## Background

This dataset collected from Kaggle tries to answer if a rookie basketball player will have a successful career or not by looking at their entire career’s statistics. I hypothesize that we should look at the minutes played instead of the games played if we want to evaluate a player better. Meaning that the rookie players do not get to spend much time on the court because the more experienced players are given more chances and hence they get the glory. Using this dataset I will try to see if the number of minutes played has something to do with the success or not.

From the article *From College to the NBA: What Determines a Player’s Success and What Characteristics are NBA Franchises Overlooking?* By Brent Evans, we see that there are a lot of factors often missed by the NBA communities and executives while drafting a player. In his article, he not only explored what factors are taken into consideration while drafting a player but also focused essentially on the determinants of player productivity and game time 3 years into a player’s NBA career [2].

Giving young players a chance to perform can significantly boost their confidence and can prepare them for future critical matches. When the new players are not able to perform for a very long time, their confidence will get low and they will be more likely to make mistakes during the crucial moment of the game. According to the authors of *Human Performance after Success and Failure: Evidence from the NBA*, people should be more careful in relying on team members who experienced a crucial success in their previous task [3], as they are likely to fail the next one, having immense pressure on them. This further adds to my point on why new players should be given a chance to perform.

People often refer to how experienced NBA players have their own momentum, and they own the game because they understand it much better. The authors of *Searching for Momentum in NBA Triplets of Free Throws* explore whether momentum exists by looking at cases in which a basketball player has three consecutive free throws [5]. And conclude that no such thing exists in basketball.

Authors of *Key Game Indicator in NBA Players’ Performance Profiles* discussed how older guards and forwards of age 26 or more, played more minutes per game compared with younger players in the same playing position, and professional basketball players aged over 30 years showed lower values in distance covered [4]. Which further proves that rookies aren’t given that many chances. Let us dive deeper into our analysis in the following sections.

## Data description

The attribute information of the dataset are stated below:

|  |  |
| --- | --- |
| Attribute | Meaning |
| 1. name | Name of Player |
| 2. gp | Games Played |
| 3. min | Minutes Played |
| 4. pts | Points Made |
| 5. fgm | Field Goals Made |
| 6. fga | Field Goals Attempts |
| 7. fg | Field Goals % |
| 8. 3p\_made | 3 Pointer Made |
| 9. 3pa | 3 Pointer Attempted |
| 10. 3p | 3 Pointer % |
| 11. ftm | Field Throw Made |
| 12. fta | Field Throw Attempt |
| 13. ft | Field Throw |
| 14. oreb | Offensive Rebound |
| 15. dreb | Defensive Rebound |
| 16. reb | Rebound |
| 17. ast | Assists |
| 18. stl | Steals |
| 19. blk | Blocks |
| 20. tov | Turnovers |
| 21. Target\_5yrs | Career Duration more than 5 years? |

The dataset was downloaded from Kaggle which was originally taken from data.world website. It consists of information of 1339 basketball players, and a total of twenty-one variables that decide if a player will have a successful career at the NBA 5 years later. However, I think there are only a few factors that affect that, and not all twenty-one variables are needed. We will see in the results section how Steals and Blocks have nothing to do with how many games were played but rather how many minutes played.

Even though the author of this dataset made this to verify a player’s future, I am going to use it to prove that minutes played matters more than the number of games played by a basketball player. To do so, I will find the correlation of the following variables.  
\* Field goals compared to games played and minutes played.  
\* Field goals attempt compared to games played and minutes played.  
\* Steals compared to games played and minutes played.  
\* Blocks compared to games played and minutes played.  
Therefore, it brings us to answer the following questions:  
\* Is the number of minutes played have anything to do with how many field goals have been scored?  
\* Is the number of games played have anything to do with how many field goals were attempted?  
\* Does the capability of blocking become better with more minutes played?  
\* Does the capability of securing more steals become easier with more minutes played?

If all these are successfully answered, then we can safely conclude that playing more minutes on the court makes one a better player.

