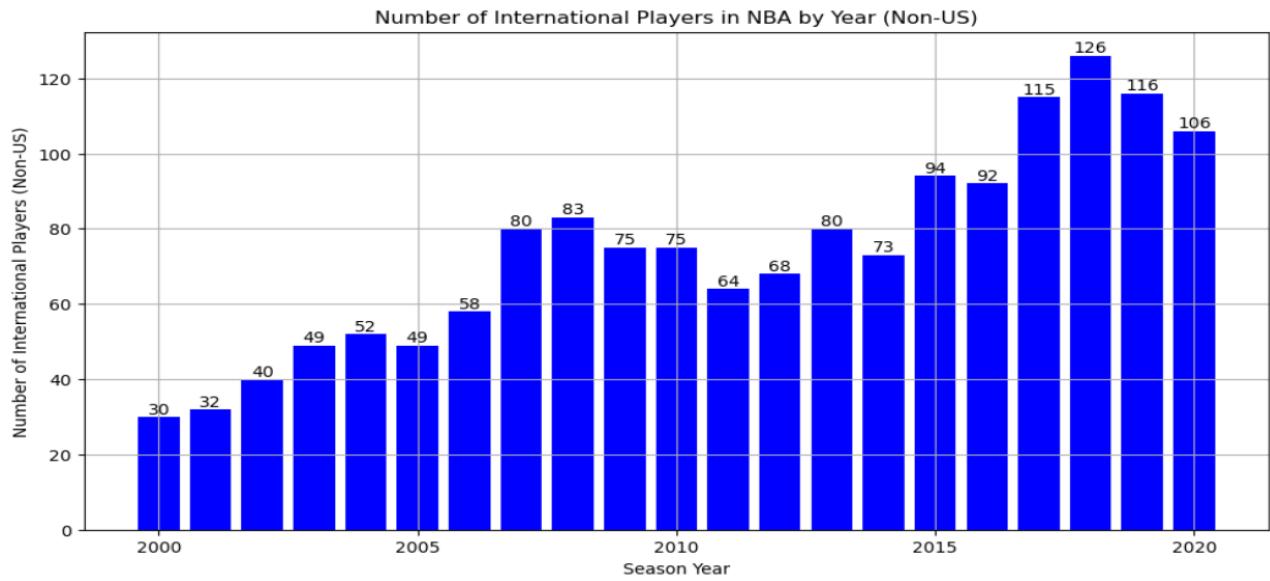


Figure 2.2 – Number of international players in NBA by year (non-US)

The reason why researchers plot the number of international players in the NBA by year is to allow readers to see the pattern of the bar which as we know, is constantly increasing over the few years. We will discuss this graph later in the discussion and results part.

2.2.2.2 Total 3PA per year

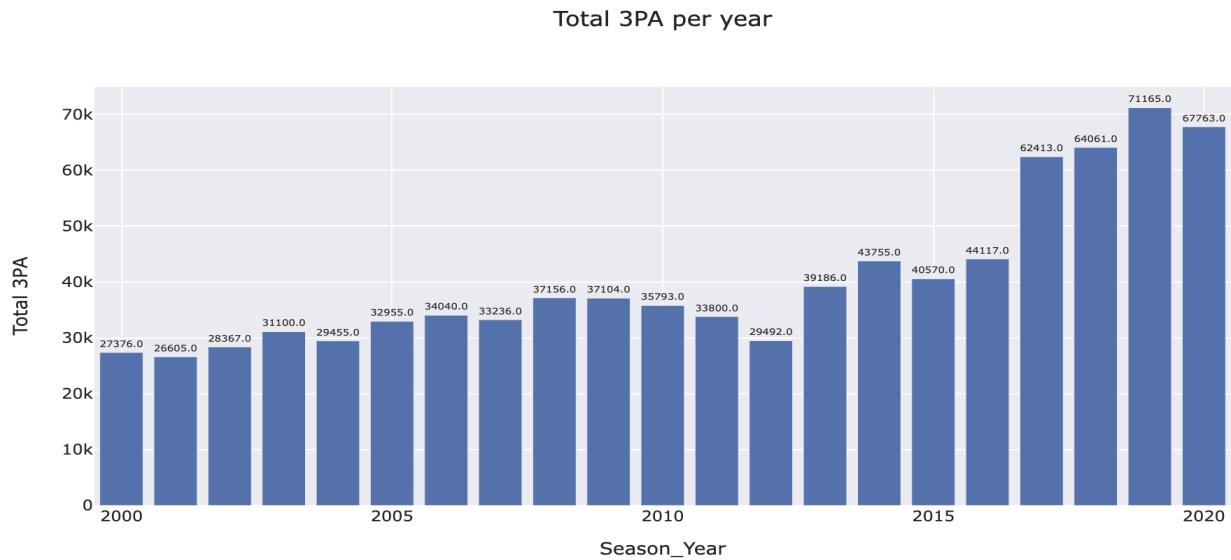


Figure 2.3 – Total 3PA per year

The reason why researchers plot this graph is also similar to the one above, which is also to discover the patterns which in this case, is the total 3PA per year, which is also constantly increasing over the years. This graph will be discussed more later on.

2.2.2.3 Total 3PA and FTA by Year

Total 3PA and FTA by Year

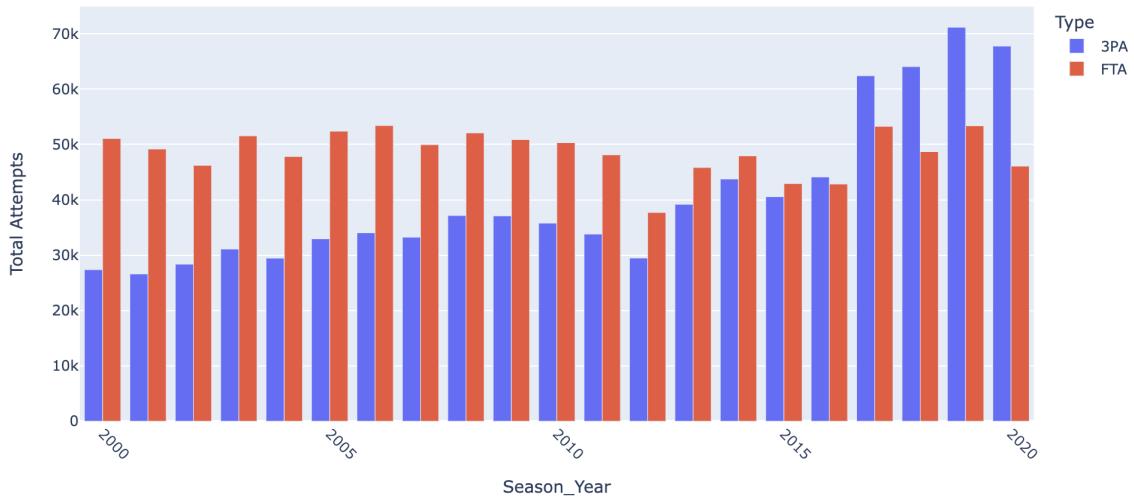


Figure 2.4 – Total 3PA and FTA by year

As we can notice, the reason why researchers create this bar chart is to show a comparison between three point attempts and free throw attempts. Same as what we discovered just now, 3PA is increasing over the years, but FTA is constant. This visualisation will be discussed later on.

2.2.2.4 Total Points for Each Pick in the First and Second Round Draft

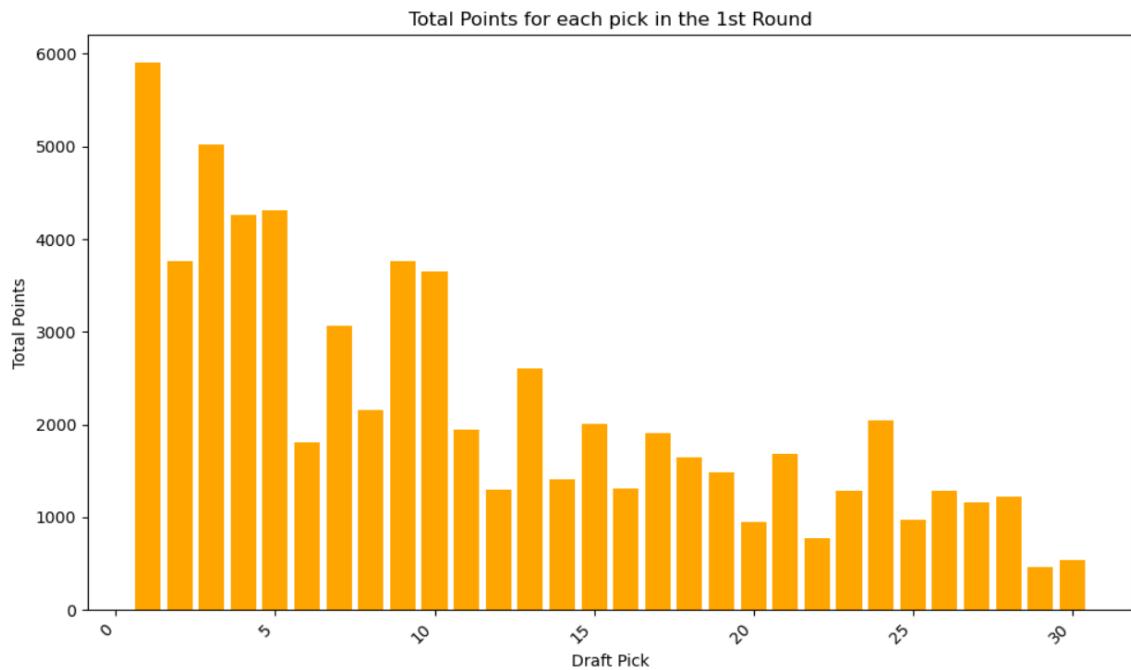


Figure 2.5 – Total points for each pick in the First Round Draft

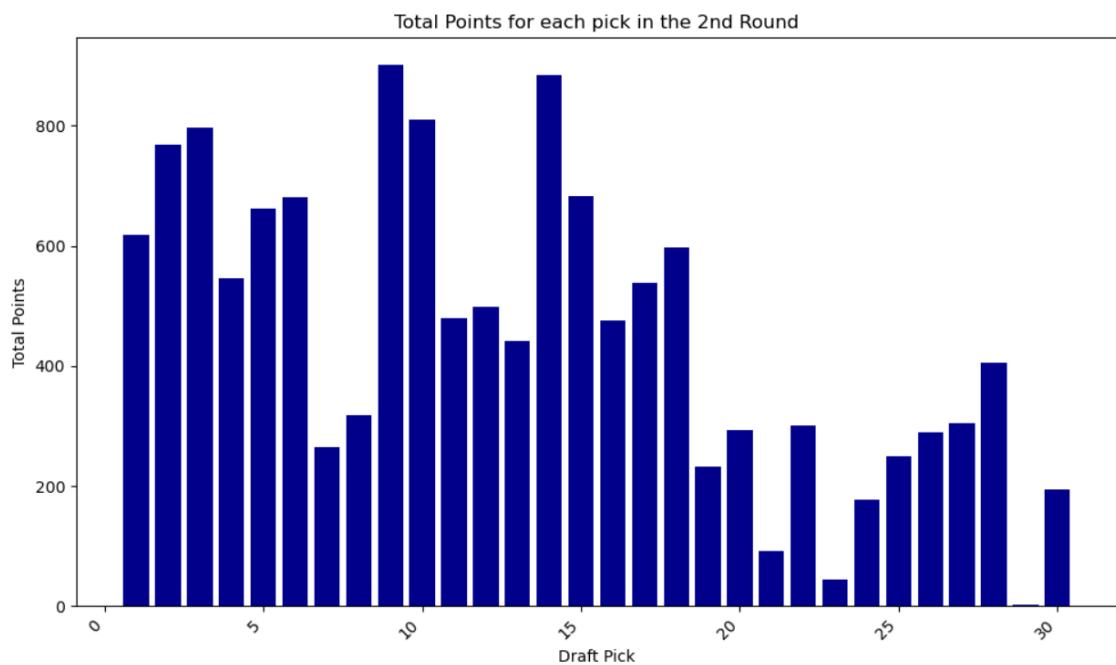
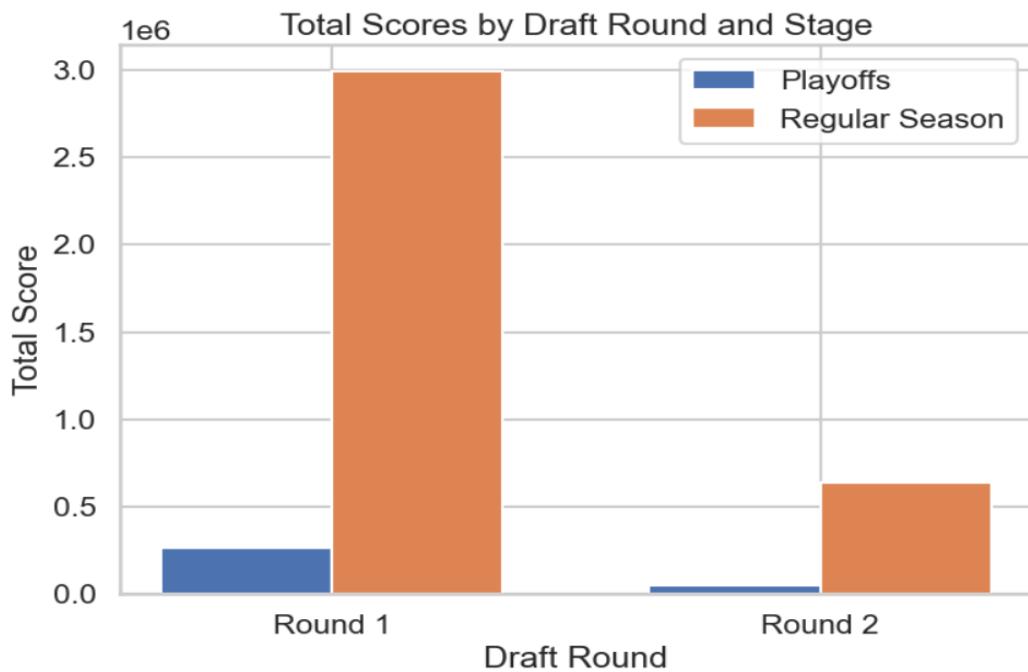


Figure 2.6 – Total point for each pick in Second Round Draft

The reason why researchers want to plot these graphs that provide total points for each draft pick is to get insights on which draft picks are the better one and what are the differences between them. Thus, these graphs will be discussed further in the results and discussion part.

2.2.2.5 Total Points Scored in Playoffs and Regular Season for both Draft Rounds



The reason why researchers plot this chart is to see the differences in terms of performance for each draft pick when playoffs time comes. Thus, this graph will be discussed further later in the research.

