# The Hot Hand Phenomenon

**Leon Wang** LJW7736@NYU.EDU
**Peter Hu** RH3659@NYU.EDU
**Yuan Li** YL10134@NYU.EDU

**Milestone 1 – Project Proposal**

## 1. Introduction

In professional sports, such as the NBA, some players, often for a short period, may reach a state of exceptionally high performance, where they consistently make a streak of successful shots that exceeds what would be expected by random chance. The hot hand phenomenon refers to the situation in which a player who has made several successful attempts is more likely to succeed in subsequent attempts. Is the hot hand phenomenon real, or is it just another type of Gambler's Fallacy, where people mistakenly believe that random events in the past can influence the probability of the future?

## 2. Motivation

The hot hand phenomenon is of particular interest today, since it is a real-world problem with real data, and it touches upon the foundational issue of pattern recognition: is there a real pattern or just a misperception of random sequences? The hot hand phenomenon also intrigues because different perspectives can be developed using various methodologies, ranging from traditional statistical analysis to models employing deep learning.

## 3. Related Works

The influential study by Gilovich, Vallone, and Tversky concluded that there is no evidence to support the hot hand phenomenon. They stated that "the belief in the hot hand and the 'detection' of streaks in random sequences is attributed to a general misconception of chance according to which even short random sequences are thought to be highly representative of their generating process." (Gilovich and Tversky, 1985)

Miller and Sanjurjo (Miller and Sanjurjo, 2018) analyzed the original work of Gilovich, Vallone, and Tversky. They introduced and formally proved streak-selection bias, and proposed corrected estimators and permutation tests that preserve the sequence structure intact. After correction for selection bias, evidence of significant streak shooting is revealed, contradicting previous conclusions of Gilovich, Vallone, and Tversky. However, their method did not include the correction for endogenous shot selection (players take harder shots when "hot") or defense adjustments.

In a new study by Pelechrinis and Winston (Pelechrinis and Winston, 2022), they utilized the two-season (2013, 2014) dataset from NBA games to analyze shooting performance in real-game situations. In their study, they focused on assessing the probability of success in the shots. A neural network with 4 hidden layers was used to predict shot made probabilities based on various factors. Their model achieved an out-of-sample accuracy of 66%. Their study showed that while individual players can exhibit the hot hand, the average player tends to regress to their mean performance after consecutive successful shots.

## 4. Methodology

We plan to analyze the hot hand phenomenon from various angles.

- First, we will implement permutation tests that preserve each player's shot sequence and spacing (respecting selection bias from conditioning on past hits/misses).
- Second, we will perform logistic regression for shot make probability with predictors: recent streak state (e.g., last k outcomes), shot distance/zone, shot clock, period, home/away, defender distance proxy (when available), and player random effects. Then we can compare the coefficients on "streak state" with and without context controls to isolate true heat from shot-selection/defense responses.
- Third, we will convert shot made probabilities to expected points added and examine win-probability deltas for shots taken in "hot" vs "neutral" states.
- Fourth, we will rebuild the deep learning model used by Pelechrinis and Winston to improve upon their model. Then we will apply other machine learning models on the data to see their performance.
- Fifth, we will use LLMs such as ChatGPT with various prompt technologies to explore their responses on the hot hand phenomenon.

