Contents

1 Introduction

Hot hands is a very talked-about phenomenon in all sports, with numerous definitions. Hot hands are highly common in sports like basketball and baseball. There are many discussions that support the idea of getting a player hot (in means of making more baskets back to back or scoring home run streaks) during a game, but there are just as many that oppose it (Gilovich et al., 1985). In this research, we look at the hot hand phenomenon in NBA 2021–2022 playoff games.

Basketball Reference is one of the largest internet resources, with information on all players and numerous seasons (LLC, n.d.). Web scraping was used to collect the data. Every match's play-by-play information is scraped, and we are aware that this information is text-based. We pre-processed the data after scraping it to create a perfect tabular form that we could utilize for later analysis and visualization. We attempt to explain the impact of a player getting hot hands on the team's win and loss using graphs and probability.

For Data formation and scrapping from web we are using R software in version 4.0.5 (R Core Team, 2021). And for data pre-processing and visualization we are using python in version 3.6.15 (van Rossum, 2021).

2 Data Formation

In this research, we mainly focused on the 5 players (Maxi Kleber, Giannis Antetokounmpo, Stephen Curry, Luka Dončić, Chris Paul) and we consider the data from previous season (2021-22) playoff matches. First, we web scrape the data from basketball reference, after that we create a comma separate file (CSV) for every player. For the scraping step we use tidyverse and rvest packages in R (Hadley Wickham, 2022). With the help of these packages we read the play-by-play stats table from the basketball reference and save the file using the write CSV package in R.

After scrapping we first pre-process the text data and convert it into a table, which we further used for visualization. For this conversion step, we use python and pandas library (pandas development team, 2020). For the data manipulation step and performing some filtering on data pandas have very useful and simple commands. As a result of this step, we get a file that has major features like makes/misses count, Short range, Series Game, Basket type, Time of short, Score and quarter of game, etc.

3 Methods

In this section we discuss about the methods that we use in this project.

3.1 Scatter plots

On a Cartesian plane, the scatter plots are used to show the relationship between two variables. An additional variable can be plotted on the same plane to demonstrate a relationship if the data points are classified by color, size, or shape. The data points are shown as Cartesian coordinates, with the value of one variable projected onto the X-axis and another variable projected onto the Y-axis (Hay-Jahans, 2018)[159].

3.2 Bar plots

To compare the estimated central tendency of numerical or categorical data, bar plots are primarily utilized. In bar plots, the width would be uniform across all bars, and the height of the bar would indicate the estimate or frequency of the variable. Although it varies from author to author, the bar is often vertical in orientation (Hay-Jahans, 2018)[111].

3.3 Point plots

Using the locations of the scatter plot points, a point plot represents an estimate of the central tendency for a numerical variable and uses error bars to show the degree of uncertainty in the estimate (Waskom, 2021a).

For all the above plots we are using pre-built libraries matplotlib and seaborn in python (Waskom, 2021b), (Hunter, 2007).

3.4 Streak probabilities

We determine the convergence probabilities for Maxi Kleber player analysis. The calculation is as follows: first, we use the groupby function to get the cumulative sum, count, and streak of shorts (pandas development team, 2020). Second, we use a built-in function in pandas crosstab to construct a frequency table between makes/misses count

and short streak (pandas development team, 2020). This frequency table tells how many shorts a player attempts during a game and how many of those short attempts are successful and unsuccessful in relation to short streaks. Using these figures, we arrive at the following calculation for the convergence likelihood of making a basket:

Probability of convergence at 1st streak is:

$$P(streak = 1) = \frac{\text{Number of shorts made in 1st attempt}}{\text{Total number makes}}.$$

Probability of convergence at 2nd streak is:

$$P(streak = 2) = \frac{\text{Number of two shorts made consecutively}}{\text{Number of shorts made in 1st attempt}}.$$