Introduction

Basketball is a team sport played on a rectangular court with two hoops or baskets, one at each end of the court. The objective of the game is to shoot a ball through the opposing team's basket while preventing the other team from scoring in your own basket. Five players make up each team, and they are free to pass, dribble, or shoot the ball wherever on the floor. You can score points by putting the ball through the goal of the opposite team. A team can receive two points for a typical field goal, three points for a field goal made from beyond the three-point arc, and one point for each free throw they make. Basketball is played at all levels from youth leagues to professional leagues, and it is one of the most popular and widely watched sports in the world. The NBA (National Basketball Association) is the premier professional basketball league in the United States, while FIBA (International Basketball Federation) governs the sport internationally. (Logan, et al., 2023)

The National Basketball Association, or NBA, is a professional basketball league comprised of 30 teams across North America and is regarded as the top basketball league in the world. The NBA was founded in 1946 as the Basketball Association of America (BAA) and adopted its current name in 1949 after merging with the National Basketball League (NBL). The league is divided into two conferences (Eastern and Western) and each conference has three divisions with five teams in each division. The regular season consists of 82 games for each team, which runs from October to April, followed by the playoffs, where the top eight teams from each conference compete in a knockout tournament to determine the NBA champion. (NBA, 2019)

Positions in basketball

In basketball, there are five positions: point guard, shooting guard, small forward, power forward, and center. Each position has its own set of responsibilities and physical requirements. (NBA, 2019)

1. Point Guard (PG): Usually the smallest player on the team, the point guard is responsible for directing the team's offense, bringing the ball up the court, and setting up plays. They are typically fast and agile, with good ball-handling and passing skills.
2. Shooting Guard (SG): The shooting guard is responsible for scoring points, both from outside the three-point line and by driving to the basket. They are typically taller than the point guard and have good shooting skills.
3. Small Forward (SF): The small forward is a versatile player who can play both inside and outside. They are typically taller than the shooting guard and have a good combination of size, speed, and shooting ability.
4. Power Forward (PF): The power forward is usually one of the tallest players on the team and is responsible for rebounding, defending the basket, and scoring points from inside the paint. They need to be strong and physical, with good jumping and dunking skills.
5. Center (C): The center is the tallest player on the team and is responsible for defending the basket, rebounding, and scoring points from inside the paint. They need to be strong and physical, with good jumping and dunking skills. They are often the team's primary rim protector and help to anchor the team's defense.

Introduction to dataset

There is a total of three dataset used in this project, which is described in 

Different stats of player are used to determine the best player in each position.

1. PG: AST, TOV, STL, `X3P.`, PTS
2. SG: `X3P.`, `FG.`, `eFG.`
3. SF: TRB, `FG.`, X3P, `eFG.`, AST, PF
4. PF: BLK, ORB, STL, `eFG.`
5. C: BLK, TRB, `X2P.`, `eFG.`, `FT.`

Scenario

As a data analyst for the Chicago Bulls, an NBA (National Basketball Association) team that had a dismal season in the previous NBA season, finishing 27th out of 30 based on win-loss record, I can speak from experience. The team's $118 million player contract budget for the next season places them 26th out of 30 teams. Finding the top five starting players that the Chicago Bulls can afford, one for each position, has been given to the data analyst by the team's general manager. The analyst may select players that were already on the Chicago Bulls roster the previous year, but they must prove that they are worth the financial commitment. The goal is to find the best players while not using up all the team's budget on just these five players, as there is still a full team roster to consider.

Aims on Objective

The aim of this project is to produce top players in each position using different game metrics that define their play. Objectively, a linear regression model is developed to find the most consistent shooter in game by considering the number of points they scored in previous session in respect to stats such as expected 2points, 3points, free throws, offensive rebounds, and assists.

1.5 Libraries

1. tidyverse: The tidyverse is a collection of open source packages for the R programming language introduced by Hadley Wickham and his team that "share an underlying design philosophy, grammar, and data structures" of tidy data. (tidyverse, 2023)

2. dplyr: dplyr is a grammar of data manipulation, providing a consistent set of verbs that help you solve the most common data manipulation challenges: (tidyverse, 2023)

3. tidyr: tidyr is new package that makes it easy to “tidy” your data. Tidy data is data that's easy to work with: it's easy to munge (with dplyr), visualise (with ggplot2 or ggvis) and model (with R's hundreds of modelling packages). The two most important properties of tidy data are: Each column is a variable. (tidyverse, 2023)

4. ggplot: ggplot2 is an open-source data visualization package for the statistical programming language R. ggplot2 is declaratively and efficient in creating data visualization based on The Grammar of Graphics. (tidyverse, 2023)

5. broom: The broom package takes the messy output of built-in functions in R, such as lm, nls, or t.test, and turns them into tidy tibbles. (Wickham, 2023)

6. naniar: naniar provides data structures and functions that facilitate the plotting of missing values and examination of imputations. This allows missing data dependencies to be explored with minimal deviation from the common work patterns of 'ggplot2' and tidy data. (Tierney, 2023)

7. caret: The caret package (short for Classification And REgression Training) contains functions to streamline the model training process for complex regression and classification problems. (Cran, 2023)

2. Cleaning the dataset

3. EDA

Pairwise complete obs (tibco, 2015)

4. Data modelling

Car:durbinwatsontest (rdrr, 2023)

Homo (scribbr, 2023)

Player Rank (BUCKLEY, 2022)

Z score (NEVIL, 2023)

Normality (statisticssolutions, 2023)