# NBA Player Data Analysis Project

## Overview

The goal of this project was to answer a few questions on a dataset to come up with the most informative and insightful takeaways I could find. In the field of analytics, often times we are tasked with finding insights regarding a dataset we have never seen before, and this includes data cleaning, data manipulation, data visualization, and ultimately data analysis to best derive meaningful value. I used a dataset on NBA players I found on Kaggle to simulate this real-world analytics problem that many of us deal with on a day-to-day basis. The code used to conduct this analysis is attached in this GitHub repository.

As per Kaggle, this dataset contains over two decades of data on each player who has been on an NBA teams' roster. It captures demographic variables such as age, height, weight, and place of birth, along with details like the team played for, draft year, and round. In addition, it has basic box score statistics such as games played, number of points, rebounds, etc.

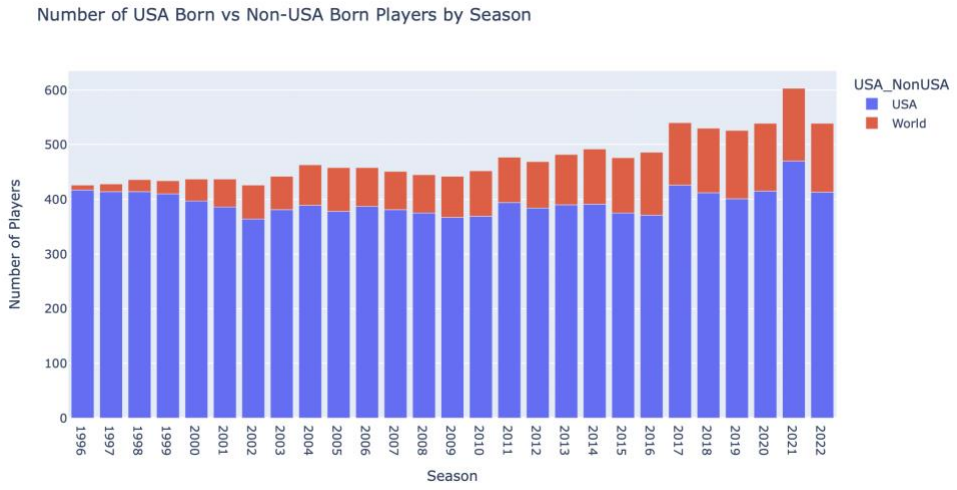
Because business case scenarios require both easily digestible information and actionable insights, I aimed to answer the following questions with further explanations as to why I felt they were important questions to tackle:

Q1. Count Number of USA Born vs Non-USA Born Players by Season
Q2. Highest PPG of USA Born vs Non-USA Born Players by Season
Q3. Average PPG of USA Born vs Non-USA Born Players by Season
Q4. Average age of Players by NBA conference and by Season
Q5. Count players that avg over 20ppg by conference and by Season
Q6. Blue Blood College vs Not PPG by Season
Q7. Count of Blue Blood College vs Not Blue Blood College Players by Draft Position
Q8. Blue Blood College vs Not Average Usage by Season
Q9. Count players by Season that have scored over 20ppg based on draft round
Q10. Count players that Avg over 20ppg by Team
Q11. NBA Player Height by Season

In order to answer these questions, there needed to be some data cleaning and data manipulation. First off, I created a column called “USA_NonUSA” that indicated whether or not a player was born in the USA or not. For the college analysis, I created a column called “BlueBlood” which indicated whether or not the player attended a “Blue Blood” university or not (Blue Blood universities in basketball include the following schools: Duke, North Carolina, Kansas, Kentucky, Indiana, UCLA). For the conference analysis, I created a column called “conference” that indicated what conference a team belongs to (eastern or western). For the season analysis, I just included the start year of the season for easier analysis (2014-2015 was simply stated as 2014 season). For the NBA draft data, I made sure all of the rows in the dataset had players that were either undrafted or drafted in the 1$^{st}$ or 2$^{nd}$ round, as this is the way the modern NBA draft works. Lastly, I dropped columns that were irrelevant for the data analysis.