2.2.3 Line Chart

A line chart is a graphical representation that shows data points connected by straight lines. It is sometimes referred to as a curve chart or line graph. This kind of chart is especially helpful for displaying relationships, trends, and changes in data over a continuous period of time. Every data point on the graph denotes a value connected to a certain category or moment in time. Viewers are able to easily identify patterns and changes in the data because of the connecting lines that connect these spots. Therefore, visualisations below are using line charts as the method of visualisation.

2.2.3.1 Average Three Point Attempt of the USA vs International Players

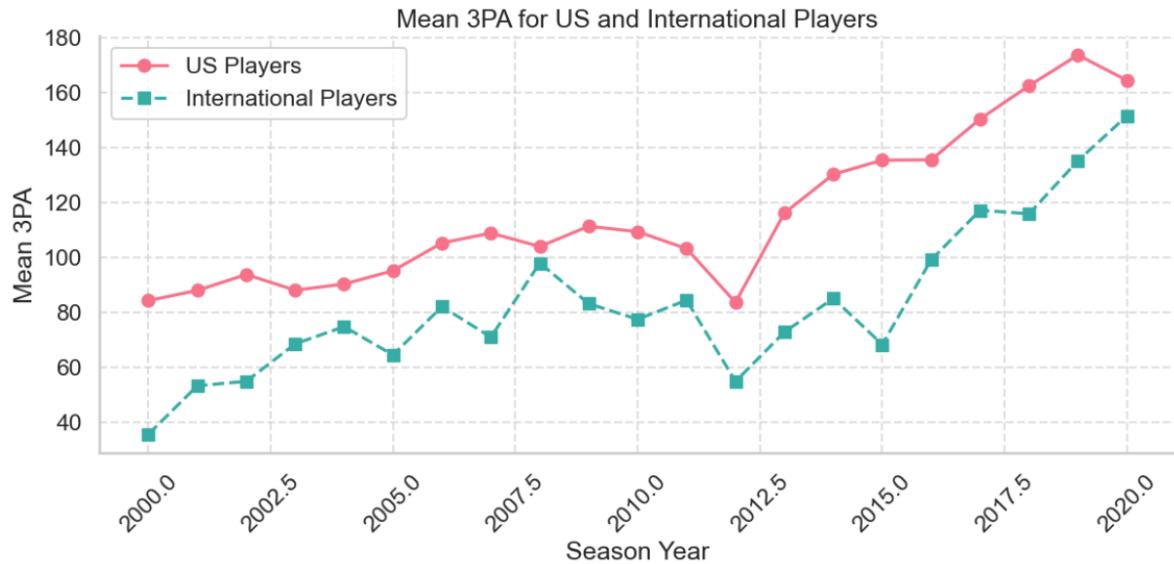


Figure 2.8 – Average Three Point Attempt of the USA vs International Players

The reason why researchers plot this chart is due to discovering the differences in average three point attempts between US players and also international players in the NBA, and further information will be discussed later on.

2.2.4 Scatter Plot

The values for two distinct numerical variables are represented by dots in a scatter plot, also known as a scatter chart or scatter graph. The values for each individual data point are indicated by the position of each dot on the horizontal and vertical axes. Relationships between variables are observed through the use of scatter plots. In addition to reporting the values of individual data points, the dots in a scatter plot can reveal patterns when the data are viewed as a whole. Hence, below are the visualisations that utilise scatter plots.

2.2.4.1 Players PPG vs FGA

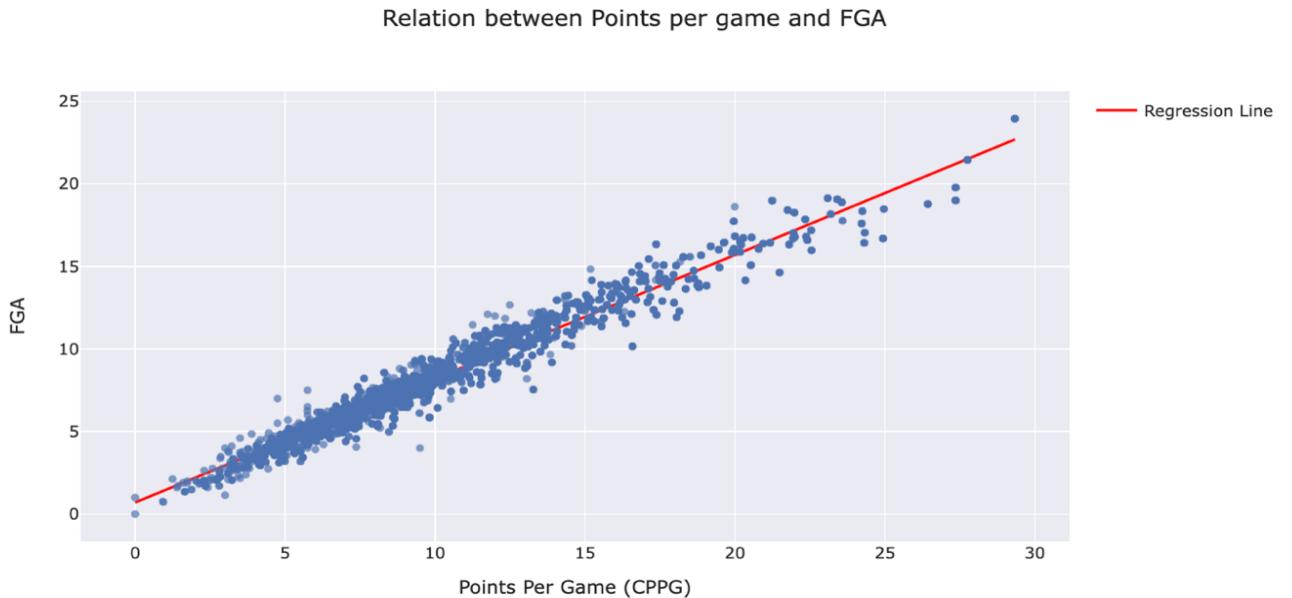


Figure 2.9 – Players PPG vs FGA

The reason why researchers plot this graph is to allow readers to understand the connection between Points Per Game and the Field Goal Attempt, since a lot of people wonder if higher PPG means higher FGA and vice versa. The regression line is also plotted in the middle to see the relationship between two variables. Thus, to comprehend this graph, researchers will explore it more later on.

2.2.4.2 Players Height vs Their Career PPG

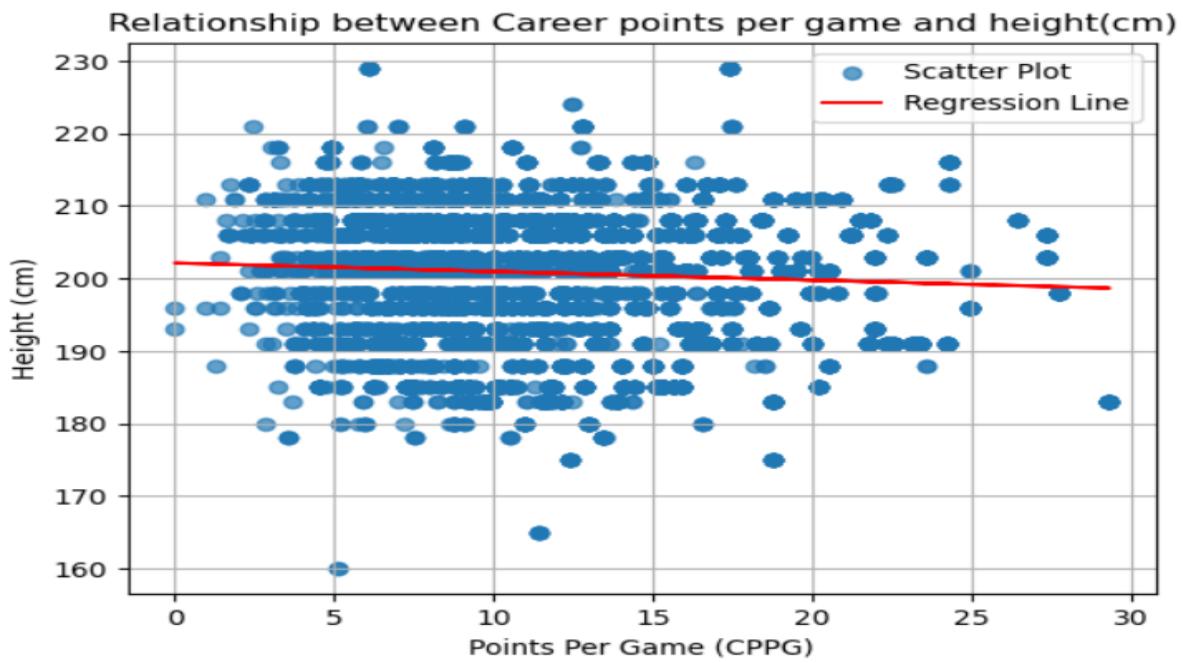


Figure 2.10 – Players Height vs Their Career PPG

As what researchers understand, the main reason for the creation of this visualisation is due to the need to comprehend in terms of how the height of a player can affect their performance in the game of basketball, and by showing their career points we can deduce if taller or shorter players are the better ones. Hence, the result of this graph will be further addressed afterward.

2.2.4.3 League's Total Free Throw Attempts (FTA) by Year

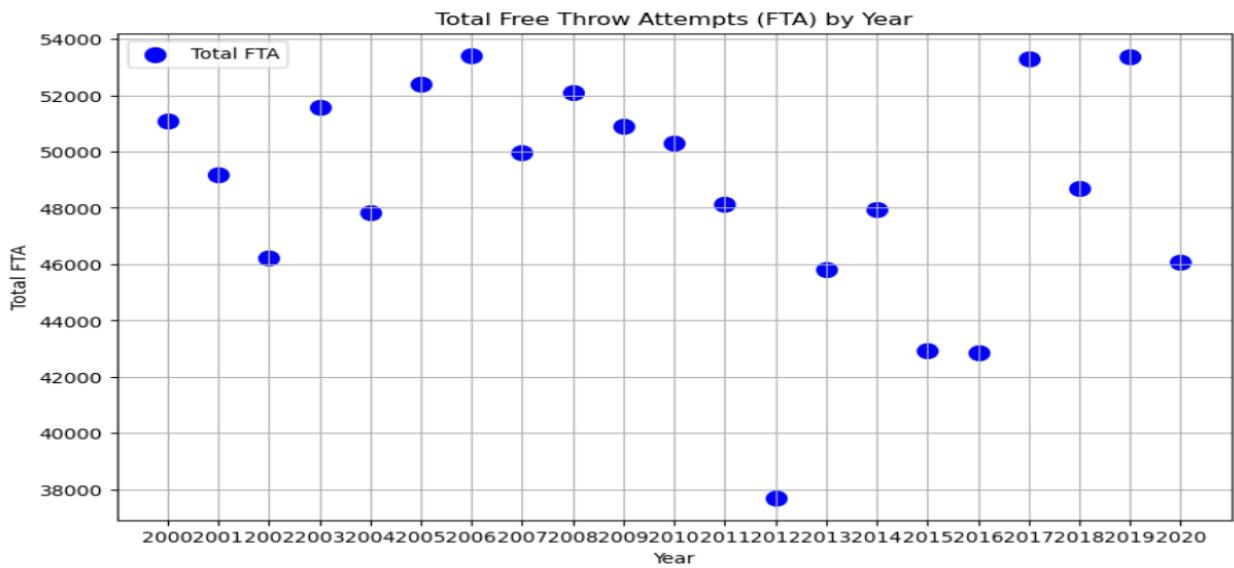


Figure 2.11 – League's Total Free Throw Attempt (FTA) by Year

The main cause of why researchers want to plot this graph is to see if there is any significance difference or changes in free throw attempts between all of those years, and this graph will also be utilised as a supporting graph to assist in explaining other graphs later on in the discussion part.

2.2.4.4 League's Field Goal Attempts by Year

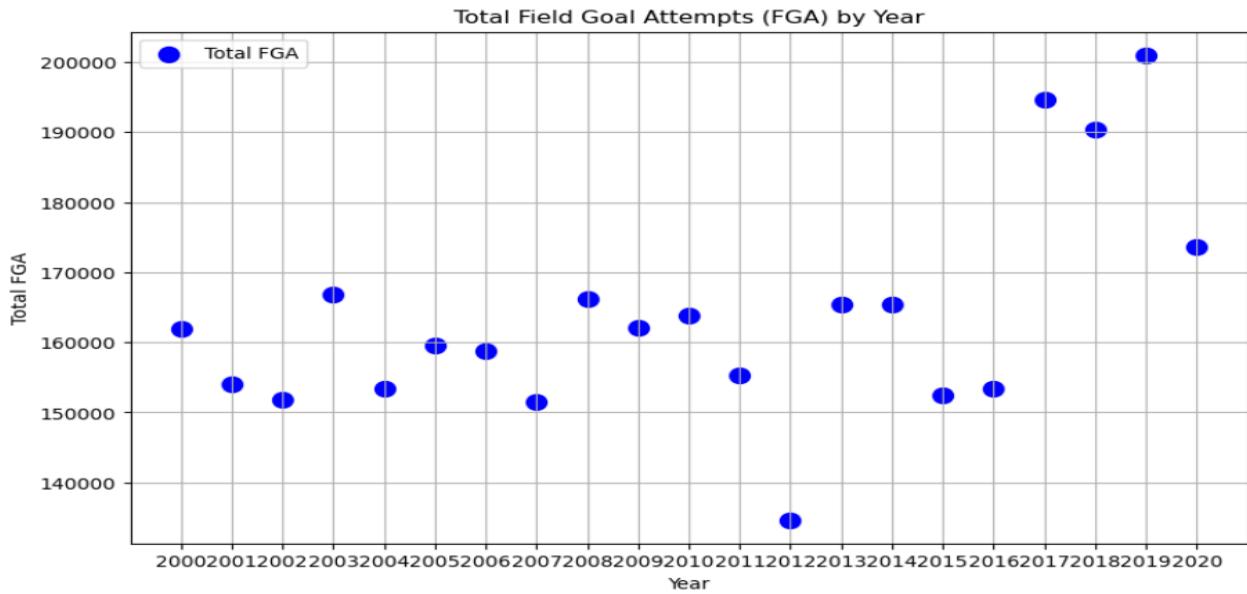


Figure 2.12 – League's Field Goal Attempts by Year

The rationale behind creating this chart is for the purpose of discovering the underlying patterns of FGA within all of those years and trying to comprehend why it fluctuates in certain years. Therefore, the information discovered will be discussed later on.