## 5. Dataset

https://www.nba.com/stats

https://github.com/shufinskiy/nba_data

https://raw.githubusercontent.com/hwchase17/sportvu/master/joined_shots_2013.csv

https://raw.githubusercontent.com/hwchase17/sportvu/master/joined_shots_2014.csv

The last two datasets were the 2013 and 2014 two-season NBA shots records used by Pelechrinis and Winston. We did some exploration of the data. Here is the basic information.

```
Index: 400671 entries, 0 to 401494
Data columns (total 42 columns):
 #   Column                       Non-Null Count   Dtype
```

```
---  ------                     --------------  -----
 0   Unnamed: 0                 400671 non-null  int64
 1   CLOSEST_DEFENDER           400671 non-null  object
 2   CLOSEST_DEFENDER_PLAYER_ID  400671 non-null  float64
 3   CLOSE_DEF_DIST             400671 non-null  float64
 4   DRIBBLES                   400671 non-null  float64
 5   FGM                        400671 non-null  float64
 6   FINAL_MARGIN               400671 non-null  float64
 7   GAME_CLOCK                 400671 non-null  object
 8   GAME_ID                    400671 non-null  int64
 9   GAME_ID.1                  400671 non-null  int64
 10  LOCATION                   400671 non-null  object
 11  MATCHUP                    400671 non-null  object
 12  PERIOD                     400671 non-null  float64
 13  PERIOD.1                   400671 non-null  float64
 14  PTS                        400671 non-null  float64
 15  PTS_TYPE                   400671 non-null  float64
 16  SHOT_CLOCK                 379873 non-null  float64
 17  SHOT_DIST                  400671 non-null  float64
 18  SHOT_NUMBER                400671 non-null  float64
 19  SHOT_RESULT                400671 non-null  object
 20  TOUCH_TIME                 400671 non-null  float64
 21  W                          400671 non-null  object
 22  GRID_TYPE                  400671 non-null  object
 23  GAME_EVENT_ID              400671 non-null  float64
 24  PLAYER_ID                  400671 non-null  float64
 25  PLAYER_ID.1                400671 non-null  int64
 26  PLAYER_NAME                400671 non-null  object
 27  TEAM_ID                    400671 non-null  float64
 28  MINUTES_REMAINING          400671 non-null  float64
 29  SECONDS_REMAINING          400671 non-null  float64
 30  EVENT_TYPE                 400671 non-null  object
 31  ACTION_TYPE                400671 non-null  object
 32  SHOT_TYPE                  400671 non-null  object
 33  SHOT_ZONE_BASIC            400671 non-null  object
 34  SHOT_ZONE_AREA             400671 non-null  object
 35  SHOT_ZONE_RANGE            400671 non-null  object
 36  SHOT_DISTANCE              400671 non-null  float64
 37  LOC_X                      400671 non-null  float64
 38  LOC_Y                      400671 non-null  float64
 39  SHOT_ATTEMPTED_FLAG        400671 non-null  float64
 40  SHOT_MADE_FLAG             400671 non-null  float64
 41  season                     400671 non-null  object
dtypes: float64(23), int64(4), object(15)
memory usage: 131.4+ MB
```

This dataset contains two seasons (2013, 2014) of NBA shot records.

1. Total number of games: 2453
2. Total number of players: 576
3. Total number of shots: 400671
4. Average number of shots per game: 163.34

The basic statistics of shot percentages are the following.

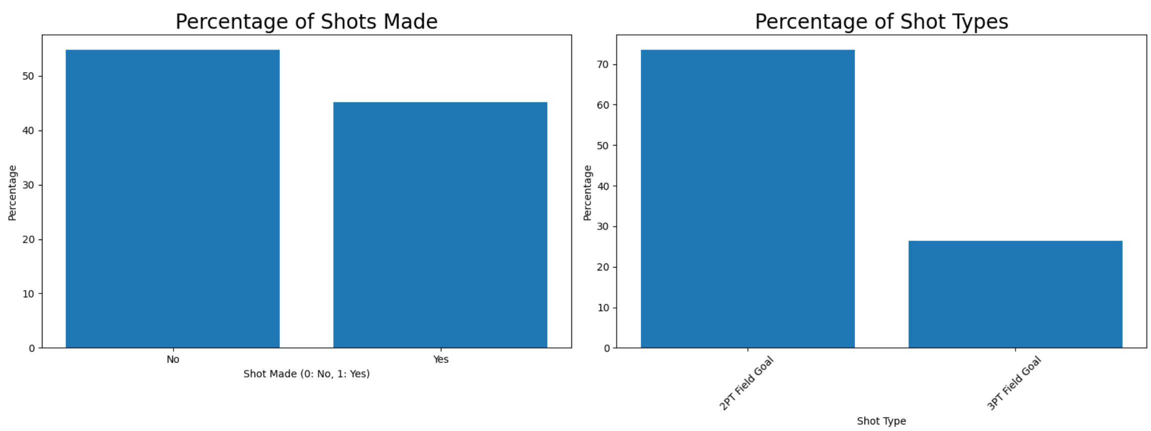


Figure 1: Shot Made and Shot Types Percentage

The majority of shots were from short distances to the basket. The shorter the distance, the easier it is to make it.

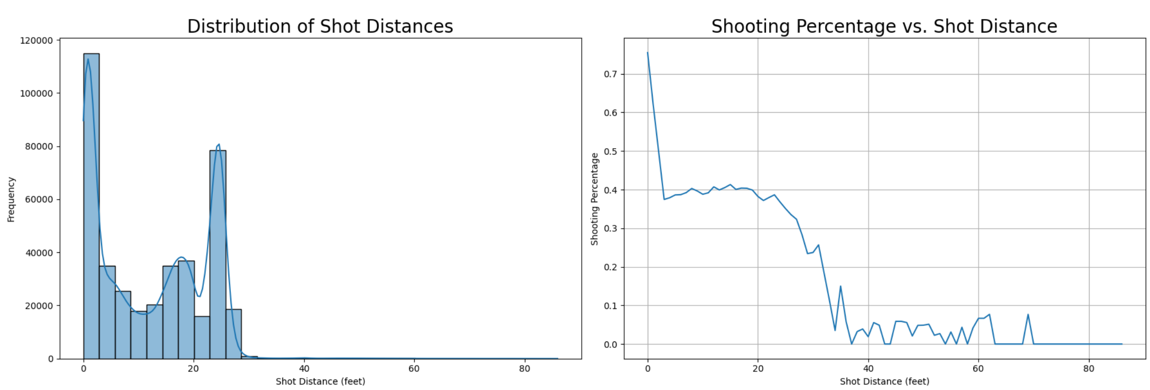


Figure 2: Shot Distance and Shots Made Percentage

Since the hot hand phenomenon is associated with streaks of shots made, we calculate the percentage of games in which at least one player made N consecutive shots.