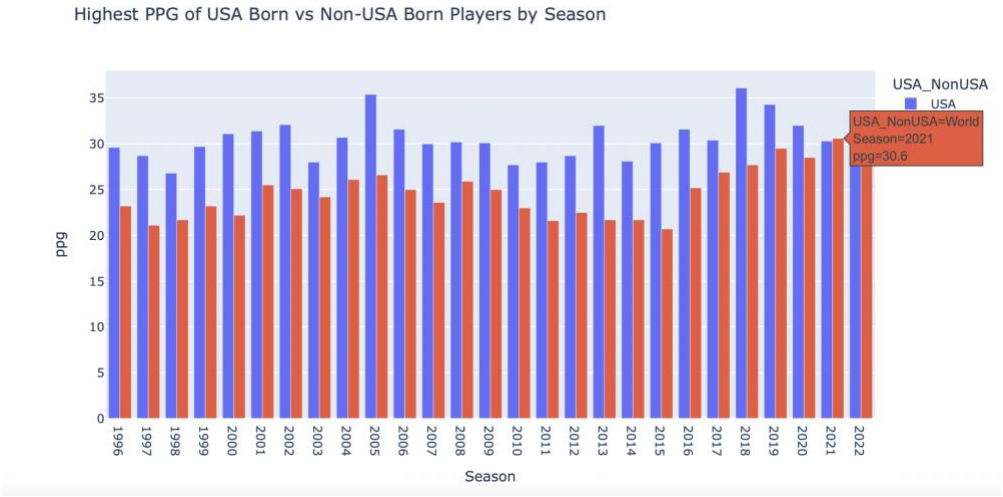
**Q1. Count Number of USA Born vs Non-USA Born Players by Season**

This graph was insightful because it showed the growing number of international born players that have been playing in the league, especially since 2014, as the number of international born players has been over 100 since that year and over 120 since 2019. It confirms the recent discussions that basketball is becoming more of a global game and could also impact drafting/free agent strategies since international players are getting better, which makes them more attractive to either draft or sign as a free agent. The graph of the data is below:



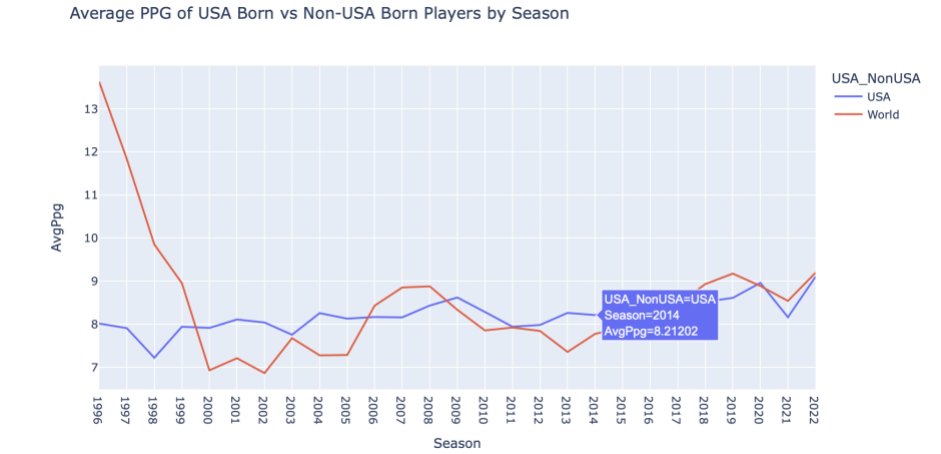

**Q2. Highest PPG of USA Born vs Non-USA Born Players by Season**

As we discussed earlier, the NBA is seeing a rise in international players in the league. This graph further illustrates the high quality of players coming to play in the NBA, as we can see especially since 2015 the highest scoring international player is scoring more than 25ppg. Additionally, both in 2021 and 2022 the highest scorer in the league was an international player. This again illustrates the previously mentioned point that top end international talent has become more attractive to teams to sign to their roster. The graph of the data is below:



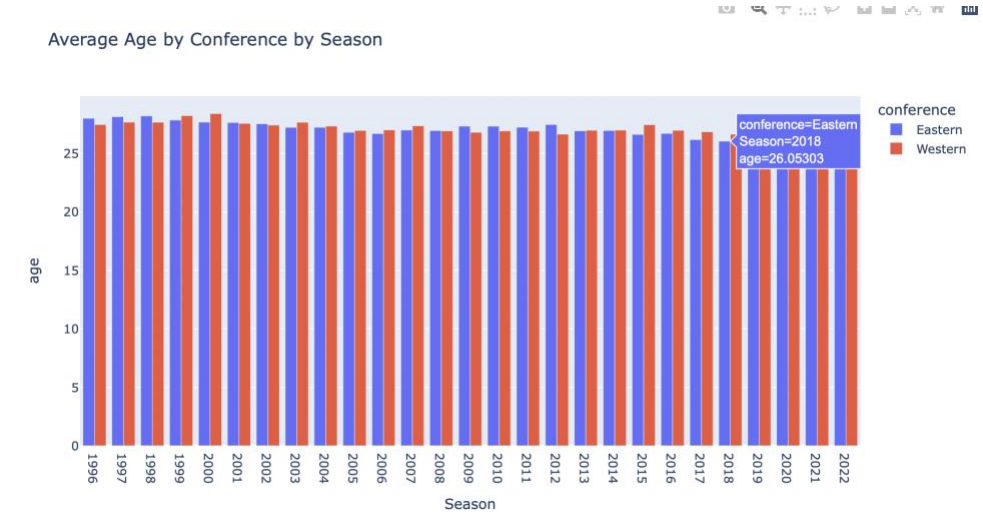
**Q3. Average PPG of USA Born vs Non-USA Born Players by Season**

This line chart shows us the average points per game of International vs USA born players over time. As we can see, the data steadies around 1999 even though from 1996-1998 international players averaged higher ppg than American players. One reason for this could be the fewer number of international players in the league at the time, and also the fact that the international players in the league then were generational talents. With the rise of scouting and more successful international leagues, it makes the game more accessible to talent evaluators, but that infrastructure was not present 25 years ago. Since 2006 the average ppg between the two groups is roughly even, further illustrating rise and success of international players in the NBA. The graph of the data is below:



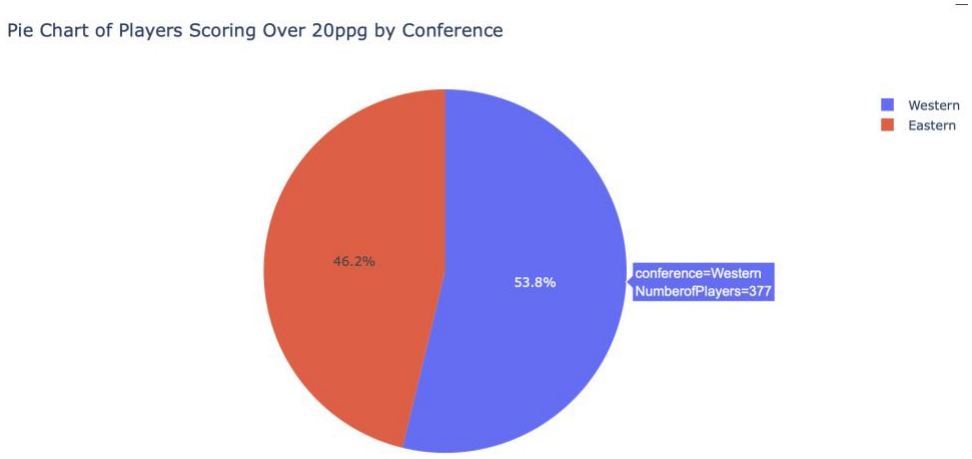

**Q4. Average age of Players by NBA conference and by Season**

This bar chart analyzes average player age over the seasons and by conference to see if there is a noticeable change over time or by conference in player ages. Unfortunately, there is no difference, which indicates neither conference prefers younger or more veteran players. This makes logical sense, as across the league team ages are determined by prior success, drafting strategies, and free agent acquisitions but this was interesting just to visualize. The graph of the data is below:



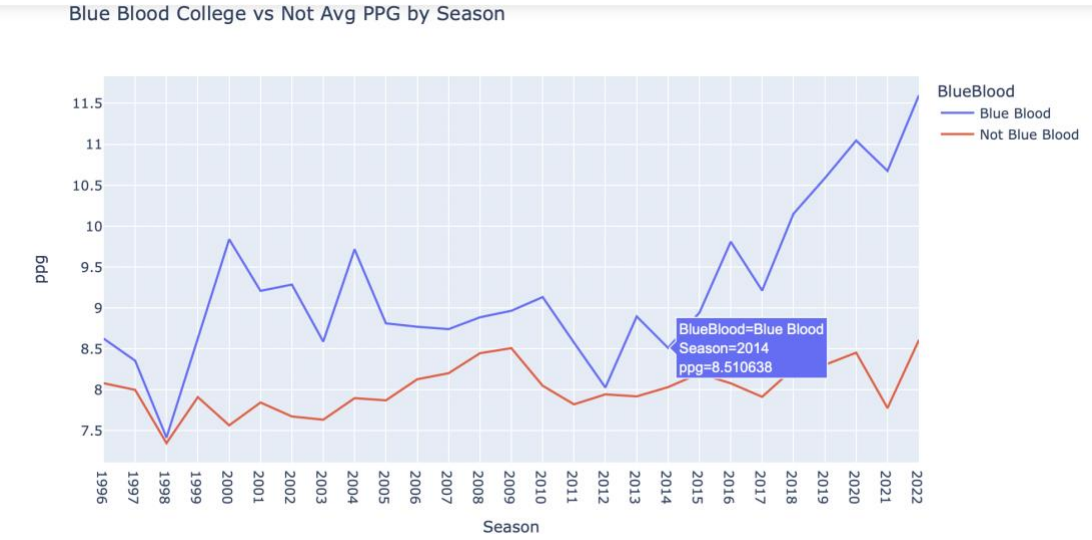
**Q5. Count Players that avg over 20ppg by conference and by Season**

This pie chart was interesting just to see if certain conferences had more higher scorers than others. As we can see from the graph, it appears the western conference had a higher percentage of players that were high scorers compared to the eastern conference, as about 54% of the high scoring players were in that conference compared to about 46% in the east. This makes sense, as the western conference has traditionally been the more offensive focused conference whereas the eastern conference is usually better on defense. The graph of the data is below:



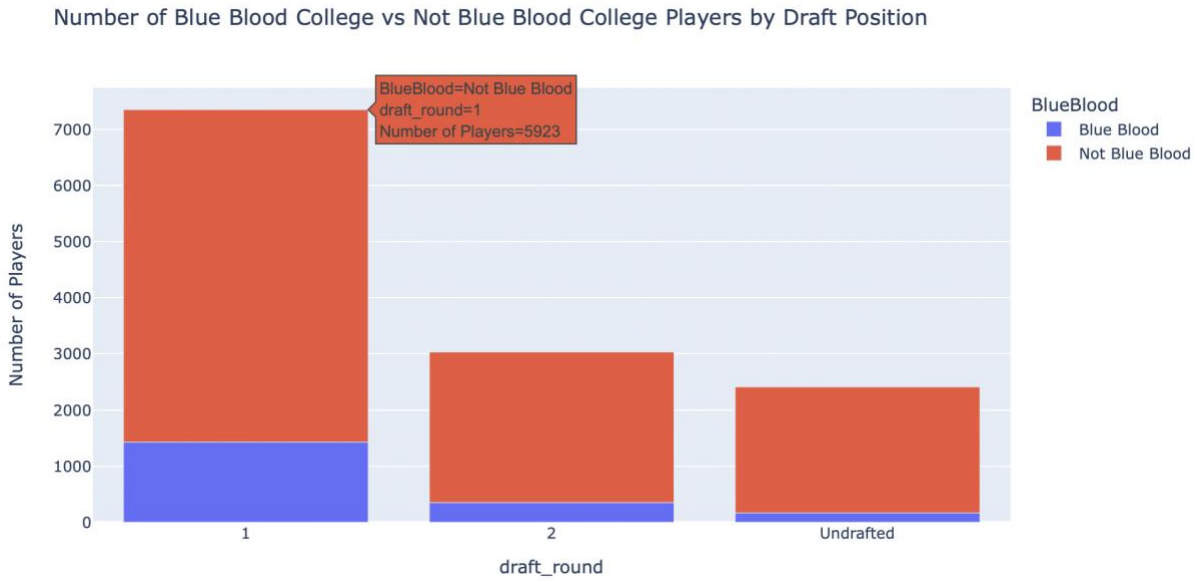

**Q6. Blue Blood College vs Not PPG by Season**

This line chart shows the average points per game for players that attended Blue Blood universities vs those that did not. Because the Blue Blood programs attract the more talented prospects, it is not surprising to see they average more ppg in the NBA than the players that did not attend those schools for every single season that this dataset has. This follows logic as usually Blue Blood schools have better coaches as well, so this could serve as incentive for front offices to err on the side of strong college team success when determining which player to draft. Additionally, it serves as further proof that high school players should seek out opportunities to play at Blue Blood schools to maximize their basketball career potential. The graph of the data is below:



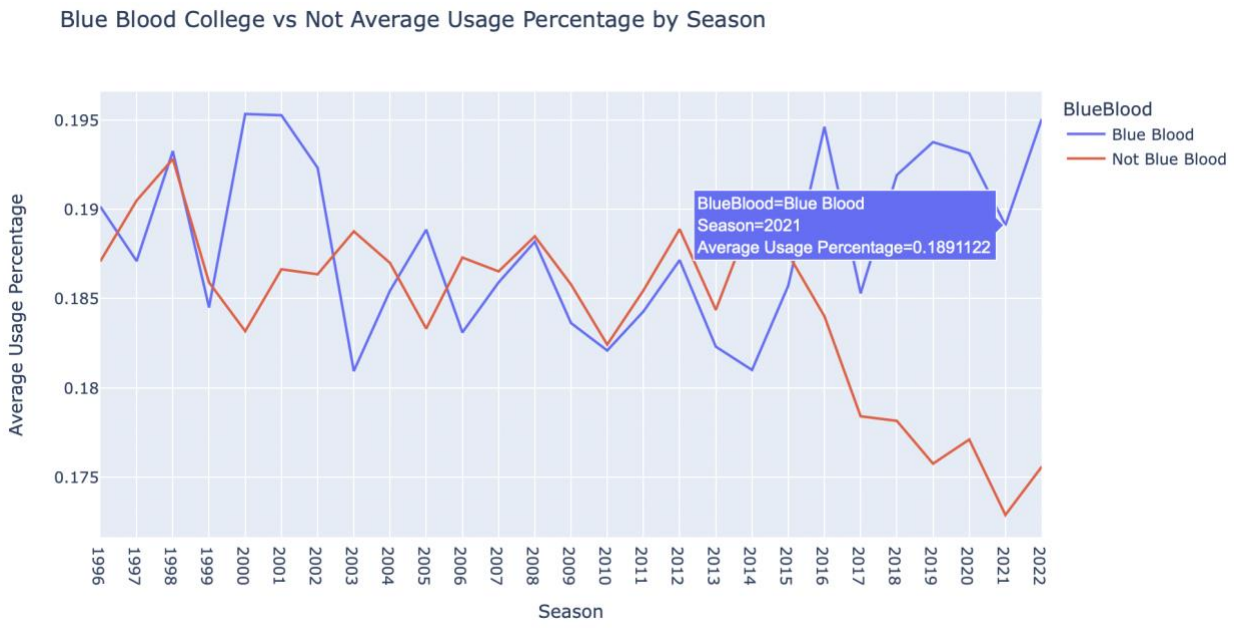

**Q7. Count of Blue Blood College vs Not Blue Blood College Players by Draft Position**

This bar chart shows the number of players for each draft position/round based on if they went to a Blue Blood university or not. As we can see, most of the Blue Blood school's players were drafted in the first round, as that is where the most talented prospects go. Additionally, it was very interesting to see how less than 10% of all undrafted players were from Blue Blood schools whereas roughly 20% of all first-round draft picks were from Blue Blood universities. As mentioned before, this goes to show the strong track record of those universities in player development and serves as further proof that high school players should seek out opportunities to play at Blue Blood schools to maximize their basketball career potential. The graph of the data is below:



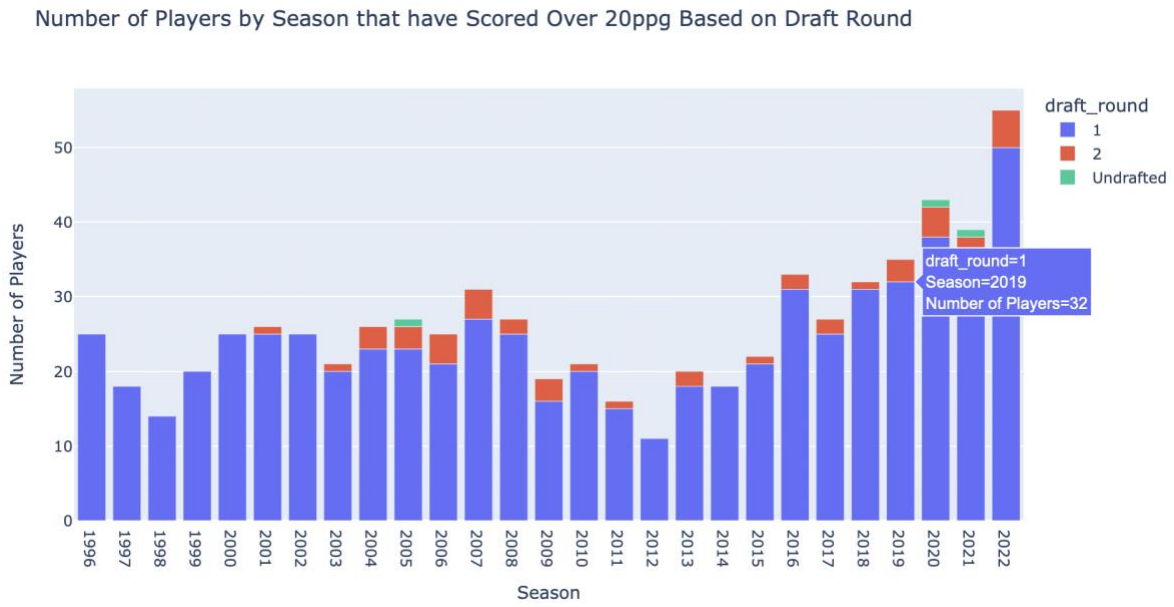
**Q8. Blue Blood College vs Not Average Usage by Season**