2.2.4.5 FGA vs FTA for Each NBA Player

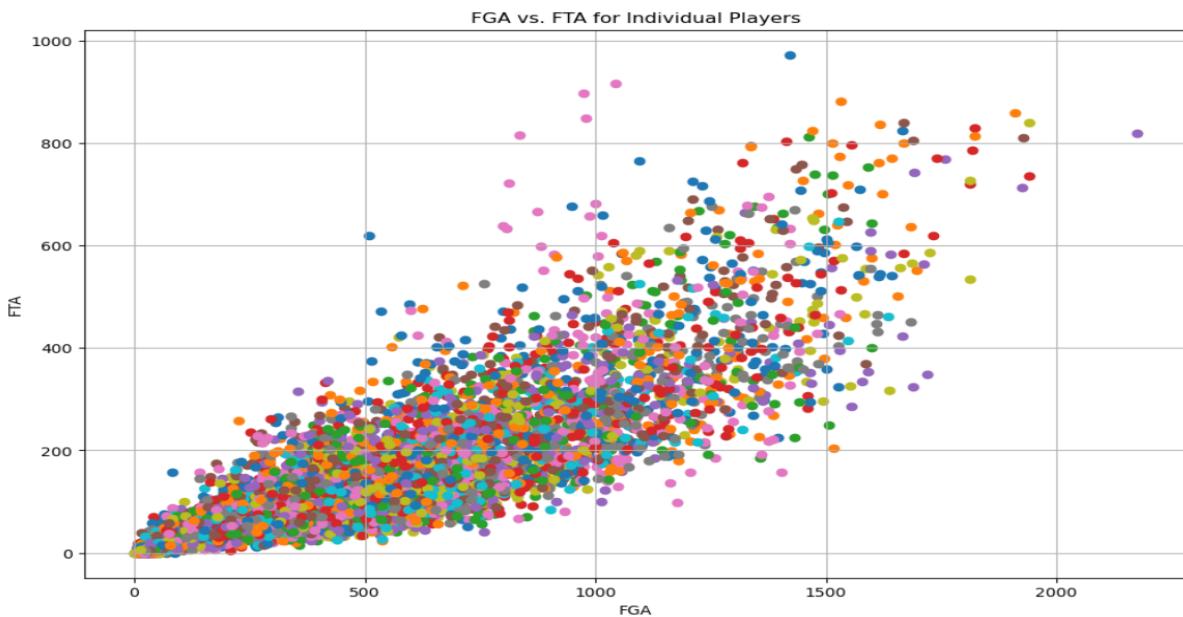


Figure 2.13 – FGA vs FTA for each NBA Player

The main cause for why researchers generate this scatter plot is due to having a better understanding in regards of the relationship between FGA and FTA for each player, and the need to know if the relationship is positively or negatively correlated. So, the results will be covered later.

2.2.4.6 Player's FGA vs TOV

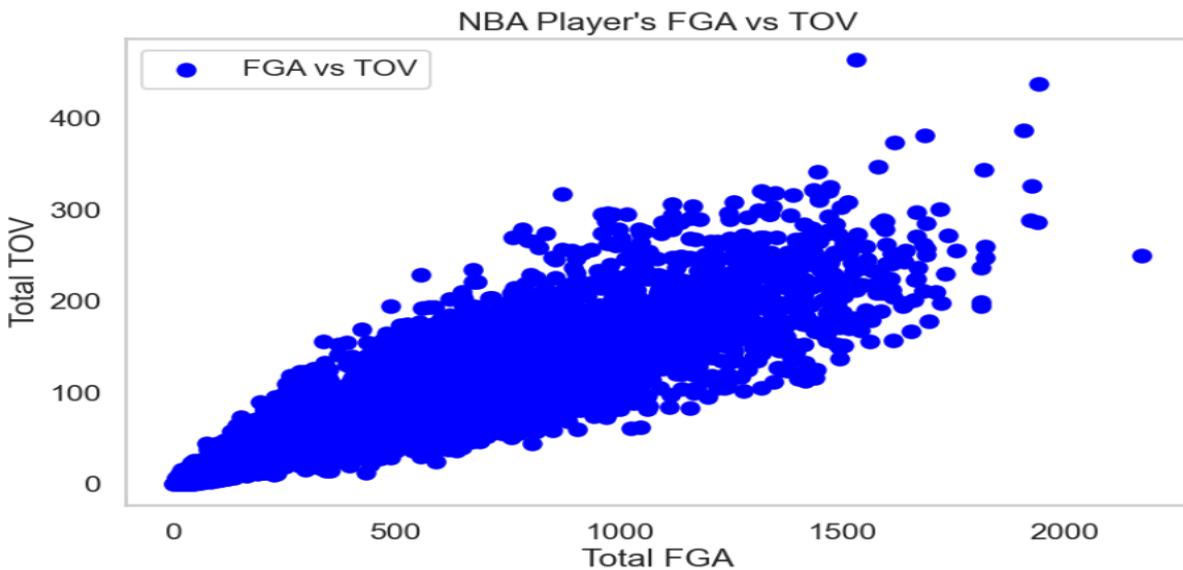


Figure 2.14 Player's FGA vs TOV

The primary reason for creating this plot is to discover any insights between players FGA and TOV, since most people do not have proper information regarding the relationship between

these two in terms of knowing if they have a positive or negative correlation. Therefore, the insights collected from this plot will be further explored in the discussion part later on.

2.2.4.7 Players PPG vs the Predicted Top 3 Players' PPG by Year



Figure 2.15 Players PPG vs the Predicted Top 3 Players' PPG by Year

The motive for researchers to plot this visualisation is to show the differences between these top players from the others, and what separates them from the rest in the context of Points Per Game. This will be explored further later in the discussion.

2.2.4.8 Players FGA vs the Predicted Top 3 Players' FGA by Year



Figure 2.16 – Players FGA vs the Predicted Top 3 Players' FGA by Year

The main cause of why researchers create this plot is to provide information to readers about the difference between total FGA in these three players from the rest of the players that are

lacking compared to them. Therefore, readers can understand how much they need to attempt in scoring in order to perform like them. Results of the insights identified will be discussed later.

2.2.4.9 Players 3PA vs the Predicted Top 3 Players' 3PA by Year

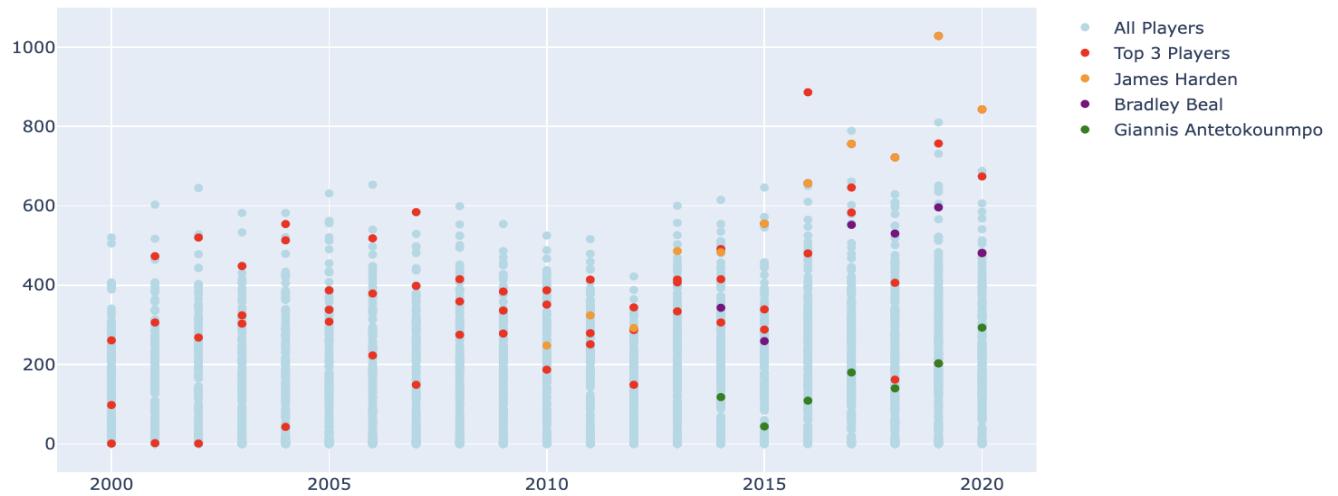


Figure 2.17 – Players 3PA vs the Predicted Top 3 Players' 3PA by Year

The justification for developing this graph is to visually compare the yearly Three Point Attempt (3PA) data for these three players with others. This will allow readers to understand how these players perform in three point shooting. The graph aims to unveil the 3PA counts for these players over a 20-year period, and the outcomes will be discussed in detail later.

2.2.4.10 Players ORB vs the Predicted Top 3 Players' ORB by Year



Figure 2.18 – Players ORB vs the Predicted Top 3 Players' ORB by Year

The reason for creating this graph is similar to the ones above, which is showing the difference in Offensive Rebounds (ORB) among these three players. Understanding the patterns of each players' ORB over the years is also the purpose of creating this visualisation. Hence, results generated will be discussed more later on.

2.2.4.11 Player's FTA vs the Predicted Top 3 Players, FTA by Year

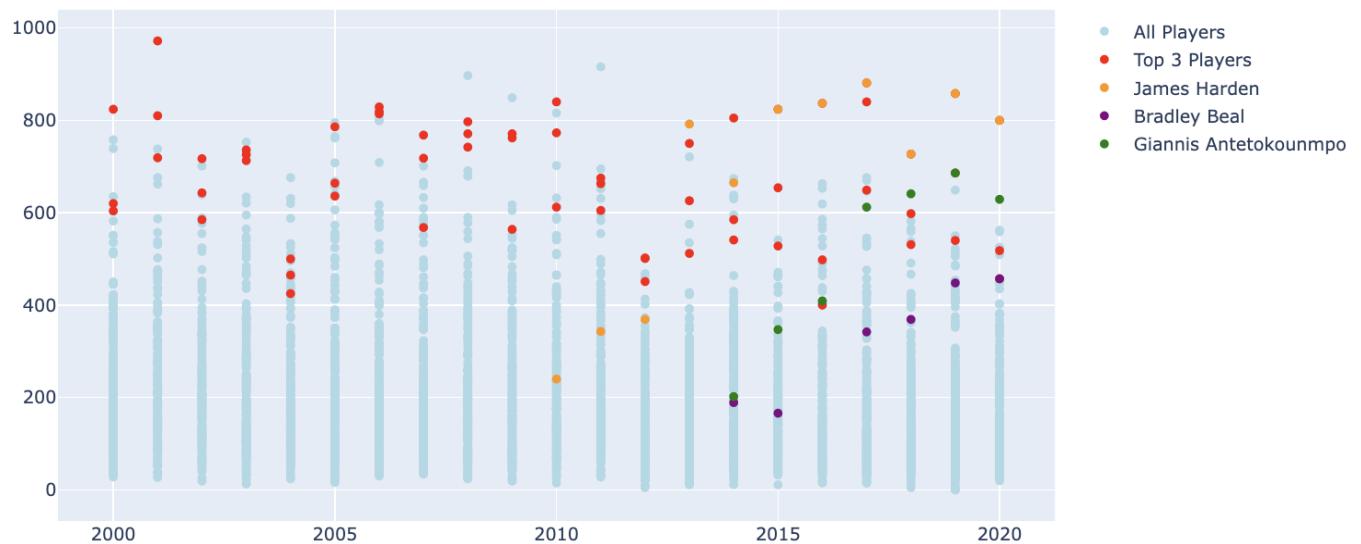


Figure 2.19 – Players FTA vs the Predicted Top 3 Players' FTA by Year

As what researchers understand, the creation of this visualisation is due to several reasons, such as to allow readers to learn more in terms of these players' Free Throw Attempt (FTA) within a 20-year period, and to discover valuable information regarding FTA among these players within this graph. Therefore, discoveries generated will be explored deeper later on.

2.5 Prediction

2.5.1 Algorithm

A statistical analysis model known as an autoregressive integrated moving average, or ARIMA, makes use of time series data to forecast future trends or to get a deeper understanding of the data set. When a statistical model forecasts future values by using historical data, it is said to be autoregressive. An ARIMA model may, for instance, anticipate a company's profitability based on historical periods or attempt to predict the future pricing of a stock based on its historical performance (Hayes, 2023).

2.5.2 Top 3 NBA Players with the Highest PPG in Regular Season in 2021

In our model, researchers will be using the ARIMA algorithm to predict the top three NBA players with the highest points per game during the season. The new dataset, created from nba_data during preprocessing, will be used in this prediction. Therefore, the following steps will outline the prediction process using ARIMA algorithm:

- Create an empty data frame to store the predictions
 - Create a for loop for the purpose of fitting each unique player into their respective ARIMA models. Thus, creating hundreds of predictions from the ARIMA models generated.
 - From all the predictions ARIMA models generated, researchers set the code so that it can choose the three highest PPG predicted from all the players PPG.
 - Store the predictions into the empty data frame created.
-

3.0 Results and Discussion

3.1 Visualisation

In this section, we will be showing the visualisation results using graphs and charts and explaining the reasoning behind them.

3.1.1 World Map Plot

3.1.1.1 NBA players across different countries

NBA Players Nationality

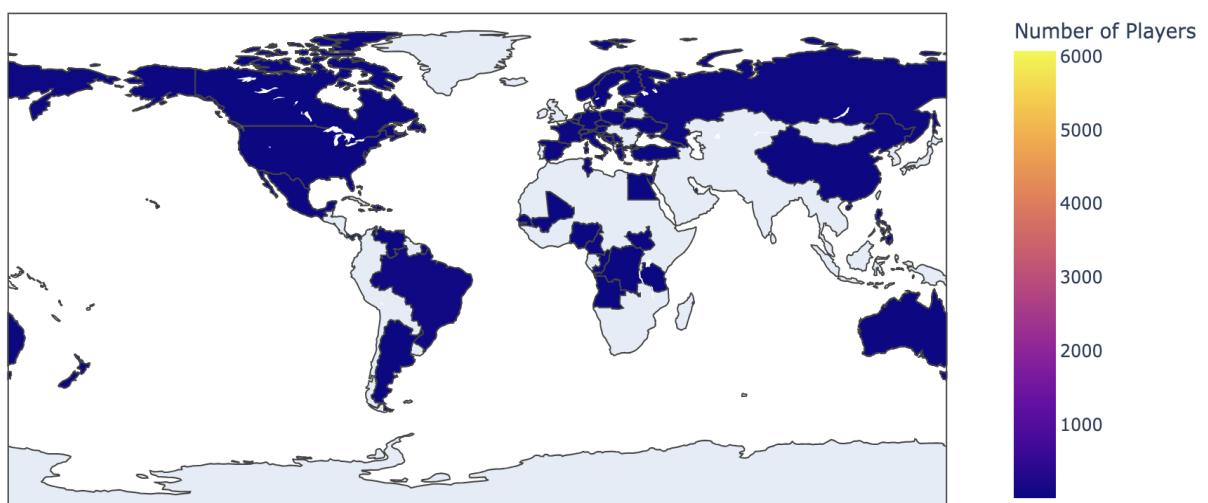


Figure 3.0 – NBA Players across different countries

Figure 3.0 provided us vital information, such as the number of players in each country that plays in the NBA. Moreover, as researchers discovered, the USA has the most number of players playing due to the fact that the NBA is the USA country's own basketball association. Furthermore, the country that occupied the second spot is France with 132 players, Canada taking the third place with 95 players, and so forth. But over the years, we have found out that international players have slowly joined the league to compete with USA players in this competitive league and will continue to increase in the future.

