NBA Players Playoffs Statistics Correlations EDA

By

Dhaval Patel

May 27, 2021

MSDS 430: Python for Data Science

Introduction

  I have been watching NBA Basketball games since childhood which led to my interest of choosing the overall NBA Players Statistics data. Based on my intuition over the past couple of years watching NBA I see a trending pattern that NBA players who are scoring more in 3-pointers offensive categories are more likely to lead total points and be in teams that make NBA playoffs. Also, is shooting 3 pointers an optimal solution over 2 pointers to make it to playoffs. This report explains these assumptions made and analyzes through data visualization process on 5 years of NBA players statistics data from 2015-2020. The data was collected from basketball-reference.com which is publicly available site and population represents overall NBA Player statistics in offensive shooting category. The categories used for analysis include player's position, team, games played, minutes played, 3 pointers, 2 pointers, total points, and fields goals per game.

The following are **three hypotheses** that are analyzed using data visualization process:

1. To analyze correlations between each shooting category so that it helps to explore which NBA players position are more likely to score points in 3 pointers category than 2 pointers per game vice-versa.

2. It is widely accepted among NBA fans and based on intuition that NBA players who have played most games and minutes have better chances of making playoffs.

3. NBA players who are the top 10 leading scorers in the average three and two points per game shooting category have more chances of being on a team that makes to the playoffs.

Data Preparation and Analysis:

The sample size of data is 3196 rows and 37 columns. Since the data does not contain playoffs information, I created a reference data of teams that reached playoffs each year from 2015-2020 to merge with current dataset. The first step in data preparation is adding year column to all the corresponding NBA seasons dataframe. Second step is to concatenate all the NBA seasons dataframe into one large dataset assigning to NBA players stats dataframe. Third step is merging NBA player stats dataframe with NBA playoffs reference dataframe which includes Year, Team, and Playoffs column into one large overall NBA playoffs stats dataset. In the final step, we did an outer join on 'Year' and 'Team' column which matches up based on the Year, Team and Playoffs column that represents with 'Y' who made the playoffs. Also, we are making the corresponding NaNs values of playoff equals to 'N' which means that these NBA players and teams did not make the playoffs in that year. We created 4 extra reference data columns such as 'MinutesPerGame', 'FieldGoalsPerGame,' '3PointerPerGame', and '2PointerPerGame'. For example, calculating minutes played per season divided by games played to get 'MinutesPerGame' columns similar calculation approach was taken for other columns. We have removed all irrelevant columns, and renamed all the remaining columns for easier interpretation. All entries with missing or nulls values were replaced with an integer value of '0'. All rows and columns with significant outliers were dropped, specifically NBA players age column with upper extreme for the age was 38 but beyond that are outliers based on data and were dropped from the table.

Table 1. Provides details on data dictionary used throughout this exploratory data analysis.

Table 1. *Data Dictionary for Variables of NBA Offensive Statistics*

Data Dictionary		
Variable Name	Data Type	Description
Player	String	Player corresponds to name of the NBA players.
Position	String	Position of the NBA players while playing the game.
Team	String	Team corresponds to team of the NBA players.
Age	Integer	Age corresponds to age of the NBA players.
GamesPlayed	Integer	The number of games played by NBA player in a season.
MinutesPlayed	Integer	The number of minutes played by NBA player in a season.
FieldGoals	Integer	The number of field goals that NBA players have made in a season. This includes both 2 pointers and 3 pointers.
3Pointers	Integer	The number of 3 pointers that NBA players have made in that season.
2Pointers	Integer	The number of 2 pointers that NBA players have made in that season.
TotalPoints	Integer	The number of total points scored by an NBA player in that season.
FieldGoalsPerGame	Float	The number of field goals that NBA players have made per game.
3PointerPerGame	Float	The number of 3 pointers that NBA players have made per game.
2PointerPerGame	Float	The number of 2 pointers that NBA players have made per game.
TotalPointsPerGame	Float	The number of total points that NBA players have made per game.
MinutesPerGame	Float	The number of minutes that NBA players have played per game.
Year	String	The year is reference data column to keep track of NBA players statistics from each season.
Playoff	String	The playoff is reference data column to keep track of which NBA players were in team made playoffs in corresponding seasons.