## Method

To answer the questions stated above, we have decided to use various data sorting techniques using R. To evaluate the dataset, even more, we have decided to use several graphical methods to answer our questions above. The reason for choosing a scatter plot for almost all the variables is because it gives us a better insight into a dataset that has so many variables (21 to be precise!). Scatterplot helps us determine which variables are positively or negatively correlated just by looking at the diagram. It shows all data points, including minimum and maximum and outliers. It also retains the exact data values and sample size. As well as show both positive and negative types of graphical correlation.

I demonstrated how the results can be looked at from a different perspective if we compare minutes played with the results instead of the games that appeared. To do that, I first plotted a graph of games that appeared vs minutes played. Then, I changed the variables to points made, field goals made, field goals attempted, steals, blocks, while keeping the x-axis constant for either minute played or games appeared. Doing so the following results showed which are discussed in the results section. The following chunk contains the code used for this report.

#installing the necessary libraries  
library(ggplot2) #for the graph plots   
library(knitr) #for creating tables  
  
  
## Reading the csv file of the data set nba-players to  
df <- read.csv("nba-players.csv") ## Assigning the data set to df  
  
  
## Mean of minutes played by all player  
m1 <- mean(df$min)  
cat("Mean of Minutes played is", m1)  
  
## Mean of games played  
m2 <- mean(df$gp)  
cat("Mean of Games played is", m2)  
  
## Mean of field goals made  
m3 <- mean(df$fgm)   
cat("Mean of Field Goals Made is", m3)  
  
  
## Mean of field goals attempted  
m3 <- mean(df$fga)   
cat("Mean of Field Goals attemped is", m3)  
  
  
## Variance of games played  
v1 <- var(df$gp)  
cat("Variance of Games Played is", v1)  
  
  
## Variance of minutes played  
v2 <- var(df$min)  
cat("Variance of Minutes Played is", v2)  
  
  
## Variance of field goals made  
v3 <- var(df$fgm)  
cat("Variance of Field Goals Made is", v3)  
  
## Variance of field goals attempted  
v4 <- var(df$fga)  
cat("Variance of Field Goals Attempted is", v4)  
  
  
  
## Plotting histogram of Games Played  
hist(df$gp, main="Histrogram of Games Played", xlab = "Games Played", ylab = "Number of Players")  
  
## Plotting histogram of Minutes Played  
hist(df$min, main="Histrogram of Minutes Played", xlab = "Minutes Played", ylab = "Number of Players")  
  
## Plotting histogram of Field Goals Made  
hist(df$fgm, main="Histrogram of Field Goals Made", xlab = "Filed Goals", ylab = "Number of Players")  
  
  
## Plotting the Scatter diagram of Games Played vs Minutes Played  
ggplot(df, aes(gp, min)) + geom\_point() +  
labs(title = "Games played vs Minutes played") +  
xlab("Games Played") +  
ylab("Minutes Played")  
  
## Plotting the Scatter diagram of Minutes Played vs Points Earned  
ggplot(df, aes(min, pts)) + geom\_point() +  
labs(title = "Minutes Played vs Points earned") +  
xlab("Minutes Played") +  
ylab("Points Earned")  
  
## Plotting the Scatter diagram of Games Played vs Points Earned using ggplot  
ggplot(df, aes(gp, pts)) + geom\_point() +  
labs(title = "Games Played vs Points Earned") +  
xlab("Games Played") +  
ylab("Points Earned")  
  
## Plotting the Scatter diagram of Minutes Played vs Field Goals Made using ggplot  
ggplot(df, aes(min, fgm)) + geom\_point() +  
labs(title = "Minutes Played vs Field Goals Made") +  
xlab("Minutes Played") +  
ylab("Field Goals Made")  
  
## Plotting the Scatter diagram of Games Played vs Field Goals Made using ggplot  
ggplot(df, aes(gp, fgm)) + geom\_point()+  
labs(title = "Games played vs Field Goals Made") +  
xlab("Games Played") +  
ylab("Field Goals Made")  
  