3.1.2 Bar Chart

3.1.2.1 Number of international players in NBA by year (non-US)

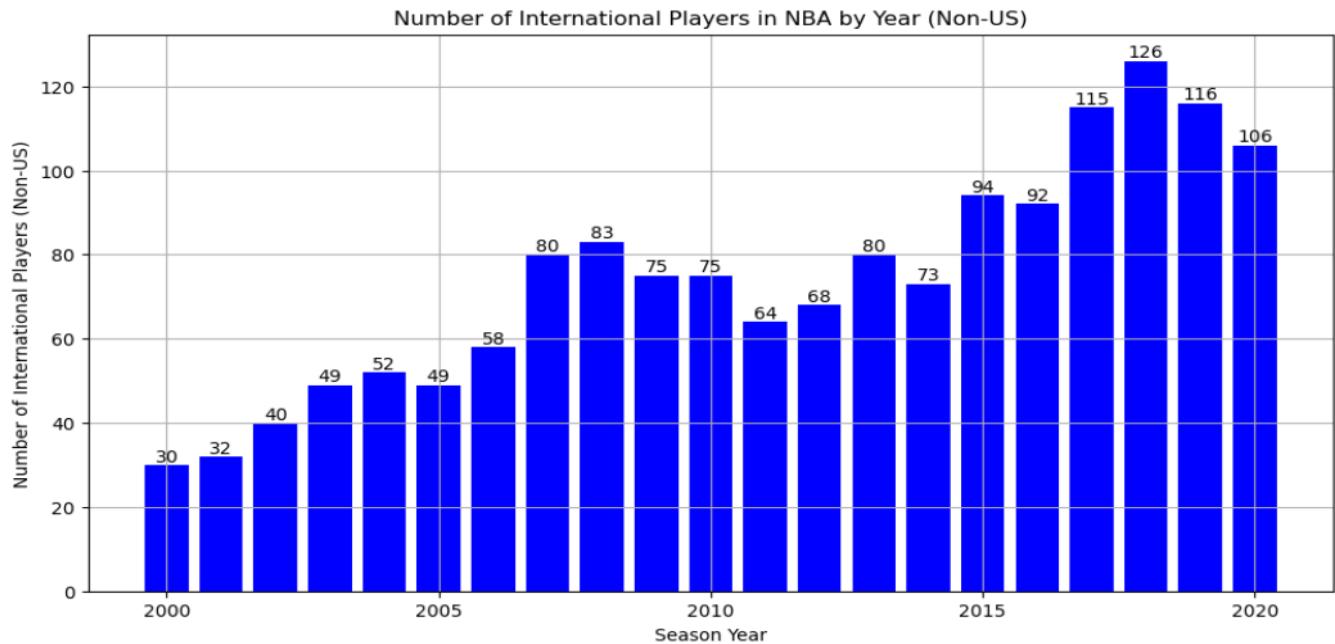


Figure 3.1 – Number of International Players in NBA by Year (non-US)

Figure 3.1 represents the number of international players (players from countries other than the United States) by year from 2000 to 2020 using a bar chart. As research had discovered, there was a fluctuating increase of international players in the NBA every year. This suggested an increase in the competitive standard of international basketball. Furthermore, the NBA is more open to drafting players from other countries in contrast to previous years.

3.1.2.2 Total 3PA per year

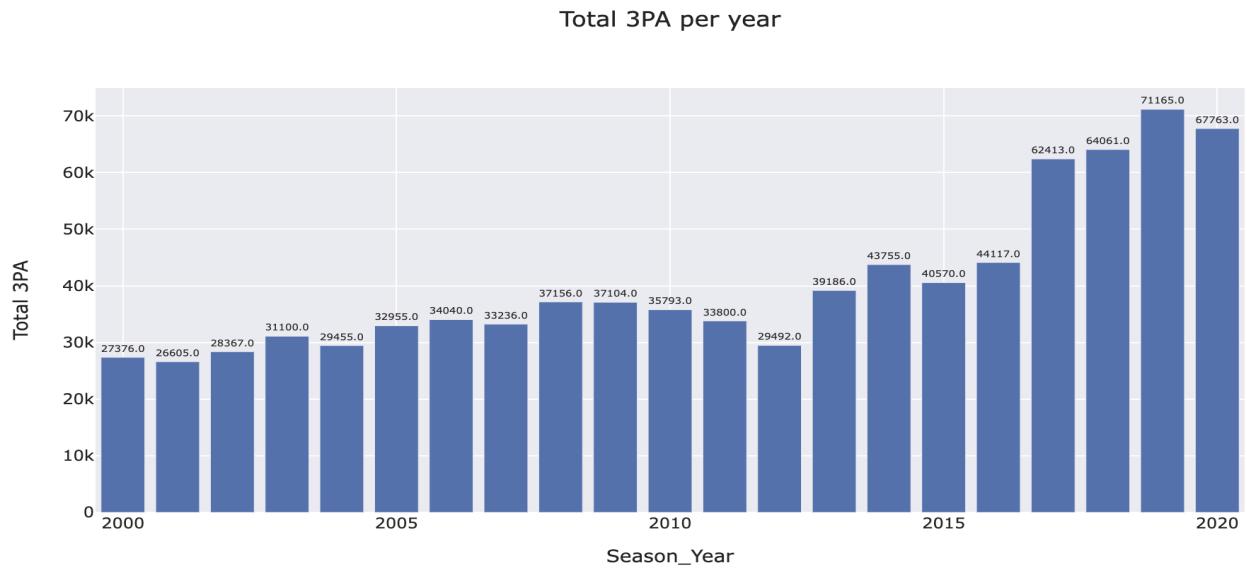


Figure 3.2 - Total 3PA per year

As you can see from Figure 3.2, it contained a bar graph comparing the total number of three point attempts made by the entire league through each year all the way from 2000 to 2020. Figure 2.2.2 showed that the number of three point attempts slowly increased from the year 2000 to 2008 and decreased slightly from the year 2009 to 2012 then increased exponentially from 2012 to 2019. This evidence showed that many players have started developing their three point shots and three point attempts have become more common than ever.

3.1.2.3 Total 3PA and FTA by Year

Total 3PA and FTA by Year

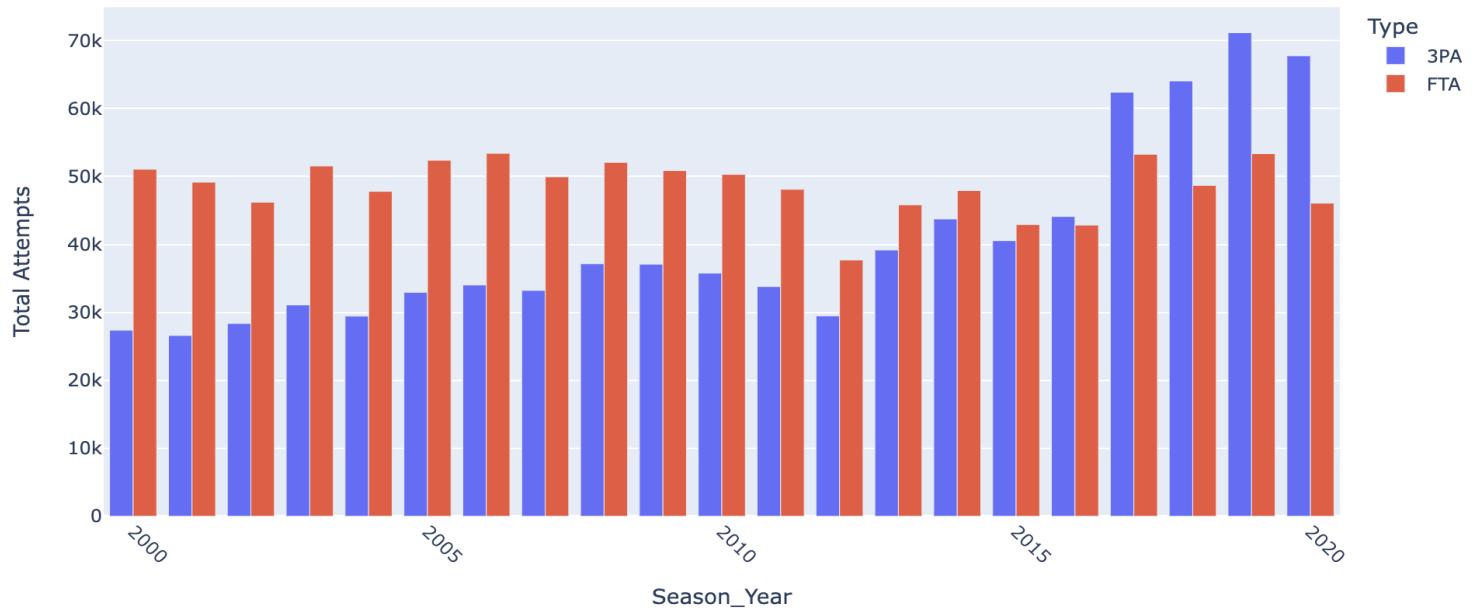


Figure 3.3 - Total 3PA and FTA by Year

As you can see from Figure 3.3, it showed the comparison between the total 3PA and the FTA by year. Before generating the bar chart, we expected the FTA to decrease every year since three point attempts were less physical. But surprisingly, the FTA remained consistent throughout the years with slight fluctuation as the 3PA increased. We came up with a theory that since players are shooting more threes now than ever, it opens up the space in the paint, which allows more attempts inside the paint.

3.1.2.4 Total Points for Each Pick in the First and Second Round Draft

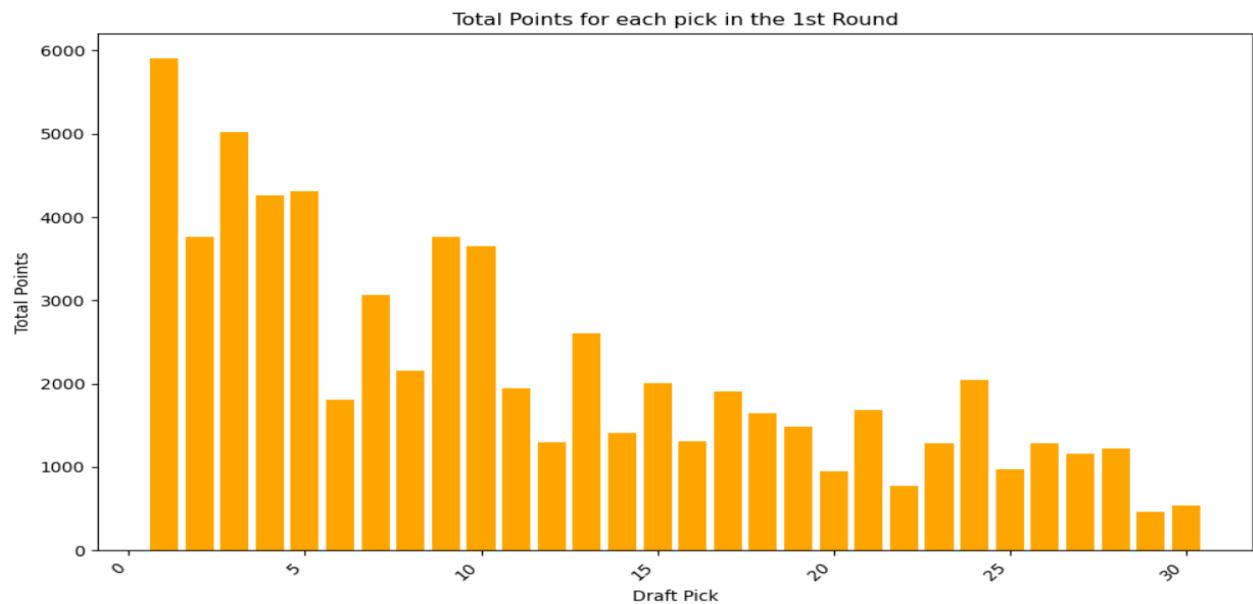


Figure 3.4 - Total points for each pick in the First Round Draft

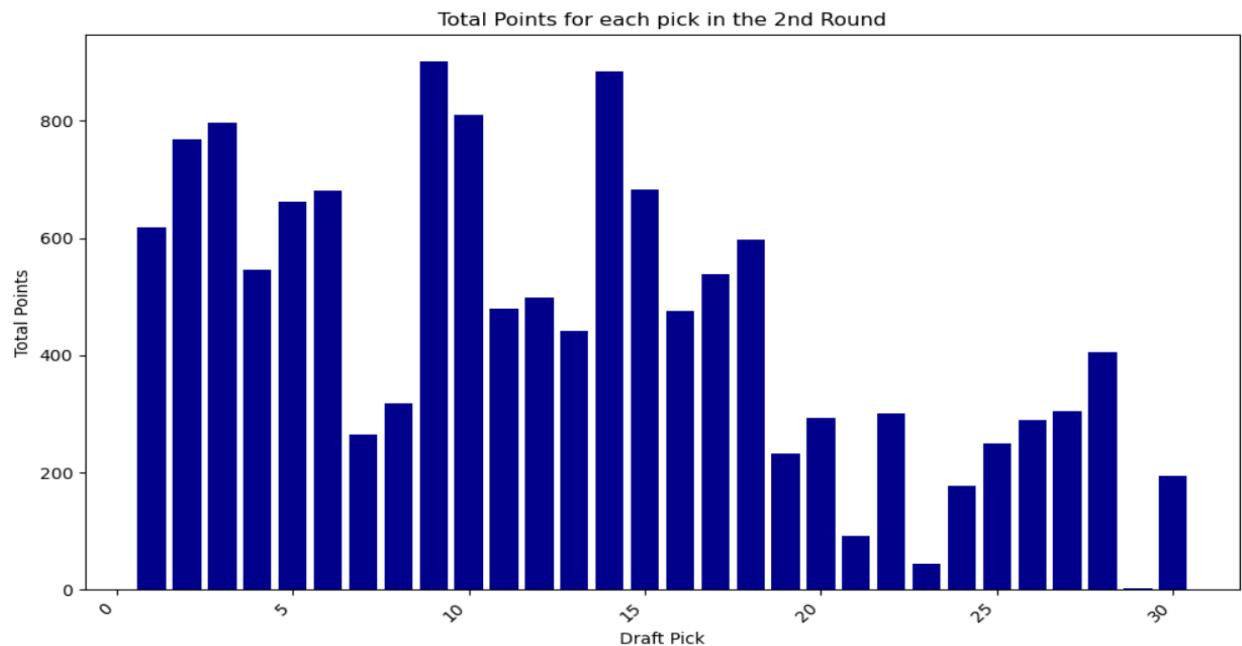


Figure 3.5 - Total points for each pick in the Second Round Draft

Based on both Figure 3.4 and Figure 3.5, you can see a bar chart showing the total points for each pick in the 1st round and 2nd draft rounds. Both figures showed that there was a significant difference in the draft picks from both rounds, from the draft picks in the first round, the total points generated from each pick decreased as the draft pick increased. However, in the 2nd draft round, the total points generated from each pick fluctuated randomly which showed that the draft pick number did not matter in the 2nd draft round. This implied that the rankings in

the 1st draft round mattered as the first few picks in it usually performed better but the rankings in the 2nd draft round did not matter.