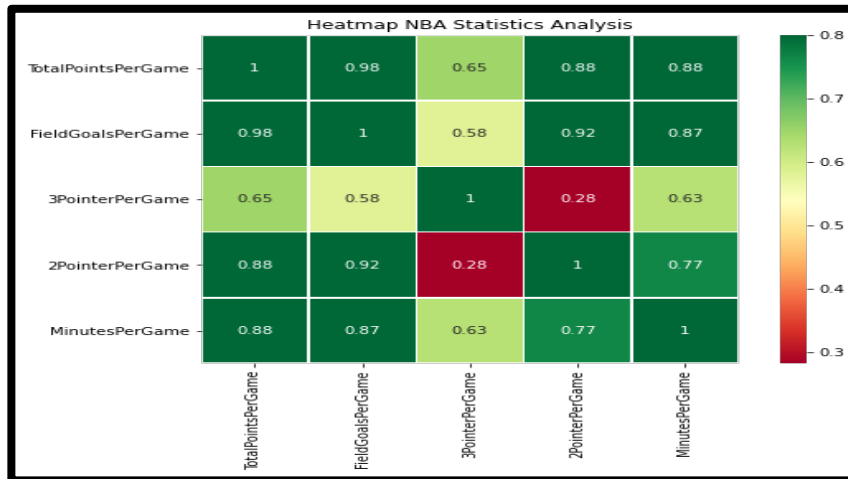**Hypothesis 1: Heatmap Analysis**

Figure 1: Heatmap Correlations of NBA Players Shooting Category Per Game

Based on correlation matrix the dark green color shows high positive correlation between Total points per game and three other columns such as field goals, 2 pointers, and minutes played by an NBA player per game. This correlation is supported based on our analysis from hypothesis 1 because when we observe an NBA overall Total points per game it depends upon these columns in terms the number of minutes played by a player, fields goals attempts in 2 pointer or 3 pointer shooting scores category by the player based on that the Most Valuable Player of the Game is awarded. Similarly, light green color show slight positive correlation as the amount of 3 pointers scored by an NBA player might be less based on player's position. The three pointer and two pointer per games shows strong negative correlation in red color as they both are two different shooting category and not dependent on each other.