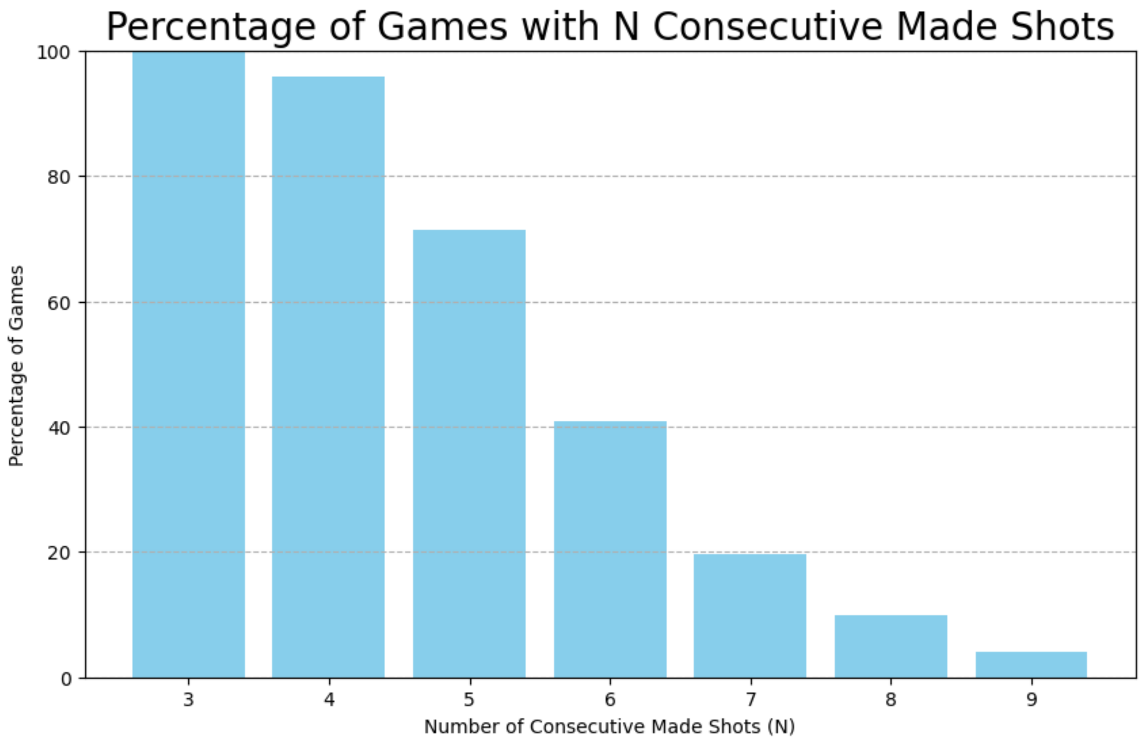


Figure 3: Percentage of Games of N Consecutive Made Shots

## 6. Work Plan

Peter suggested the topic of the hot hand phenomenon, and he also recommended the paper by Miller and Sanjurjo (Miller and Sanjurjo, 2018). The paper by Pelechrinis and Winston (Pelechrinis and Winston, 2022) was recommended by Leon, and he also did the EDA with visualization using the data from that paper. Yuan reminded us to connect the topic with the current GenAI research trend. Peter and Yuan provide materials, and Leon prepares the draft of this Project Proposal.

We will all be fully involved in the project. Peter and Yuan will more focus on the first three items listed in the Methodology. Leon will focus more on the last two items.

Tentative weekly plan:

1. Week of 10/12: Study the papers
2. Week of 10/19: Study the papers, examine the data and results in the paper
3. Week of 10/26: Build models, prepare the data

4. Week of 11/02: Build models, test models
5. Week of 11/02: Test models, analyze the results
6. Week of 11/02: Organize the results, prepare Milestone 2
7. Week of 11/09: Evaluate the progress, plan for additional work
8. Week of 11/16: Perform additional work
9. Week of 11/23: Evaluate the progress on the additional work
10. Week of 11/30: Finish the additional work
11. Week of 12/07: Wrap up, write the final report

## References


Vallone R. Gilovich, T. and A. Tversky. The hot hand in basketball: On the misperception of random sequences. *Cognitive Psychology*, 17(3):295–314, 1985.

J. B. Miller and A. Sanjurjo. Surprised by the hot hand fallacy? a truth in the law of small numbers. *Econometrica*, 86(6):2019–2047, 2018.

K. Pelechrinis and W. Winston. The hot hand in the wild. *PLoS ONE*, 17(1):e0261890, 2022.