3.1.2.5 Total Points Scored in Playoffs and Regular Season for both Draft Rounds

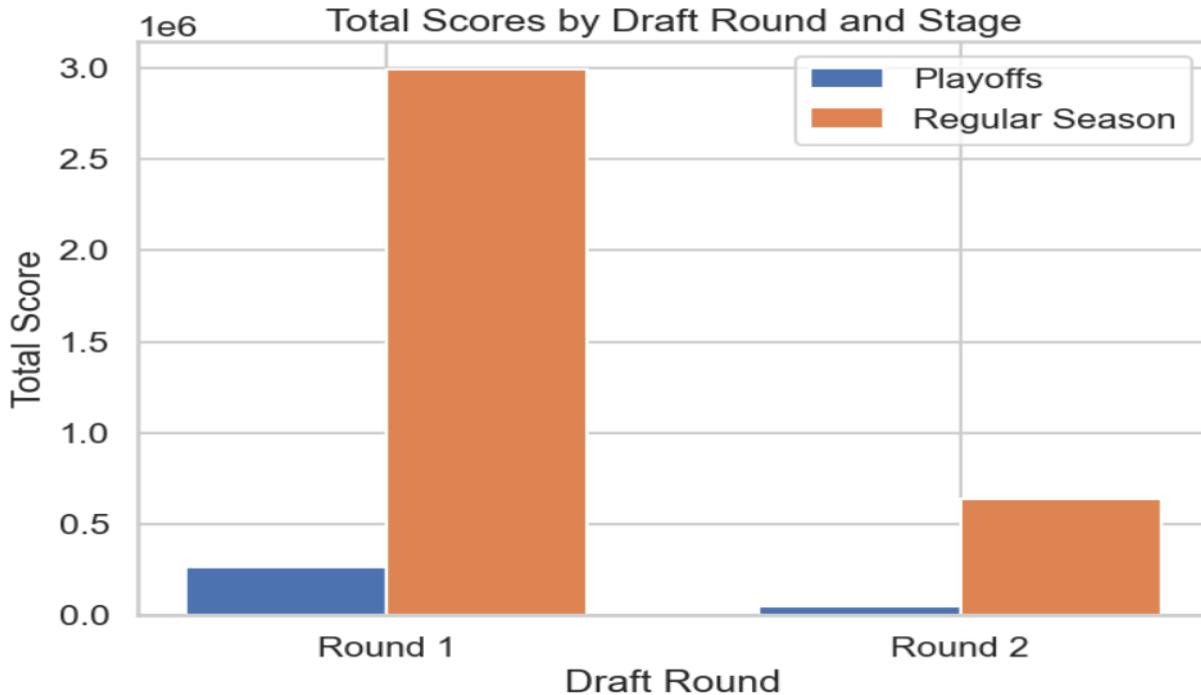


Figure 3.6 – Total Points Scored in Playoffs and Regular Season for both Draft Rounds

In Figure 3.6, you can see a bar chart that shows the total points scored by all the picks in both draft rounds in the playoffs and the regular season. In Round 1, the total points generated is nearly 4 times more than the points generated in Round 2 in both the playoffs and regular season. This showed that the players in Round 1 performed significantly better than the players in Round 2, which should be expected.

3.1.3 Line Chart

3.1.3.1 Average 3 Point Attempt of the USA vs International Players

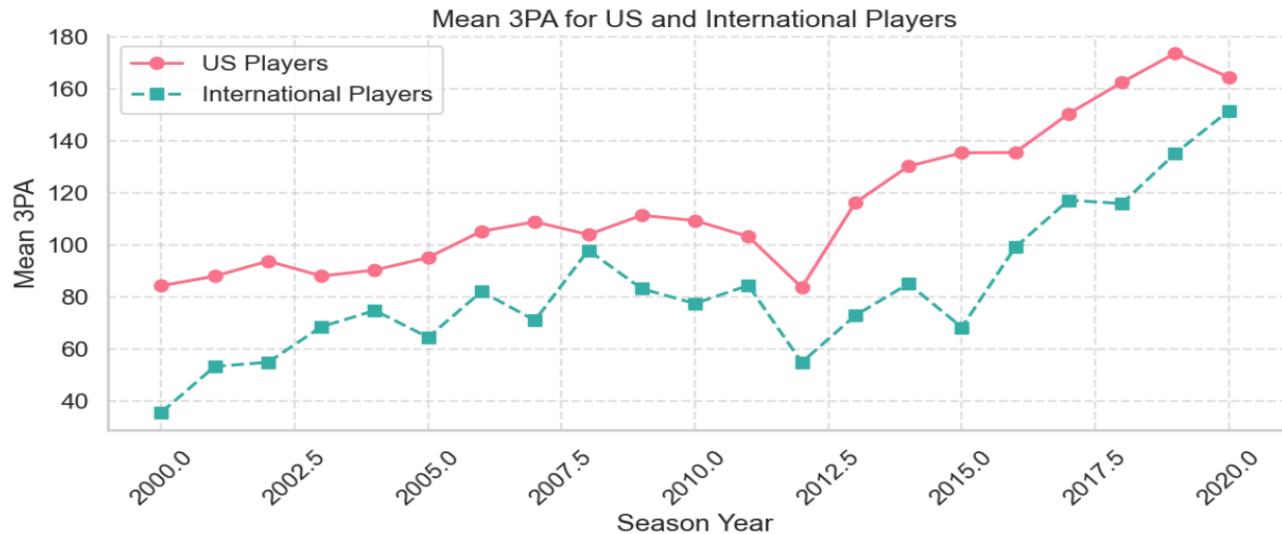


Figure 3.7 – Average Three Point Attempt of the USA vs International Players

Based on Figure 3.7 which is a line chart consisting of 2 lines, you can see the difference in the average three point attempts between the national players and the international players. The line chart indicated that national players attempted a slightly higher frequency of three-pointers compared to the international players.

3.1.4 Scatter Plot

3.1.4.1 Player's PPG vs FGA

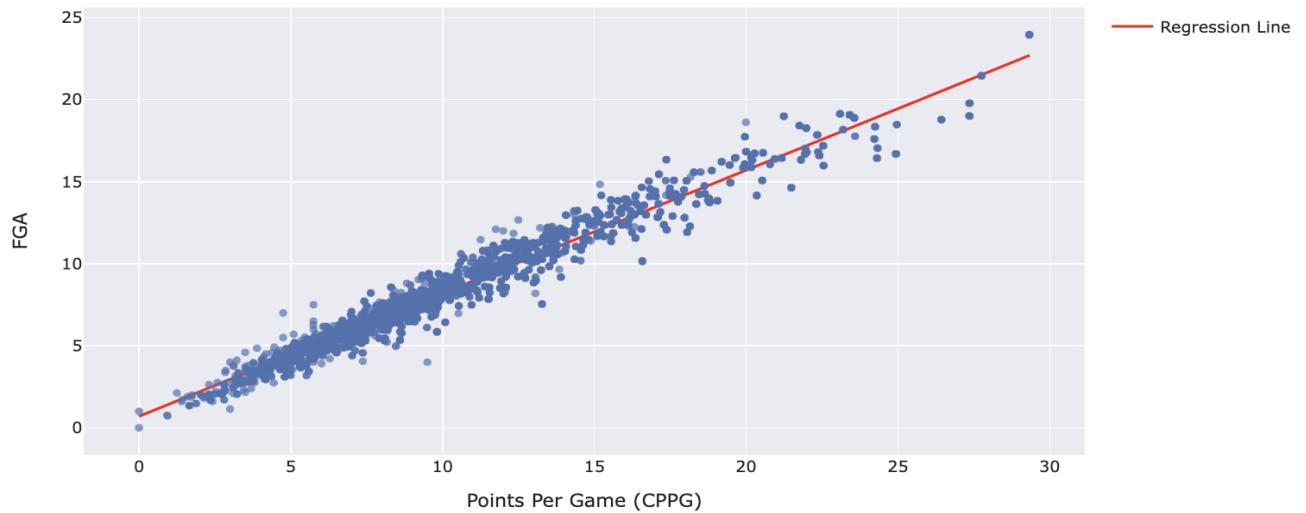


Figure 3.8 – Players PPG vs FGA

Figure 3.8 shows a scatter plot of the player's PPG vs their FGA. As anticipated, an increase in a player's shooting frequency is correlated with a rise in their points per game. Generally, players who performed well were more confident to shoot more shots, thus the reason for efficient field goals.

3.1.4.2 Player's Height vs Player's Career PPG

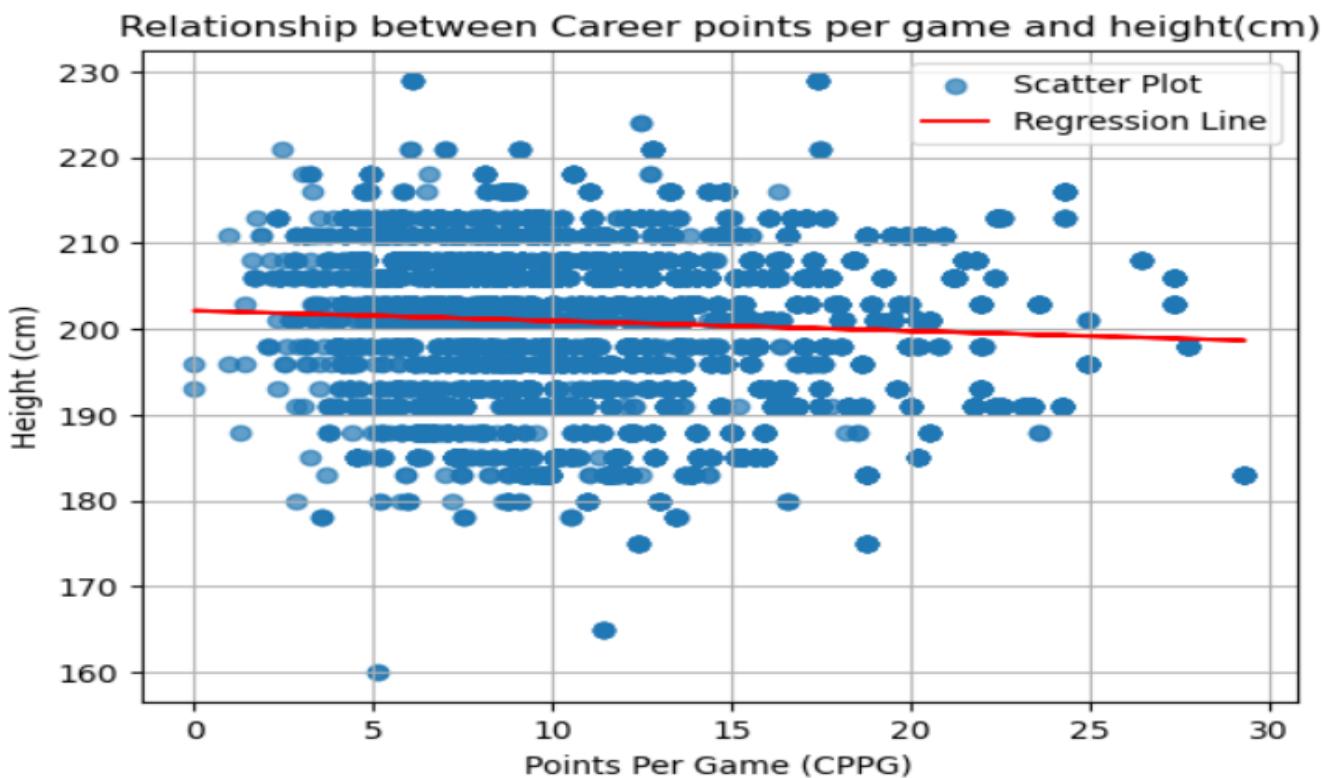


Figure 3.9 – Players Height vs Their Career PPG

Based on Figure 3.9, it represented the correlation between the player's height and their career points per game using a scatter graph. The graph showed that there was little to no correlation between the player's height and their career points per game. We came up with the theory that because the minimum height of an NBA player was already around 180cm (which is taller than the average person's height), the height did not matter much anymore in the NBA. Height was needed to enter the NBA but was not needed to perform well in the NBA.

3.1.4.3 League's Total Free Throw Attempts (FTA) by Year

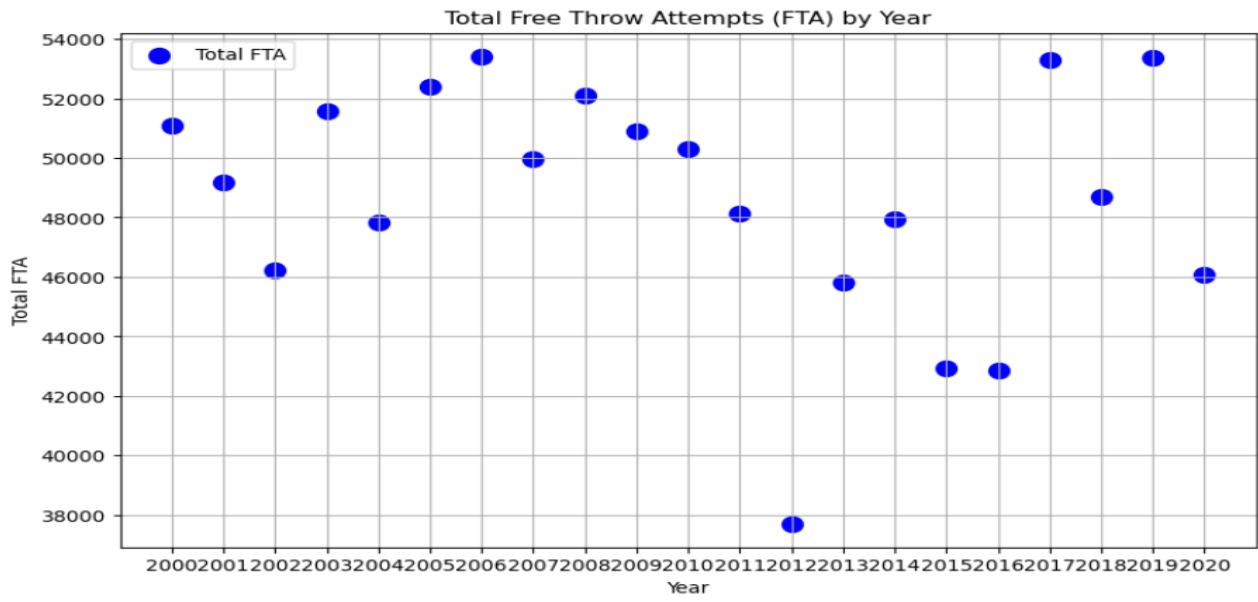


Figure 3.10 – League's Total Free Throw Attempt (FTA) by Year

From Figure 3.10, you can see a scatter plot of the league's total free throw attempt each year. The figure showed that the league's total FTA fluctuates every year but drastically dropped to around 38000 in 2012. Based on the finding, we came up with a theory that the defence that year was exceptionally good with way less fouls committed.