## Plotting the Scatter diagram of Minutes Played vs Field Goals Attempted using ggplot  
ggplot(df, aes(min, fga)) + geom\_point()+  
labs(title = "Minutes Played vs Field Goals Attempted") +  
xlab("Minutes Played") +  
ylab("Field Goals Attempted")  
  
## Plotting the Scatter diagram of Games Played vs Field Goals Attempted using ggplot  
ggplot(df, aes(gp, fga)) + geom\_point() +  
labs(title = "Games played vs Field Goals Attempted") +  
xlab("Games Played") +  
ylab("Field Goals Attempted")  
  
## Plotting the Scatter diagram of Games Played vs Free Throw using ggplot  
ggplot(df, aes(gp, ft)) + geom\_point() +  
labs(title = "Games played vs Free Throw") +  
xlab("Games Played") +  
ylab("Free Throw")  
  
## Plotting the Scatter diagram of Minutes Played vs Free Throw using ggplot  
ggplot(df, aes(min, ft)) + geom\_point() +  
labs(title = "Minutes played vs Free Throw") +  
xlab("Minutes Played") +  
ylab("Free Throw")  
  
## Plotting the Scatter diagram of Games Played vs Block using ggplot  
ggplot(df, aes(gp, blk)) + geom\_point() +  
labs(title = "Games Played vs Block") +  
xlab("Games Played") +  
ylab("Blocks")  
  
## Plotting the Scatter diagram of Minutes Played vs Block using ggplot  
ggplot(df, aes(min, blk)) + geom\_point() +  
labs(title = "Minutes Played vs Blocks") +  
xlab("Minutes Played") +  
ylab("Blocks")  
  
## Plotting the Scatter diagram of Games Played vs Steals using ggplot  
ggplot(df, aes(gp, stl)) + geom\_point() +  
labs(title = "Games Played vs Steals") +  
xlab("Games Played") +  
ylab("Steals")  
  
## Plotting the Scatter diagram of Minutes Played vs Steals using ggplot  
ggplot(df, aes(min, stl)) + geom\_point() +  
labs(title = "Minutes Played vs Steals") +  
xlab("Minutes Played") +  
ylab("Steals")

## Results

In this section, we can see the graphical representation of the methods we applied above. We can see how the two parameters compare in each case stated above. Plotting a scatter plot diagram helps us see all the points in the dataset. Most importantly, it also helps us figure out if any outlier exists.

The results show to some extent that more minutes on the field does ensure a better scoring chance, and thus it is a better comparison than games played.

The mean of minutes played, games played, field goals attempted, field goals made are stated below.

## Mean of Minutes played is 17.62463

## Mean of Games played is 60.41418

## Mean of Field Goals Made is 2.629104

## Mean of Field Goals attemped is 5.885299

The Variance of minutes played, games played, field goals attempted, field goals made are stated below.

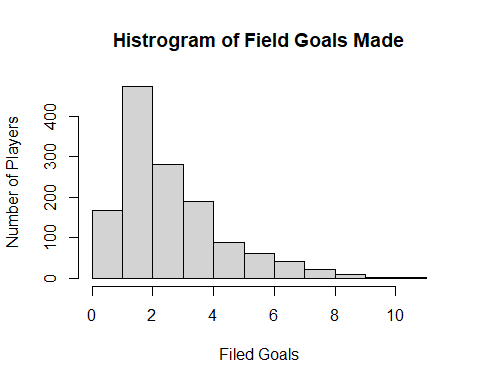
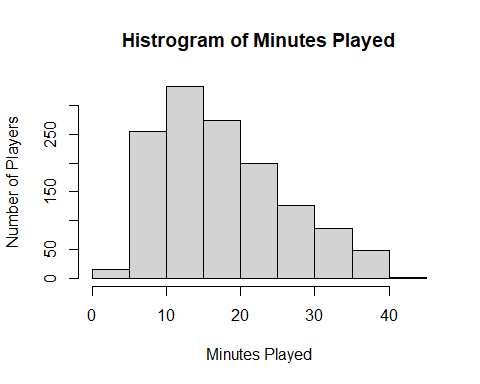
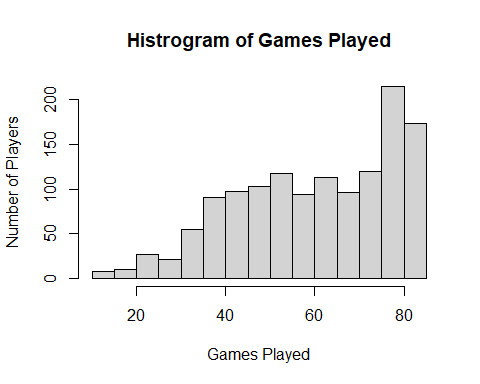
## Variance of Games Played is 303.9441