Swarm Plot and Pie Chart Analysis: The total 3 pointers NBA players have made per seasons and per game from years 2015-2020 is categorized by NBA players' position. The following are the NBA players position for better interpretation of data analysis: Power Forwards (PF), Shooting Guards(SG), Small Forwards(SF), Center(C), and Point Guards(PG).
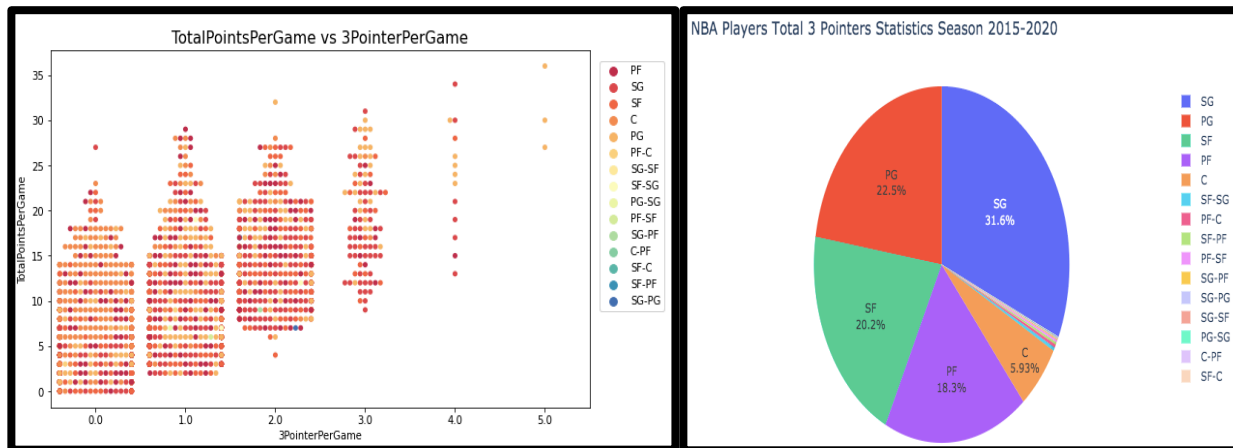
Figure 2: TotalPointsPerGame vs 3PointerPerGame Categorized By NBA Players Positions
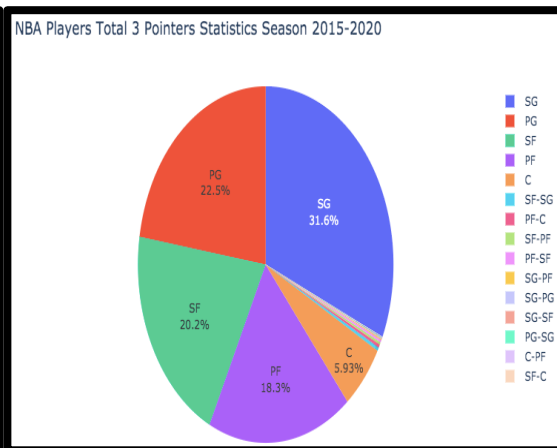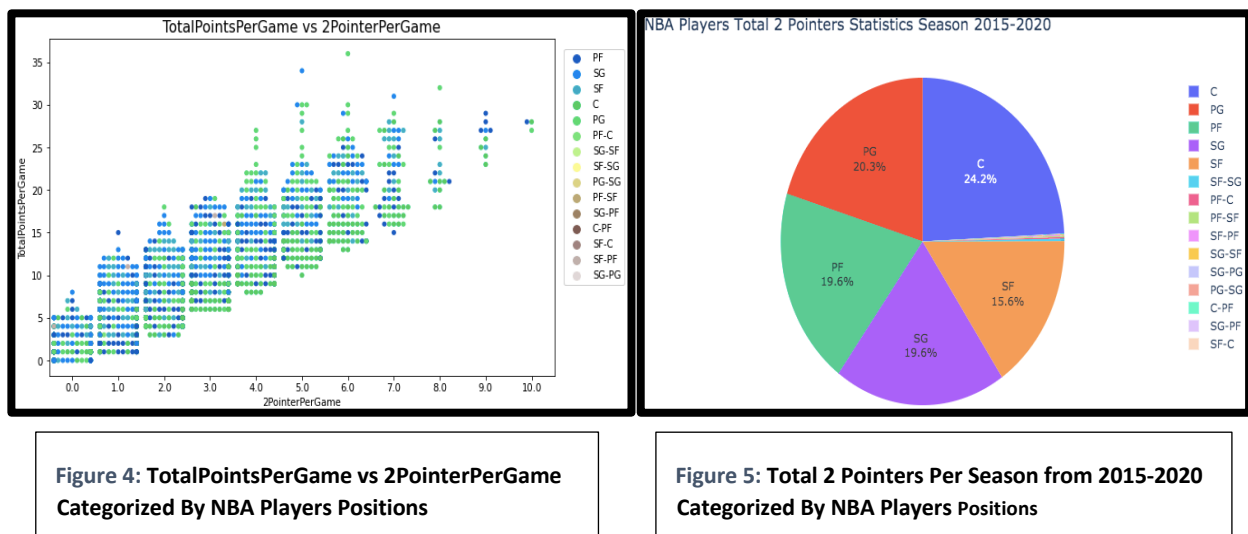

Figure 3: Total 3 Pointers Per Season from 2015-2020 Categorized By NBA Players

Based on the swarm plot in **Figure 2** the red dots in the data corresponds NBA Shooting Guard (SG) position it shows that players in this position have a high average scoring range around 2 to 4 three pointer per game which leads them to having an average of 10 to 25 total points per games. The light orange dots in the data corresponds to NBA Point Guard (PG) position which indicates that players in this position have a high average scoring range around 3 to 5 three pointer per game which leads them to having an average of 10 to 35 total points per games shooting three pointers compare to other positions.

To further confirm my hunch either correct or incorrect that the overall data will reveal positive correlation between NBA players in Point Guard (PG) and Shooting Guard (SG) position are likely to score an average of higher three pointers compared to other position. As we can see in pie chart **Figure 3** that 31.6% of the 3 pointers are scored by players in the Shooting Guard (SG) position from our historical data. Similarly, 22.5% of 3 pointers are scored by NBA players in Point Guard (PG) position it clearly shows that Shooting Guard position have always scored more three pointers per season than any other position in NBA. The data indicates that our hunch was correct, and supported by hypothesis 1 based on swarm plot and pie chart. It is surprising to

see there are is not much difference in between Small Forwards, Power Forwards, and Point Guards positions shooting three pointer categories. In this case players might try to learn how to improve in both category over the years which led to us seeing closer percentages. It interesting to note that players who play multiple positions do not make significant amount in both plots.