3.1.4.4 League's Field Goal Attempts by Year

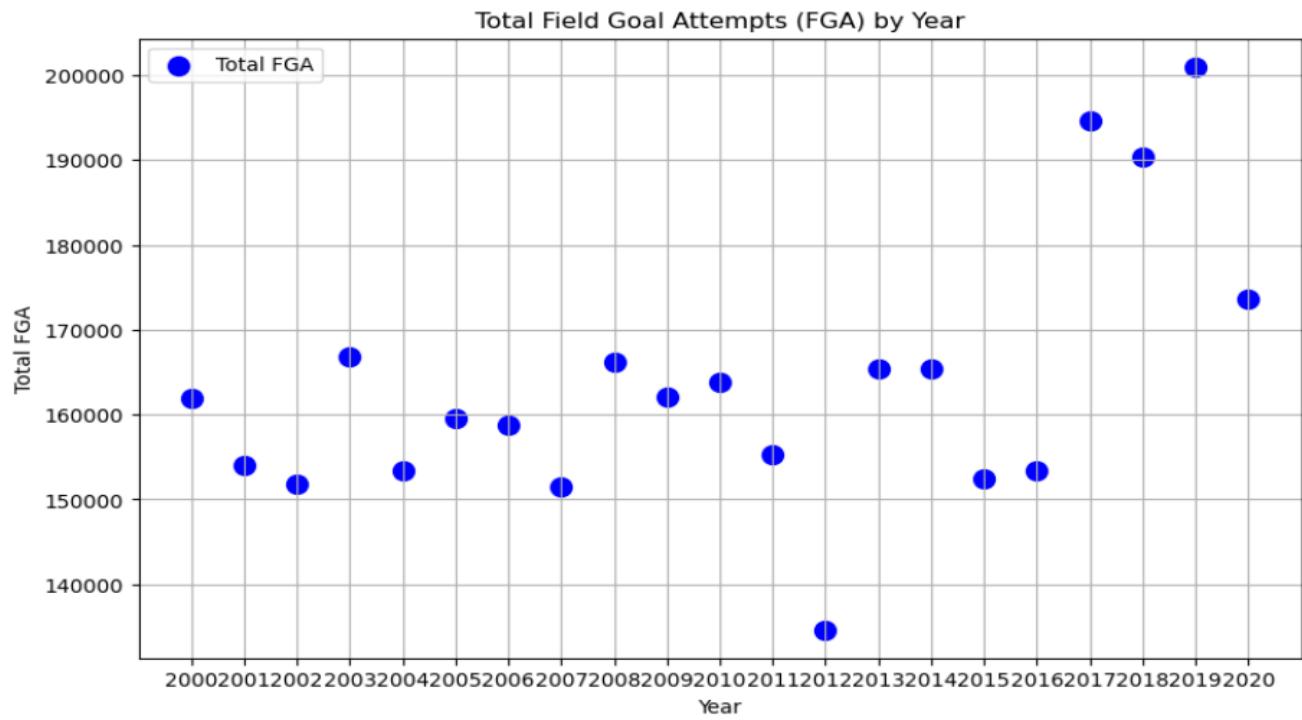


Figure 3.11 – League's Field Goal Attempts by Year

In Figure 3.11, it showed a scatter graph of the league's FGA every year from 2000 to 2020. Based on the graph, the league's total FGA fluctuated randomly until 2017. From 2017, there was a sudden jump in the league's total FGA that increased to around 195000. Based on the information provided, we concluded that one of the reasons of the sudden change was also related to the increasing number of 3PA every year. Additionally, FTA in figure 3.10 also increased suddenly in 2017, indicating that a lot of players had been fouled, thus players were eager to go to the paint to score two pointers, which added more in two points attempts (2PA). Therefore, FGA would increase in accordance with the increased in 3PA and 2PA.

3.1.4.5 FGA vs FTA for Each NBA Player

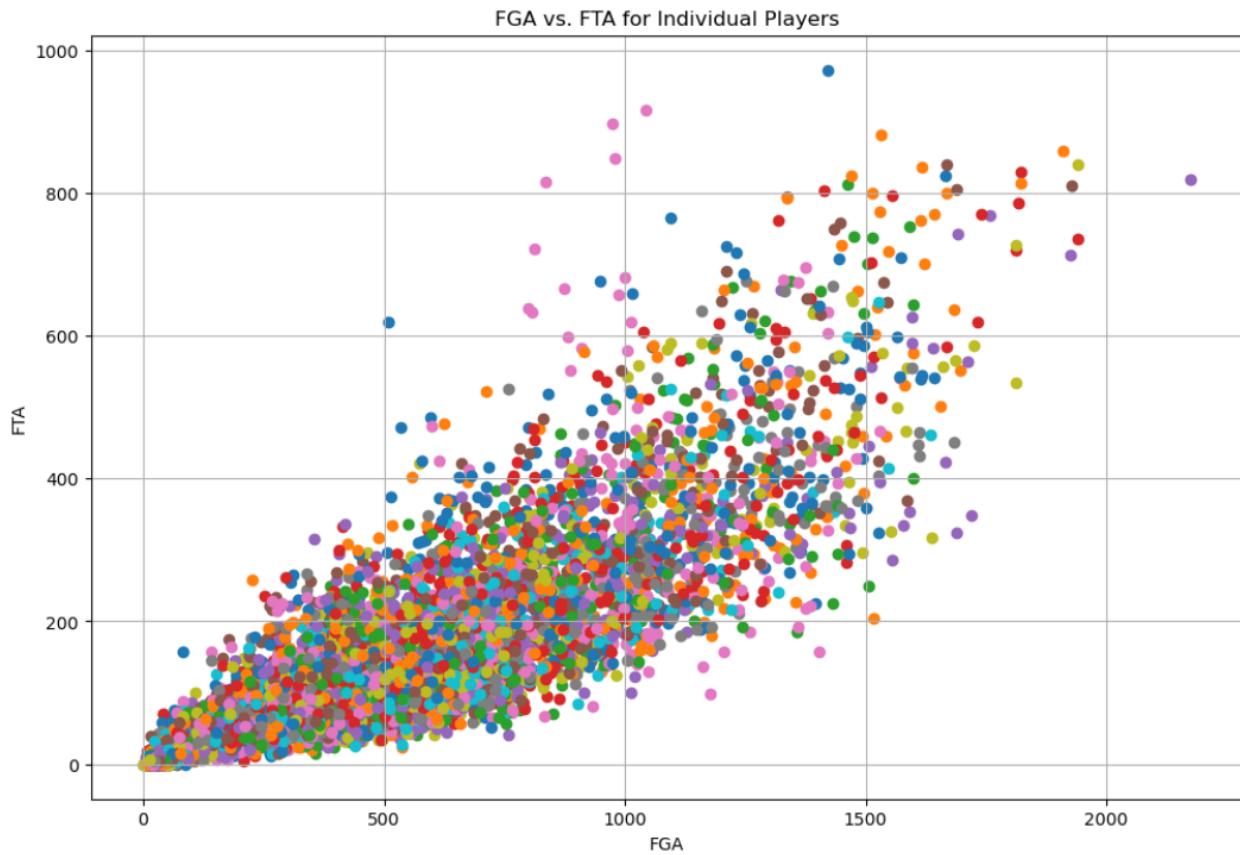


Figure 3.12 – FGA vs FTA for each NBA Player

In Figure 3.12, it showed a scatter plot comparing each player's field goal attempted with the free throw attempted for their whole career. Based on the graph, the number of free throws each player attempted increases as the number of field goal attempts increases. This implied that their inclination to take more shots not only increased the potential for drawing fouls but also increased the chances of getting onto the free throw line.

3.1.4.6 Player's FGA vs TOV

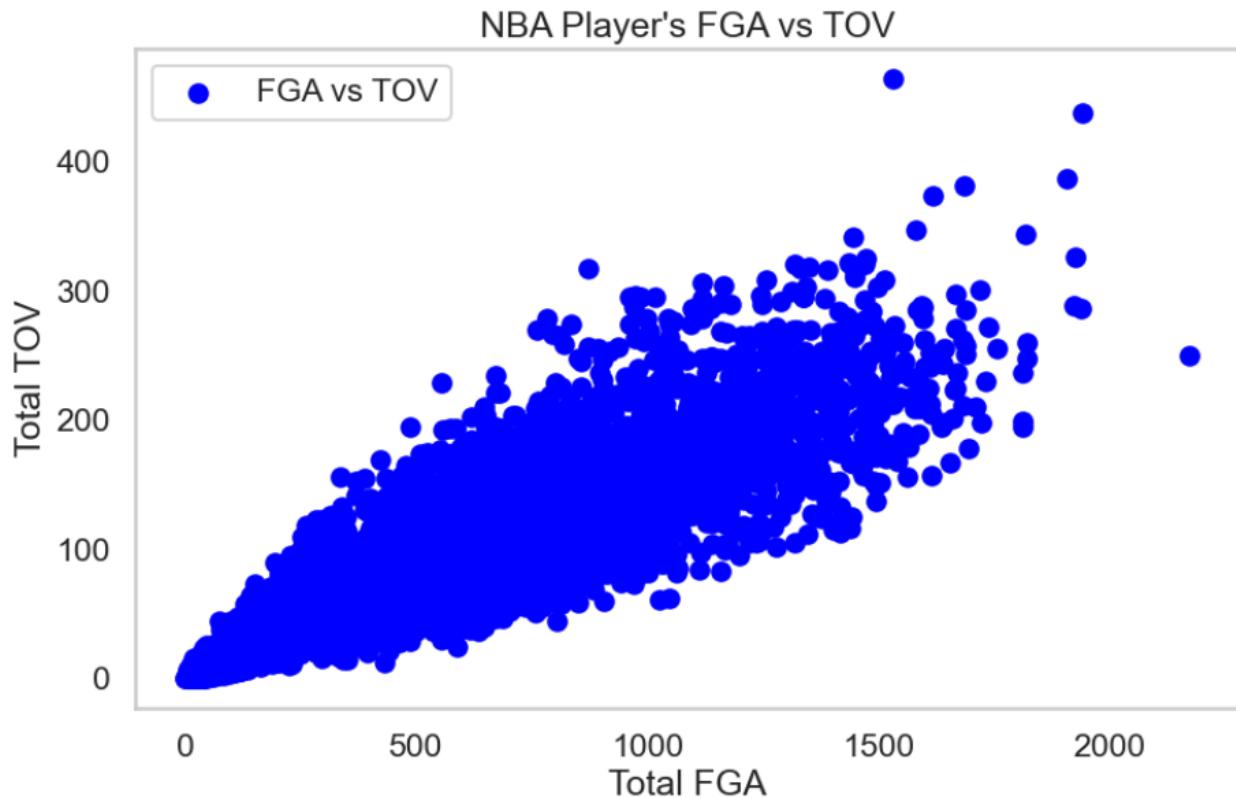


Figure 3.13 Player's FGA vs TOV

In Figure 3.13, it showed a scatter plot of the NBA player's FGA vs their TOV. As expected, the total number of TOV increased as the player's FGA increased. Most of the time, when a player had a higher Field Goals Attempted (FGA), it indicated a greater chance of them handling the ball, thus elevated the risk of turnovers.

3.3 Prediction

In this section, we will be explaining the results forecasted using the Arima algorithm and compare them with the actual results. Hence, before we delve into the results, the reason why we use comparative analysis to provide explanation about the result generated by the ARIMA model is due to our massive data size that contains hundreds of players' information with their PPG in every season they have played, the ARIMA model will perform loops for each unique players and predict the future year points per game (PPG), then from the lists of predicted outputs, the model will choose three players with the highest PPG as the result. Therefore, it is impossible to provide evaluation metrics in each ARIMA model generated, thus that is the reason for our decision to utilise a comparison method to evaluate the model's performance.

3.3.1 Top 3 NBA Players with the Highest PPG in Regular Season in 2021

3.3.1.1 Predicted Results

	Player	Predicted_PPG
131	James Harden	34.234559
223	Bradley Beal	33.209660
246	Giannis Antetokounmpo	29.442734

Figure 3.14 - Prediction of Top 3 players

Figure 3.14 showed the results of the predicted top 3 players with the highest PPG in the regular season in 2021. James Harden appeared as the leading scorer with 34.23 PPG, closely trailed by Bradley Beal with 33.21 PPG, and finally, Giannis Antetokounmpo secured the third position with 29.44 PPG.

3.3.1.2 Actual Results

Stephen Curry • GSW	32.0
Bradley Beal • WAS	31.3
Damian Lillard • POR	28.8

Figure 3.15 – Actual Top 3 players

Figure 3.15 showed the actual top 3 players with the highest PPG in the regular season in 2021. Only one player, Bradley Beal ,was predicted right as he was also in the second position with an actual PPG of 31.3. Based on what we have found, the reason why Stephen Curry wasn't predicted into the top 3 was because he did not play in the 2020 NBA regular season due to a serious injury. However, we were not able to find a reasoning for Damian Lillard as to why he was not predicted into the top 3. Furthermore, the reason why James Harden was not in the actual top 3 was due to the fact that the NBA made huge adjustments to their rule where players are not allowed to draw foul calls by flopping on purpose, which affected James Harden's scoring ability.

3.3.2 Justifications of the Prediction

3.3.2.1 Players PPG vs the Predicted Top 3 Players' PPG by Each Year



Figure 3.16 Players PPG vs the Predicted Top 3 Players' PPG by Year

Figure 3.16 showed every player's PPG throughout the years compared with the predicted top 3 players. James Harden, Bradley Beal, and Giannis Antetokounmpo exhibited a consistently rising Points Per Game (PPG) trend, outperforming the majority of players consistently over the past four years.