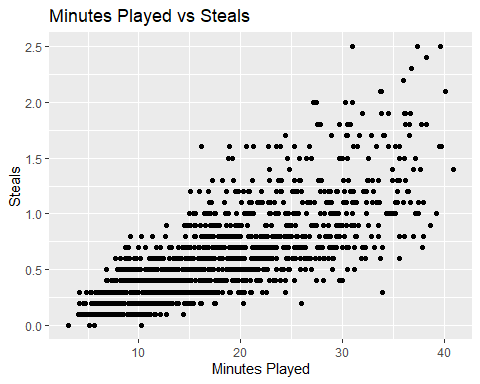
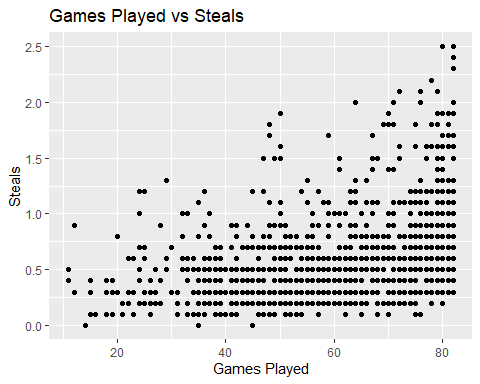
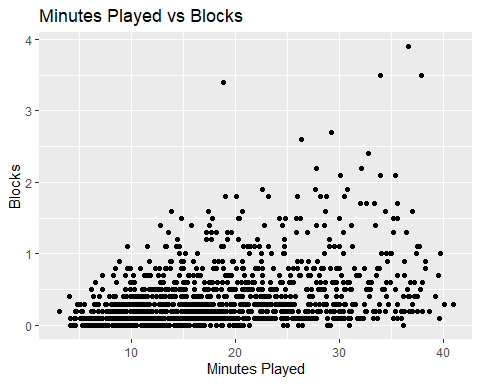
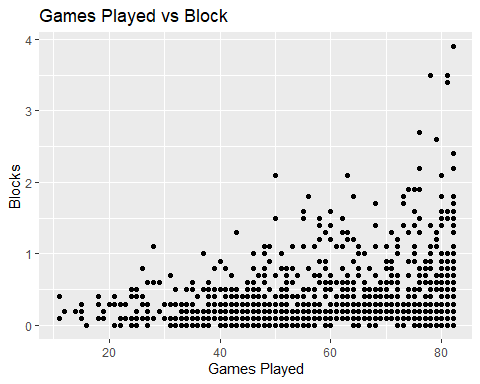
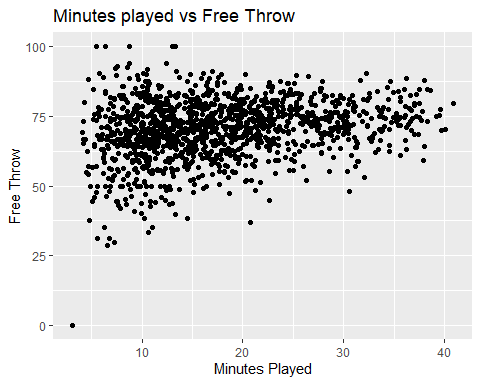
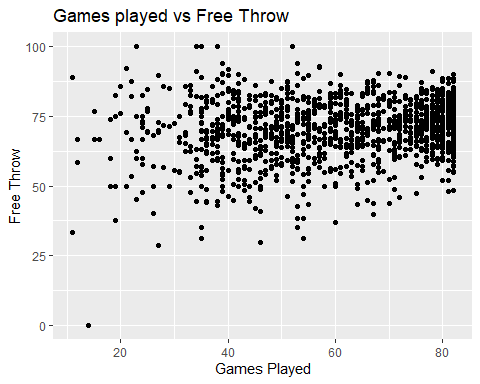
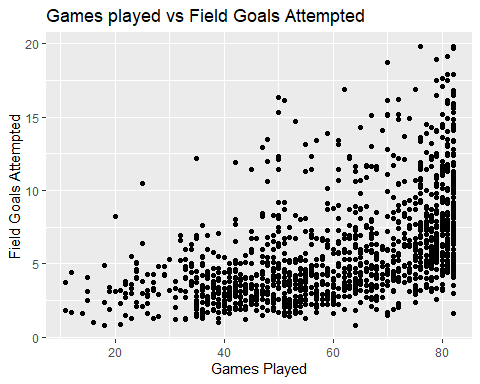