Swarm Plot and Pie Chart Analysis Part 2: The total 2 pointers NBA players have made per seasons and per game from years 2015-2020 is categorized by NBA players' position.

Figure 4: TotalPointsPerGame vs 2PointerPerGame Categorized By NBA Players Positions

Figure 5: Total 2 Pointers Per Season from 2015-2020 Categorized By NBA Players Positions

Based on the swarm plot in **Figure 4** shows the green and light green dots in the data corresponds to NBA Center (C) and Point Guard (PG) position which indicates that players in this position have a high average scoring range around 2 to 7 two pointer per game which leads them to having an average of 5 to 30 total points per games shooting two pointers. The blue dots in the data corresponds NBA Power Forward (PF) position it shows that players in this position have a high average scoring range around 1 to 6 two pointer per game which leads them to having an average of 5 to 20 total points per games shooting two pointers. To further confirm my

hunch either correct or incorrect that the overall data will reveal positive correlation between NBA players in Power Forward (PF), Small Forward (SF), and Center (C) position are likely to score an average of higher two pointers compared to other position. As we can see in pie chart **Figure 5** that 24.2% of the 2 pointers are scored by players in the Center position from our historical data. Similarly, 20.3% of 2 pointers are scored by NBA players in Point Guard (PG) position it clearly shows that Center position have always scored more two pointers per season than any other position in NBA.

The data indicates that our hunch was correct, was partially supported by hypothesis 1 based on swarm plot and pie chart for two of the NBA position such as Center and Point Guard. It is surprising to see there is tie between Power Forwards and Shooting Guard position in the shooting two pointer category. After analyzing all the four plots the data shows players position had light positive correlation with shooting category.

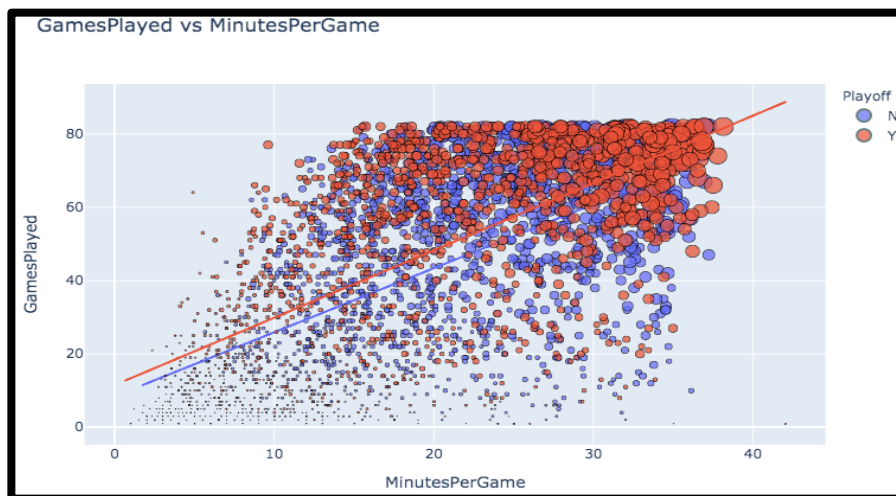**Hypothesis 2: Scatter Plot Analysis**

Figure 6: GamesPlayed vs MinutesPerGame Categorized By NBA Players Made Playoff or Not

Based on the assumptions made in hypothesis 2 how games and minutes played per game affects players chances of making playoffs or not. As we look at the trendline in **Figure 6** of the scatter plot we see that a lot of red dots in the upper right quadrant of the plot where NBA players who have played around range of 80 games and have an average of greater than 30 minutes per game have more chances of making playoffs, and scoring more points based on the historical data. On the other hand, trendline for the blue dots starts at the lower left quadrant to mid upper right quadrant of the plot where NBA who have played around range of 20 - 65 games and have an average of less than 30 minutes per game have less chances of making playoffs, and scoring points based on the historical data. But there are some outliers in the scatter plot for instance there are blue dots in the upper right quadrant of the plot that have met all criteria but still not make playoffs. So, based on our analysis we can conclude that there is a light position correlation between GamesPlayed and MinutesPerGame would affect players chances of making playoffs.