3.3.2.2 Players FGA vs the Predicted Top 3 Players' FGA by Each Year

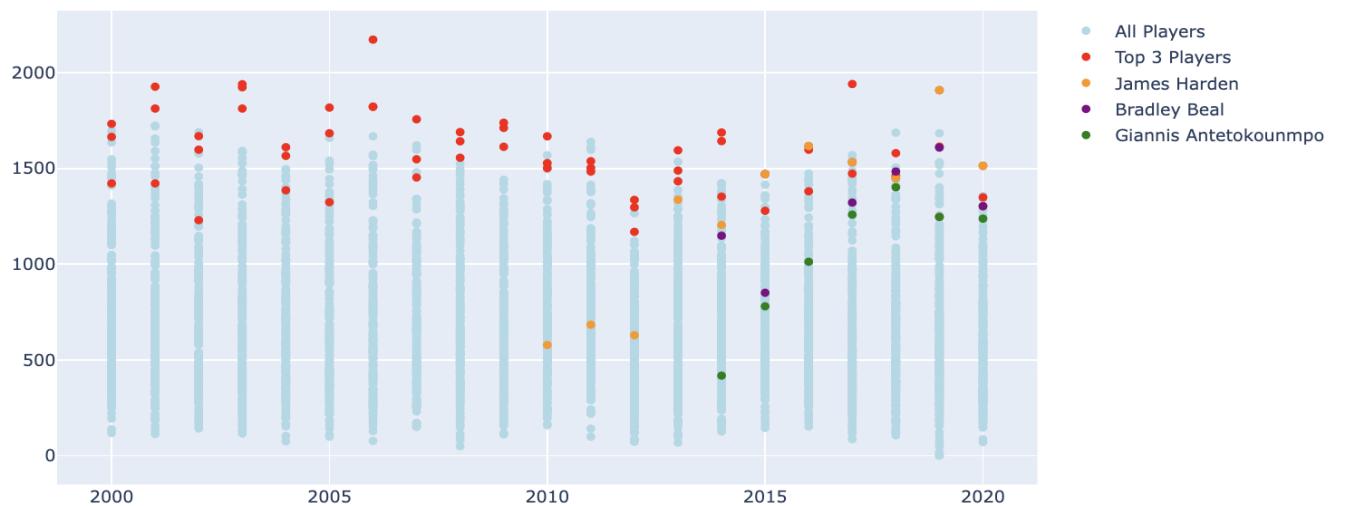


Figure 3.17 – Players FGA vs the Predicted Top 3 Players' FGA by Year

Figure 3.17 illustrated the players' FGA every year compared to the predicted top 3 players. The graph showed that James Harden, Bradley Beal, and Giannis Antetokounmpo had consistently maintained high levels of Field Goal Attempts (FGA), particularly in the last four years. This observation implied that an increased volume of attempts correlated with greater scoring opportunities for these players which led to a much higher PPG compared to other players.

3.3.2.3 Players 3PA vs the Predicted Top 3 Players' 3PA by Each Year



Figure 3.18 – Players 3PA vs the Predicted Top 3 Players' 3PA by Year

As shown in Figure 3.18, over the past four years, James Harden and Bradley Beal had consistently secured positions in the league's upper echelon for 3-point attempts (3PA), signifying that a notable portion of their scoring is attributed to their proficiency beyond the arc. However, Giannis Antetokounmpo's comparatively lower 3PA in contrast to James Harden and Bradley Beal implied that a significant share of Giannis' points is generated through 2-point shots.

3.3.2.4 Players ORB vs the Predicted Top 3 Players' ORB by Each Year

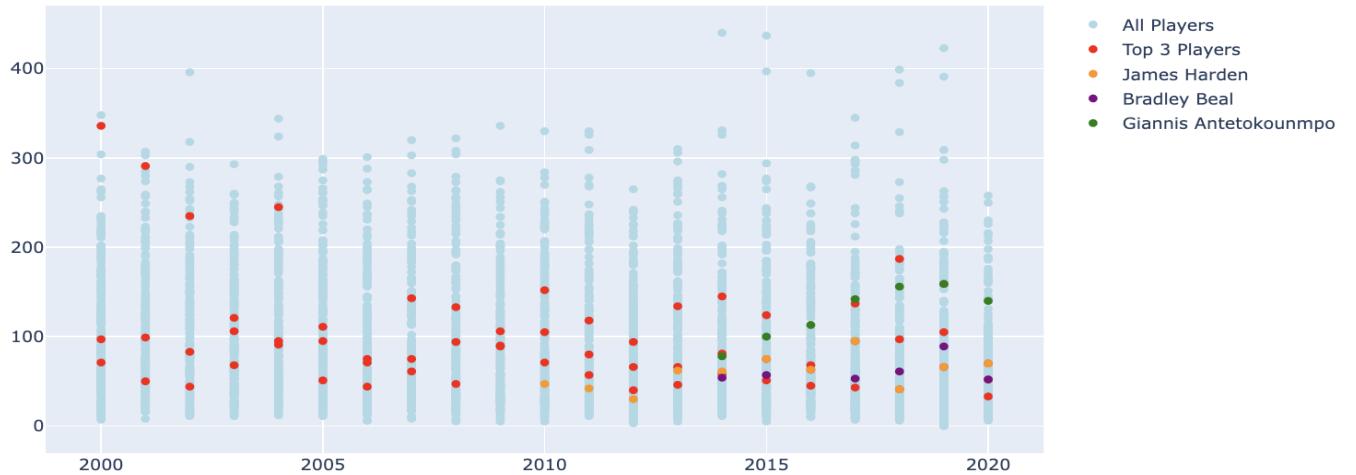


Figure 3.19 – Players ORB vs the Predicted Top 3 Players' ORB by Year

Based on Figure 3.19, you can see that James Harden and Bradley Beal consistently found themselves at the lower echelon of the league when it came to offensive rebounds. In contrast, Giannis Antetokounmpo boasted a significantly higher Offensive Rebound (ORB) count compared to the other two players, thus the increased chances to score in the paint. However, this observation suggested that offensive rebounding might not have a substantial impact on elevating James Harden and Bradley Beal's points per game (PPG), given that none of these players rank among the league's top offensive rebounders.

3.3.2.5 Players FTA vs the Predicted Top 3 Players' FTA by Each Year

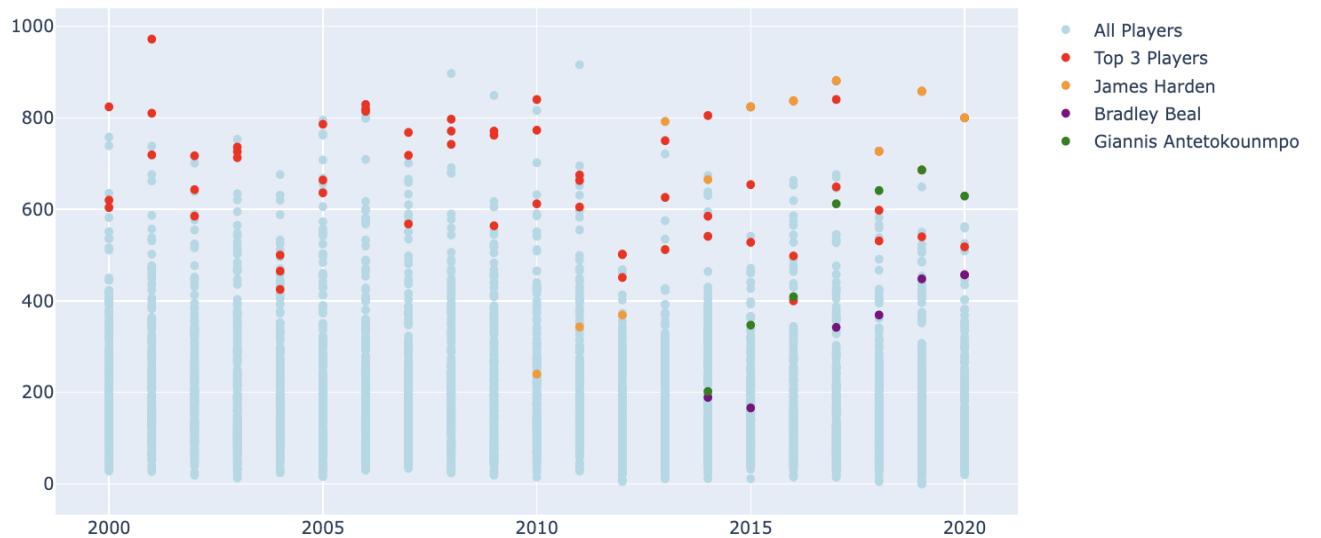


Figure 3.20 – Players FTA vs the Predicted Top 3 Players' FTA by Year

By examining Figure 3.20, it revealed a consistent pattern where James Harden consistently had a higher FTA in comparison to Bradley Beal and Giannis Antetokounmpo. This variation may serve as a potential factor contributing to Harden's position within the highest Points Per Game (PPG) among the three of them. Additionally, Bradley Beal stood out with the lowest FTA among the three, implying that a significant portion of his points is derived from successful field goals rather than free throws.

3.3.3 Study Comparison Analysis

In contrast to the study done by Miller, R. et al, which examined the forecasting performance of Regression, Holt-Winters Multiplicative Model (HWMM), and Seasonal Autoregressive Integrated Moving Average (SARIMA) models for the overall popularity index scores of the NFL and the NBA, this study focuses on predicting the performance of the top 3 NBA players using AutoRegressive Integrated Moving Average (ARIMA) models. While their research explored a variety of forecasting techniques, including regression-based methods and time series models, our research focuses on the prediction of player performance. Additionally, the similarities between our study and theirs lie in the data preprocessing stage and the time series models. However, the key difference is the emphasis on ARIMA models and their application to forecast the performances of specific NBA players, which contrasts with the wider scope of the study done by Miller, R. et al, which encompassed entire sports leagues.

4.0 Conclusion

The visualisation and prediction of NBA players' and teams' performance has been a great way to get insight into the field of basketball analytics. By means of thorough data analysis and advanced visualisation tools, we have achieved a deeper understanding of the players' and teams' statistics, trends, and possible outcomes. Even if our final forecast had a few inaccuracies, that was to be expected given the excessive number of variables in the NBA and the constantly changing data every minute and every second. In conclusion, the combination of current analytics and visualisation methods has improved our comprehension of NBA's players and teams and opened the door for more possibilities in the ever-changing state of the NBA.

Looking ahead, the future development of ARIMA models holds promise for further enhancements of NBA predictions. As technology advances and the field of basketball analytics continues to evolve, incorporating ARIMA models into forecasting techniques can contribute to more accurate insights into players' and teams' performance. The adaptability of ARIMA models works well with the ever-changing nature of sports analytics. The ongoing development of ARIMA models provides a good future for the field, where advanced analytical tools will play a crucial role in shaping how we understand and predict performance in the world of basketball.

5.0 References

- Hayes, A. (2023, September 29). *Autoregressive integrated moving average (ARIMA) prediction model*. Investopedia.
<https://www.investopedia.com/terms/a/autoregressive-integrated-moving-average-arma.asp>
- Lawton, G. (2022, January 31). *Data preprocessing: Definition, key steps and concepts*. Data Management.
<https://www.techtarget.com/searchdatamanagement/definition/data-preprocessing>
- Li, B., & Xu, X. (2021). Application of artificial intelligence in basketball sport. *Journal of Education, Health and Sport*, 11(7), 54–67.
<https://doi.org/10.12775/jehs.2021.11.07.005>
- Miller, R., Schwarz, H., & Talke, I. S. (2017). Forecasting sports popularity: Application of time series analysis. *Academic Journal of Interdisciplinary Studies*, 6(2), 75–82.
<https://doi.org/10.1515/ajis-2017-0009>
- Wang, Y., Liu, W., & Liu, X. (2022). Explainable AI techniques with application to NBA gameplay prediction. *Neurocomputing*, 483, 59–71.
<https://doi.org/10.1016/j.neucom.2022.01.098>
- What is data visualization?*. IBM. (n.d.). <https://www.ibm.com/topics/data-visualization>

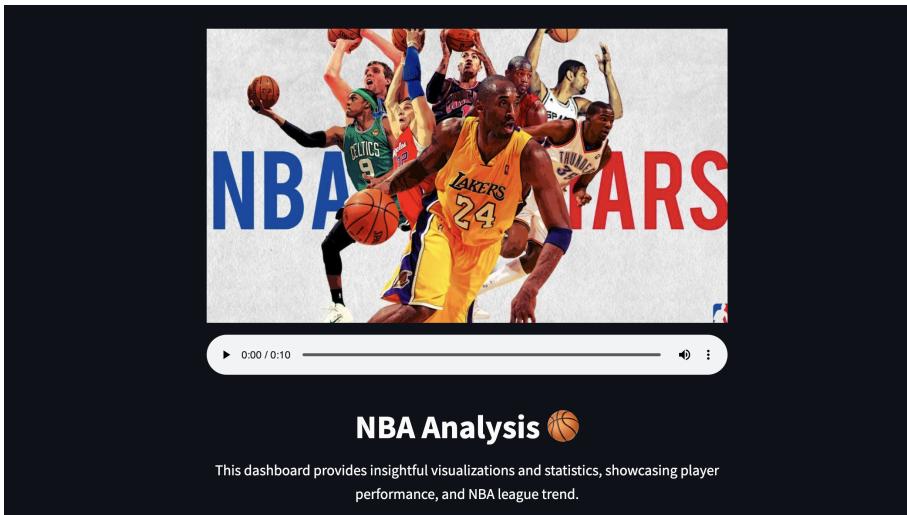
Appendix

Website Link

<https://nbaanalysistoolbox.streamlit.app/>

About Us

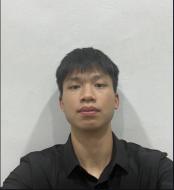
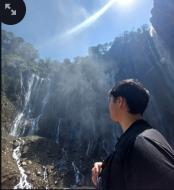
In this section, we only provide general information about the members of the group.



Topics

- Will number of international players influence the 3PA?
- Is height matters in NBA?
- Is it easier to get foul nowadays in NBA?
- Prediction on top 3 points per game player in NBA
- Does draft pick matters?
- Can NBA player improves their points per game by making more FGA?