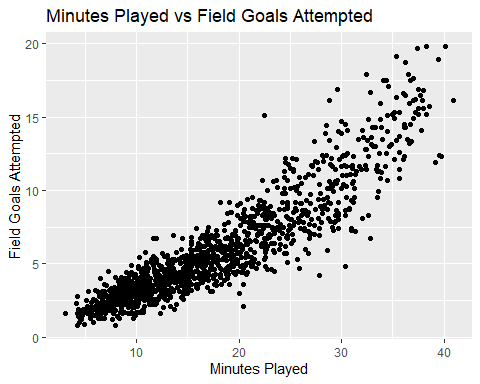
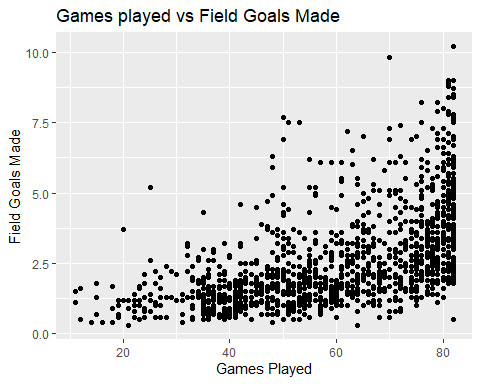
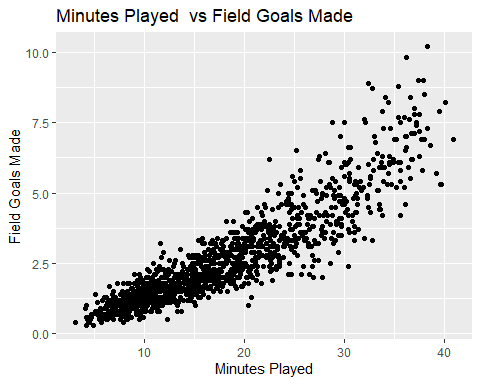
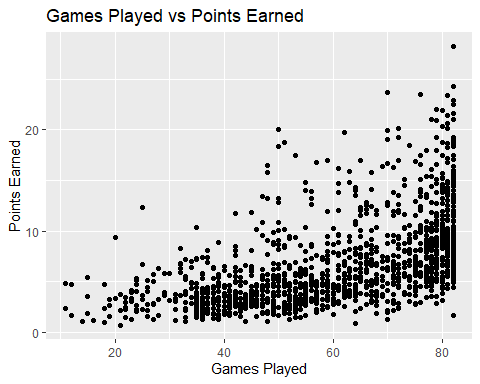
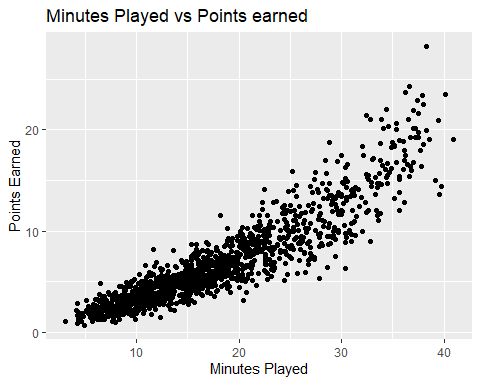
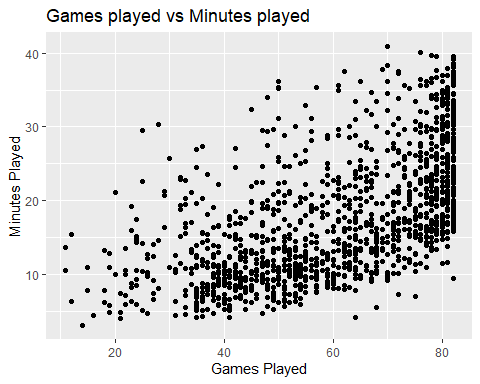
## Variance of Minutes Played is 69.02226