Hypothesis 3:

					3PointerPerGame_Mean
Player	Team	Position	Playoff	Year	
Stephen Curry	GSW	PG	Y	2018-2019	5.0
James Harden	HOU	PG	Y	2018-2019	5.0
Stephen Curry	GSW	PG	Y	2015-2016	5.0
Dāvis Bertāns	WAS	PF	N	2019-2020	4.0
Duncan Robinson	MIA	SG	Y	2019-2020	4.0
R.J. Hunter	BOS	SG	Y	2018-2019	4.0
Buddy Hield	SAC	SG	N	2019-2020	4.0
Damian Lillard	POR	PG	Y	2019-2020	4.0
D'Angelo Russell	TOT	PG	N	2019-2020	4.0
	GSW	PG	N	2019-2020	4.0

					2PointerPerGame_Mean
Player	Team	Position	Playoff	Year	
Anthony Davis	NOP	PF	Y	2017-2018	10.0
		C	N	2016-2017	10.0
Russell Westbrook	HOU	PG	Y	2019-2020	10.0
Anthony Davis	NOP	C	N	2018-2019	9.0
LaMarcus Aldridge	SAS	C	Y	2017-2018	9.0
Giannis Antetokounmpo	MIL	PF	Y	2017-2018	9.0
Karl-Anthony Towns	MIN	C	N	2016-2017	9.0
Giannis Antetokounmpo	MIL	PF	Y	2018-2019	9.0
Anthony Davis	NOP	C	N	2015-2016	9.0
DeMar DeRozan	TOR	SG	Y	2016-2017	9.0

Figure 7: Top 10 NBA Player Statistics based on Average 3 Pointer Per Game.

Figure 8: Top 10 NBA Player Statistics based on Average 2 Pointer Per Game.

Based on the **Figure 7** show the data with top 10 NBA players having a high average of 3 pointer per game we can see that some of the players have been consistently performing well in their individual statistics shooting like Steph Curry leading in points in the year 2018 and 2015, also being in teams that playoffs. Similarly, 'James Harden' have tied shooting an average of 5 three pointers per game. On the other hand, there are players for instance D'Angelo Russell who has been consistently scoring 4 three pointer per game in the year '2019' for two different teams but did not make the playoff.

When we examine the data in **Figure 8** to see which top 10 NBA players having a high average of 2 pointer per game, we can see that some of the players have been consistently performing well in their individual statistics shooting like Anthony Davis leading in points in the year 2017 and 2016, but only made playoffs once. At the same time Anthony Davis have scored an average of 9 two per games in the year 2018 and 2015 but did not make playoffs with same team.

The data indicates that overall team's performance plays a vital role in making playoffs then just individual players statistics. In addition, the data reveals that 4 out of 10 players in the leading three and two pointer shooting category were not in the teams that made the playoffs which is 40% of the data above. It further proves that based on the assumptions made in hypothesis 3 individual players statistics shows neutral correlation with shooting category to their chances of making the playoffs or being in the teams that makes the playoffs.

Conclusion:

Based on the visualization data revealed that scoring a high average in shooting category does not necessarily guarantee a spot in the playoffs there are other categories in the statistics

that also play vital role to achieve that goal. For instance, hypothesis 1 was partially supported as shown in the swarm plot and pie chart that some position like Shooting Guard (SG) and Center (C) have scored higher three and two pointer per game which was correct based on our hunch. But the other NBA players positions have been fluctuating because players might be trying to learn how to improve in both three and two pointer shooting category over the years which led to us seeing tie and close percentages

The hypothesis 2 was partially supported NBA players who have played around range of 80 games and have an average of greater than 30 minutes per game have more chances of making playoffs, and scoring more points. The overall analysis shows light position correlation between GamesPlayed and MinutesPerGame column as they have been categorized based on playoffs. Assumptions made in hypothesis 3 were partially supported as well because the data revealed that 4 out of 10 players in leading average three and two pointer shooting category were not in the teams that made the playoffs which is 40% of the data. It further proves that based on the individual players statistics shows neutral correlation with shooting category to their chances of making the playoffs because its dependent on the overall team's statistics data.

I did not encounter difficulties with the data, but it would be interesting to gather overall team's statistics data from 2015 to 2020 and merge with current dataset of player's playoffs statistics data so that we can do analysis on other various categories which would lead to seeing better correlations between team and players columns. Also, how it affects their chances of making playoffs that would possibly improve the outcomes of the analysis. To take it a step further I would look into analyzing the dynamics of each NBA players positions as well because the players are improving in both three and two pointer shooting category over the years.