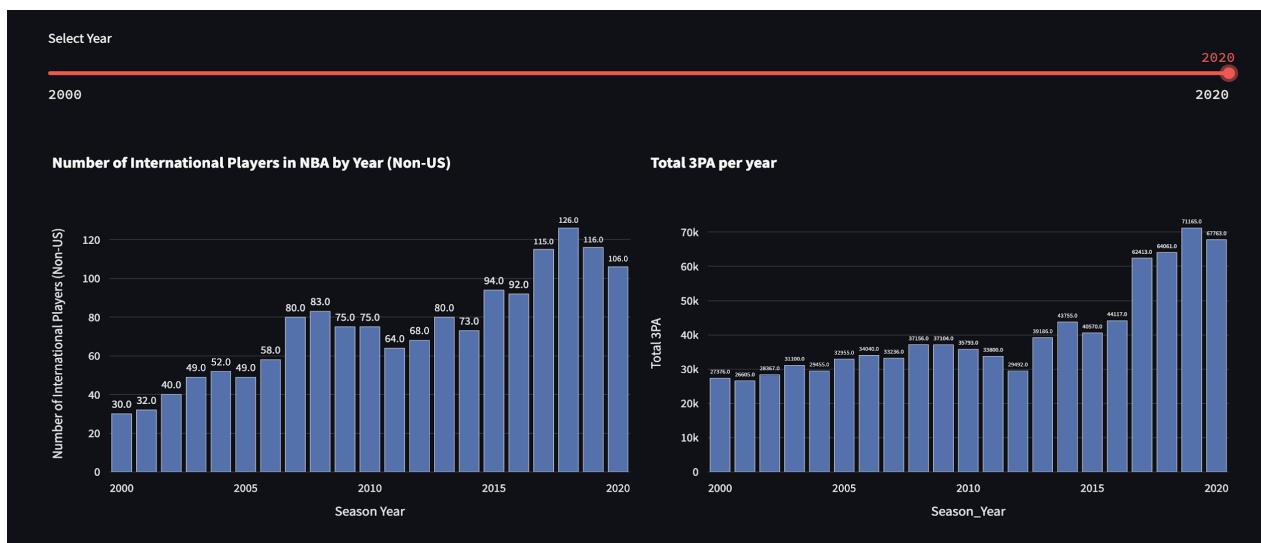
Our Team



Wong Yen Chik Marco Setiawan Ling Yang En

Main Topic

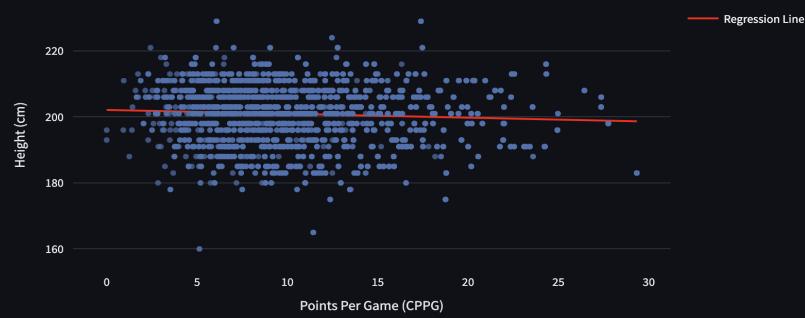
This is the section where we show all of the visualisations and predictions that we have discussed above. Thus, these are the website's image snippets





Is height matters in NBA?

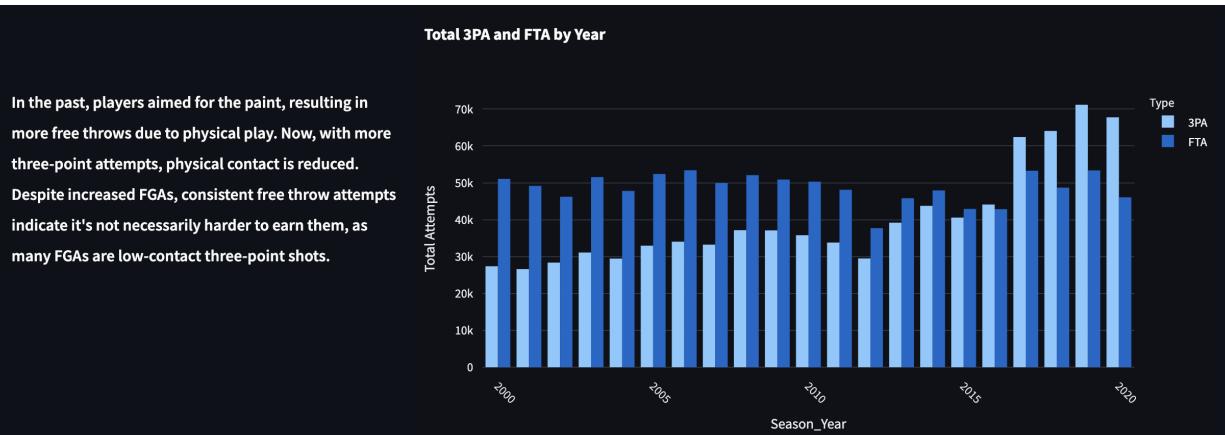
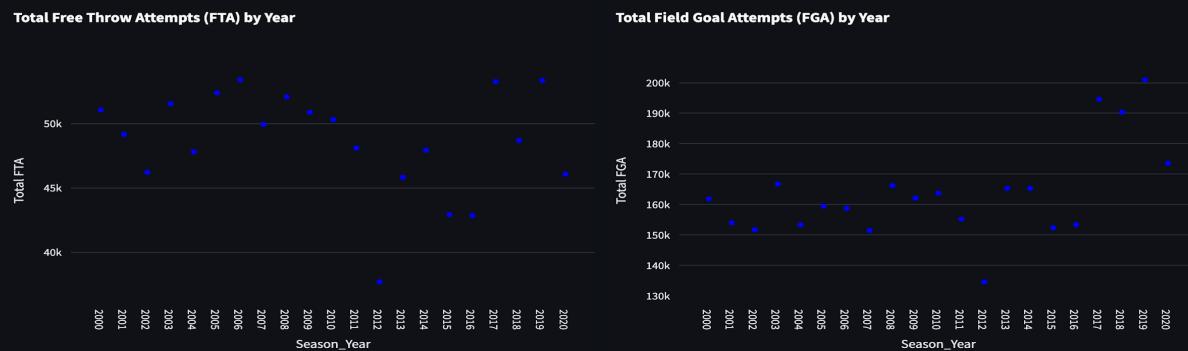
Relationship between Career points per game and height(cm)



Height may not significantly affect NBA performance, but it seems crucial for entry. Individuals around 180cm have a higher chance of making it to the league.

Is it easier to get foul nowadays in NBA?

We can see that the number of FTA each year does not fluctuate a lot but the number of FGA each year increased a lot in the recent 4 years. Hence, does it proves that it's harder to get free throw nowadays?



Prediction on top 3 points per game player in NBA

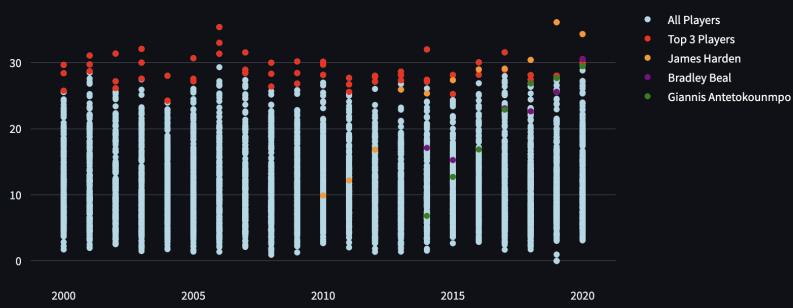
Select Year

2000

Year	Player	PPG
2000	Shaquille O'Neal	29.670886
2000	Allen Iverson	28.414286
2000	Grant Hill	25.756757

Why James Harden, Bradley Beal, and Giannis Antetokounmpo are predicted to be the top 3 points per game player in NBA in 2021?

PPG by each players each year



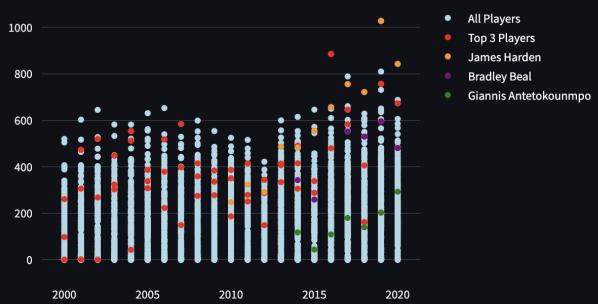
The PPG trend for James Harden, Bradley Beal, and Giannis Antetokounmpo is consistently upward, surpassing the majority of players over the last 4 years.

FGA by each players each year

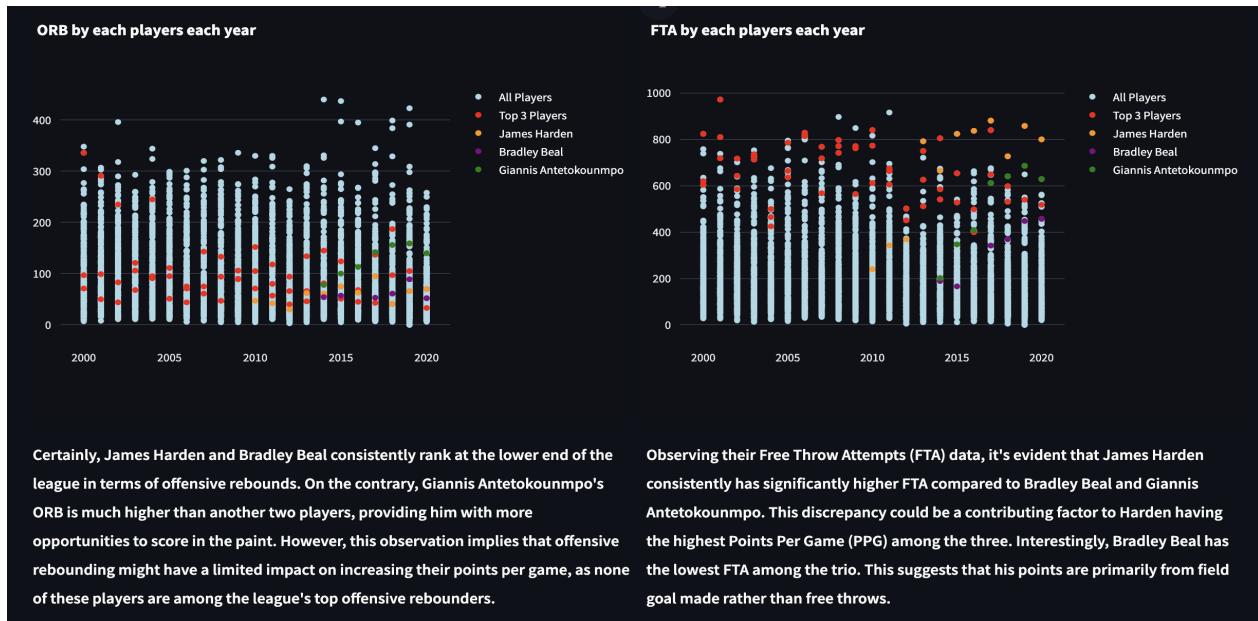


The graph clearly illustrates that James Harden, Bradley Beal, and Giannis Antetokounmpo have consistently high Field Goal Attempts (FGA), especially in the last 4 years. This suggests that a higher number of attempts leads to more scoring opportunities for these players.

3PA by each players each year

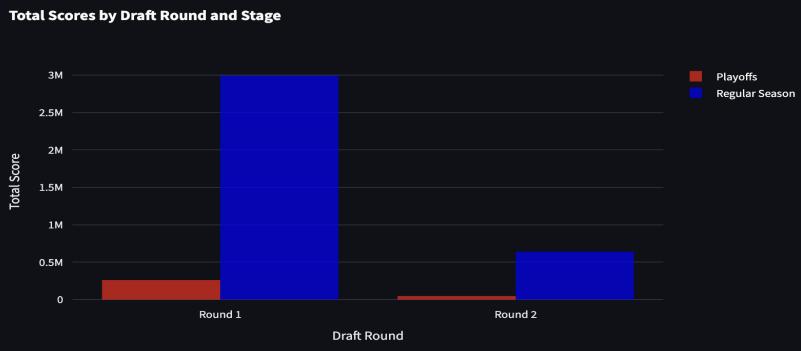


In the last 4 years, James Harden and Bradley Beal consistently rank in the league's top tier for 3-point attempts (3PA), indicating that a significant portion of their points come from 3 point line. However, Giannis Antetokounmpo's lower 3PA compared to James Harden and Bradley Beal suggests that a substantial portion of his points are derived from 2-point shots.

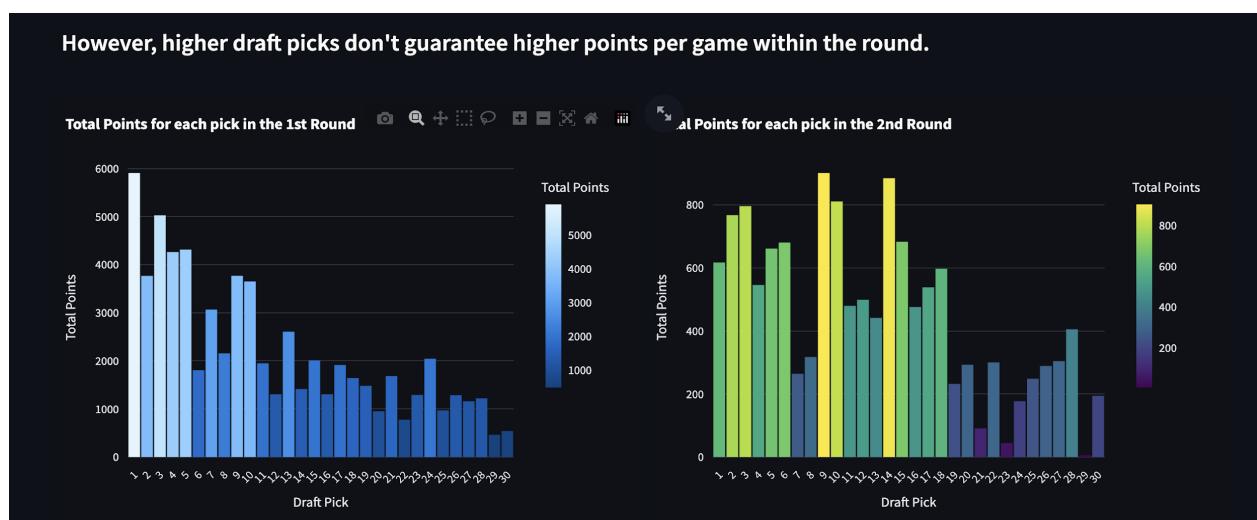


Does draft pick matters?

We can prove that round pick does matter as 1st round players will score more than 2nd round players.



However, higher draft picks don't guarantee higher points per game within the round.



Can NBA player improves their points per game by making more FGA?