Our last analysis on player university attendance shows the success of players once they are in the league. We have already established that the Blue Blood university attending players are drafted higher and score more points, but usage rate is another important stat to look at to determine the dominance and success of a player, because usually the better the player the higher their usage rate. This is because coaches will want to maximize the amount of touches the best players get in order to maximize their chances to win. As we can see in the data, there is no clear separation in usage percentage for these groups until we get to 2016, where the usage rate of players that went to the Blue Blood schools clearly is higher than the other players that did not. We can see that trend has continued up until 2022, and also coincides with a sharp decline of usage rate for players that did not attend the elite colleges. This goes to suggest that especially with recent developments in NIL in college athletics, the schools that have the most resources, history, and pedigree have a much better chance of attracting, developing, and retaining top talent, which in turn results in better careers in the NBA. The graph of the data is below:



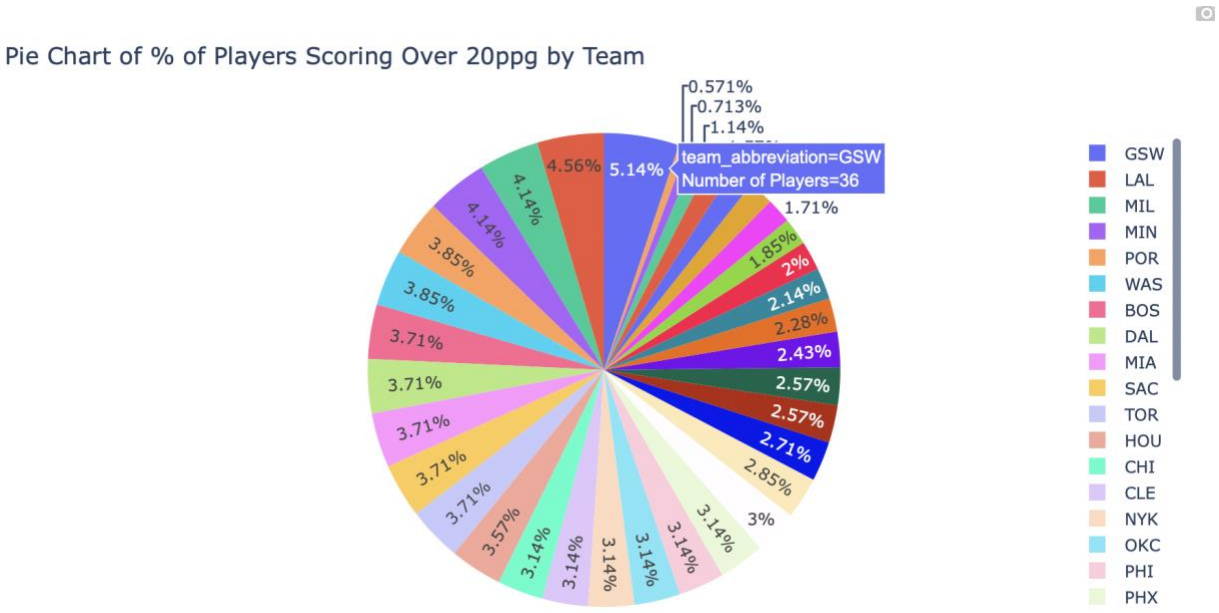
**Q9. Count players by Season that have scored over 20ppg based on draft round**

This graph counts the number of players by season that have averaged more than 20 points per game based on where they were drafted. As we can see, for the most part up until about 2016, most if not all of the players were first round draft picks. However, as we can see with the recent years, it has expanded to include more second round and undrafted players. This could be for a few reasons, firstly being the increased offense and scoring in the game as a whole. This could also be due to advancements in player development, overall better basketball talent as the game has become more global, and also more resources for players themselves to get better with the advancements in technology and rest/recovery science in the NBA. The graph of the data is below:



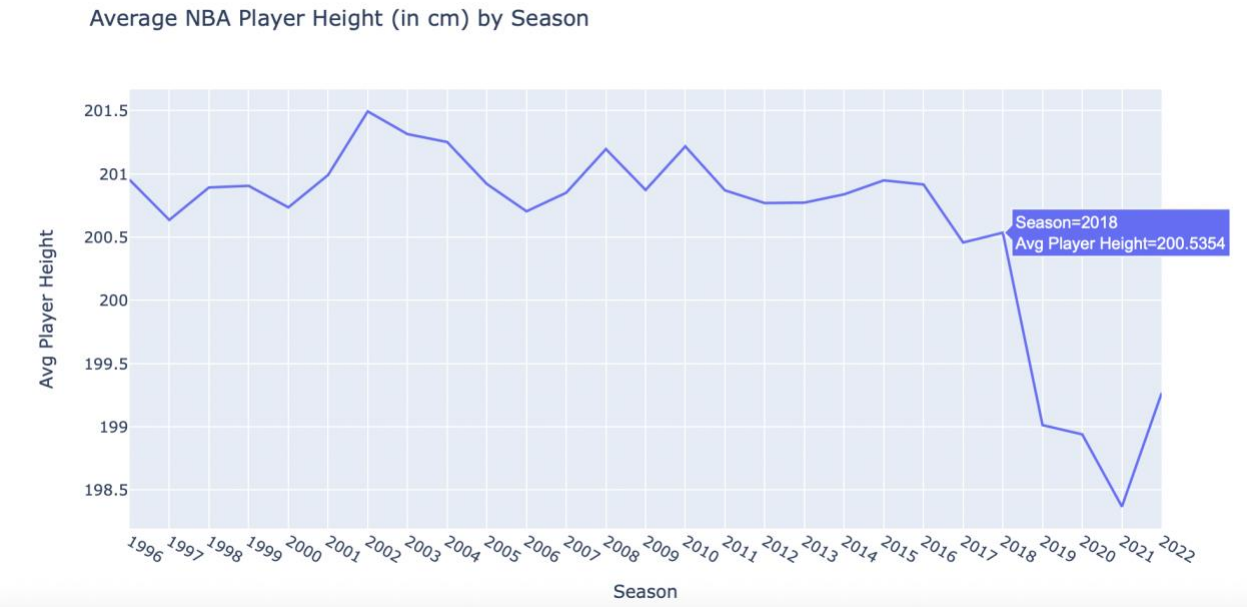
**Q10. Count players that Avg over 20ppg by Team**

This pie chart shows us the percentage of players that average over 20ppg based on the teams. This was done to see if certain successful teams have had more higher scorers than others. As we can see, the Golden State Warriors have the most players in the league that have averaged more than 20ppg, at roughly 5% of all players that have done so in the dataset. The top 5 teams with the most players that average over 20 ppg are GSW, LAL, MIL, MIN, POR, and based on team data from the time frame of this data set, does not show a correlation with more team success. A variety of factors could be the reason for this, first of them being the high scorers were not all on the same team at the same time. Additionally, just because a team has a lot of high scorers does not mean they will be successful, as strong bench teams are required to win championships as well as good defense. The graph of the data is below:




**Q11. NBA Player Height by Season**

The last graph for this analysis was just looking at average NBA player heights by season (measured in cm). As we can see, starting from about 2016 player heights have clearly fallen drastically. This coincides with the rise of “positionless” basketball and also the rise of the 3-point shot, as teams have shifted their focus away from traditional big men at power forward and center and prioritized more versatile players at those positions that may be smaller but can shoot better from 3 and can handle the basketball. This ties into what we previously discussed with the rise in scoring across the NBA over the past few years, as now teams are smaller but can shoot more often and more effectively which increases offensive output. Though this comes with a tradeoff (poor paint defense), it will be interesting to see how this trend continues, as we are already seeing 2022 average heights climb back up for the biggest year over year increase since 2002. The graph of the data is below:


Average NBA Player Height (in cm) by Season
Avg Player Height
201.5
201
200.5
200
199.5
199
198.5
Season=2018
Avg Player Height=200.5354
1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015 2016 2017 2018 2019 2020 2021 2022
Season