NBA Project

Group 24

De Sanctis Riccardo, Scognamiglio Gabriele, De Sanctis Matteo

Table of Contents

Introduction

The intent of this project is to get a deeper understanding of NBA basketball and a wider picture of the game that is not limited to or influenced by personal beliefs and opinions.

That was reached through the analysis of numerous statistics and data; and through the application of the statistical tools that have been studied this year.

Main targets of the Project:

- Finding the most efficient Field Goal
- Understanding the change in three-pointers style of game in modern NBA
- Understanding the change in triple doubles style of game in modern NBA
- Predicting the Rookie of the Year
- Simulating a Match
- Simulating Playoffs
- Clustering players to find similarities
- Predicting the Most Valuable Player

The plenitude of data used was collected from the official NBA site and thus it was totally reliable. To actually retrieve data it was used nba_api: an API Client package to access the APIs for NBA.com. This API analyzed and mapped as many endpoints on NBA.com as possible, hence it allows to reach 'easily' plenty of NBA data and statistics.

A lot of work was put into data cleaning and manipulation to make it usable to the project's aims.

The Project is a Jupyter Notebook file with some extensions such as: RISE, Split Cells Notebook and Hide Input. These extensions are not required to run the code but just to properly visualize the presentations slides that are integrated within the code.

Most Efficient Field Goal

The intent of this section is comparing the different types of shots in Basketball and understanding which one results in the highest production of points.

Effective Field Goal metric (eFG%) is used to rank the different types of shot. It weights more the shots worthing three points with respect to normal shots worthing only two.

Shooting trends and shooting percentages over different seasons are then visualized.

The field is divided into different zones and the effective shooting efficiency for each zone is compared.

The shot that has the highest eFG% will hence be the most efficient.

Evolution of the game of modern NBA

Change in three points style of game

This section explores how the modern league is changing style of game subsequently to three points shots success these later years.

Shooting percentages and trends are visualized.

Stephen Curry contribution to the evolution of the game is analyzed and compared to NBA history players.

Change in Triple Double style of game

Moreover, the increment in number of Triple Doubles is displayed as well.

Russell Westbrook iconic seasons are analyzed.

Rookie of the Year Prediction

This section aim is to understand if it was possible to predict who is awarded with the Rookie of the Year award.

The award should be given to the rookie with the best statistics but sometimes personal preferences of the voters alter the choice.

Logistic regression has been the tool used.

It weighted every box stat and understood the correlation between them and the award winning.

It trained over 2007 to 2020 seasons and predicted the 2021 Rookie of the Year.

The player with the highest probability of winning is selected as the winner.

Match Simulation

This section allows to simulate an NBA match by selecting the two teams and the season.

Points scored and points allowed by the two teams during the season selected are retrieved and distributions are created.

Then the game is simulated. To win a game the score of a team must, obviously, be higher than the opponent's one. The score is calculated by adding a sample from the team points scored distribution and a sample from the opposite team points allowed distribution.

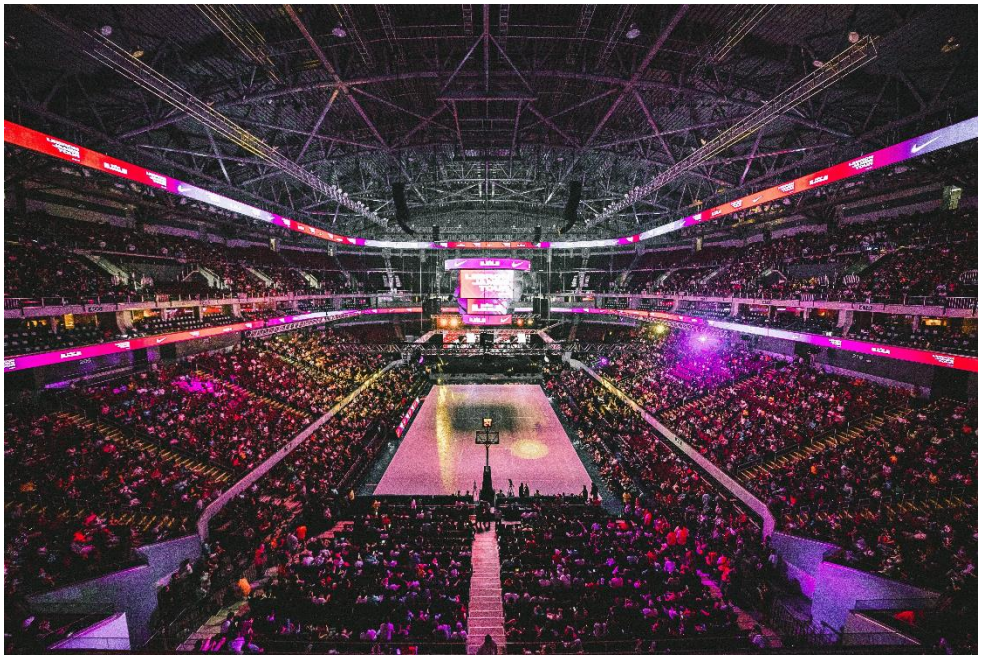
The simulation can be carried out by performing either a Parametric Bootstrap or a Non-Parametric one.

In case of Non-Parametric Bootstrap, the score of each team is simply calculated by sampling of the two distributions.

In case of Parametric Bootstrap, mean and standard deviation are calculated for each distribution. Then the score is calculated by sampling from a two Normal distribution with these calculated mean and standard deviation as parameters for both points scored and points allowed.

Playoffs Simulation

This section is based on the previous one. Its intent is to simulate the NBA Playoffs to get an accurate prediction of the NBA Champion based on the result of the Regular Season.



The data of each team are retrieved for the desired season. To win a playoff series any team must win four matches out of seven. The first that reaches four wins passes the round and go straight to the next.

Parametric Bootstrap Match Simulation is used to simulate every game of the Playoffs.

Players Clustering

This section is aimed at finding similarities among players and group them is homogeneous groups. These groups can be identified as position roles or even as worth to the team.

The player are grouped by the main basketball statistics.

K-Means is the algorithm used to perform the clustering.

Most Valuable Player Prediction

Similarly to the Rookie of the Year prediction, this section aims to predict which player is going to be awarded with the MVP award.

The MVP choice is influenced by the media even more then the ROTY award, where popularity and personal preferences supress actual talent and better performances.

Logistic Regression was used once more. The player that resulted with the highest probability was selected to be the Most Valuable Player of the year.

This time, the accuracy of the prediction was verified. The prediction was made for each of the season from 1970 to 2021, predicting the given season and training on the others.

References

- Predicting the 2020 NBA Champion with Machine Learning
- Make a Simple NBA Shot Chart with Python