## Variance of Field Goals Made is 2.834358

## Variance of Field Goals Attempted is 12.91316

I used graphs and mainly scatter plots to understand the dataset better. The following are some histograms plotted to better understand the dataset.



The scatter diagrams are given below: 

## Conclusion

From the above dataset and discussion, it can be concluded that more minutes spent playing can significantly boost a player’s confidence [3]. This especially benefits newbie players as they need more time and experience to gain the same level of expertise as the professional, senior players.

Spending more minutes playing helps a player grasp the situation better and can prepare them for future critical situations. It is not always wise to rely on players who did well in the past [3] as they too can crack under pressure and there is statistically no such thing as momentum in basketball, as it is not a game where such scenarios can be benefited from or even created. It can be concluded that, basketball matches in general are not a suitable environment for activation of momentum that might have applied to other games such as boxing, etc.[5].

What further supports my analysis is that players who are over 30 years old always do not perform as well as they should despite being so experienced [4] but the players in the mid-age range perform better than the younger ones only because they are at the peak of their career and get more chances to perform.

Therefore, while drafting players, NBA should look into other factors and not just rely on previously experienced players [2]. Such as who was in the high school team and how many trophies they had previously.

A lot of the time we fail to acknowledge the true potential of an NBA player because we haven’t seen enough of them. *Evans*, also showed in his paper how older players outperform younger players in the NBA. This I believe is mainly because the young ones do not get enough chance to prove themselves and are mostly benched.

As we have seen from the results above, giving players a chance to play on the court makes them more experienced and helps increase their overall confidence and understanding of the game. Therefore, minutes played should be considered as a determining factor of a players’ input. From the above results, we have also seen that steals has nothing to do with the minutes played and therefore it should not be used as a determining factor if a player is qualified enough to play even 5 years later or not.

## Reference

[1] “NBA Players Career Duration.” *Kaggle*, 22 Sept. 2021, www.kaggle.com/sveneschlbeck/nba-players-career-duration.  
[2] Evans, Brent A. “From College to the NBA: What Determines a Player’s Success and What Characteristics Are NBA Franchises Overlooking?” Applied Economics Letters, vol. 25, no. 5, 2017, pp. 300–04. Crossref, <doi:10.1080/13504851.2017.1319551>.  
[3] Bühren, Christoph, and Stefan Krabel. “Human Performance after Success and Failure: Evidence from the NBA.” The International Journal of Human Resource Management, vol. 32, no. 16, 2019, pp. 3402–27. Crossref, <doi:10.1080/09585192.2019.1634121>.  
[4] Dehesa, Rubén, et al. “Key Game Indicators in NBA Players’ Performance Profiles.” Kinesiology, vol. 51, no. 1, 2019, pp. 92–101. Crossref, <doi:10.26582/k.51.1.9>.  
[5] Morgulev, Elia, et al. “Searching for Momentum in NBA Triplets of Free Throws.” Journal of Sports Sciences, vol. 38, no. 4, 2019, pp. 390–98. Crossref, <doi:10.1080/02640414.2019.1702776>.