

ORIE 4741 Project Proposal

Genghis Shyy (gs484), Heesun Chang (hc483), Mohammad Kamil (mk848)

October 4, 2020

The Problem

The goal of this project is to analyze the different relationships between NBA players' statistical performance and the contracts they receive. More specifically, we plan on exploring a number of questions such as: How much does a player's shooting efficiency typically influence the general perception of a player's points-per-game statistic? How does a player's position and/or age affect the statistical expectations that player receives? Can we even confidently assert that there does indeed exist a positive correlation between a player's points-per-game statistic and that player's salary or contract length? By exploring such potential relationships between players' statistical profiles and their contracts, we can ultimately hope to reliably predict the standard market contract any given NBA player should be awarded.

Dataset Overview

In order to accurately predict future contract details for any given NBA player, we have gathered two paired datasets—the first of which stores all major 2017-18 regular season statistics for 540 NBA players; and the second of which stores contract details for 523 NBA players from the 2018-19 season to the 2023-24 season. More specifically, the first dataset provides information on each player's primary position(s), age, and team(s), along with 25 different statistics including various per-game statistics (ex: points, rebounds, assists, steals, blocks, etc.) and full season statistics (ex: field goal percentage, three-point percentage, free throw percentage, number of games played, etc.) The second dataset similarly provides each player's team (at the start of the 2018-19 season), along with each player's 1) base per-season salary, over the seasons of 2018-19 through 2023-2024; 2) contract type; and 3) total guaranteed salary. Therefore, with this information in mind we can directly study and visualize the various impacts each player's statistical performance in the 2017-18 regular season had on his subsequent contract.

Both of the above datasets were found on basketball-reference.com, a third-party website created by a team of engineers, statisticians, sport analysts, and computer scientists. The raw data itself is made available by National Basketball Association and its staff, who collects and stores relevant information after every game.

Importance of Problem

As fans all over the world watch NBA championship games, it is often wondered what factors actually determine who gets paid the most in the NBA. Indeed, even a cursory browse through the salaries and statistics of past and present NBA players reveals that it is not necessarily the players who play the most minutes, secure the most rebounds, or even score the most points who receive the league's most rewarding contracts. Therefore, by developing machine learning models and various forms of statistical analysis, this project aims to give us a more nuanced means of predicting NBA player contracts. This in turn will reduce the chance we give players either under- or overpriced contracts, and additionally facilitate effective team building by accurately identifying which players are most statistically valuable to team success. Last but not least, with our findings and prediction models, we also plan to determine which players have been over- or undervalued by comparing their given salaries vs. predicted salaries—thereby allowing us to more easily sign or trade for better valued players.