# Research Proposal: Predicting Shot Quality in the NBA Using Defensive Pressure, Shot Creation, and Contextual Factors

## Objective:

This study aims to develop models that predict the quality of a shot, represented as the probability of success, by analyzing factors like defender distance, shot creation methods, and situational contexts such as the shot clock and dribbles taken before the shot. The analysis will provide insights into good vs. bad shots and variations in shot quality across players and play types.

## Research Questions and Corresponding Methodologies/Models:

### Research Question 1:

Can we predict the probability of a shot being made based on player behavior and situational factors such as defender proximity, shot clock timing, and dribble count? Are "bad" shots for most players still good shots for elite shooters?

* Methodology:
  + Use logistic regression as a baseline model to predict the likelihood of a shot being successful (made or missed) based on various features such as:
    - Defender Distance: Categorical distance ranges (e.g., "tight," "open").
    - Shot Clock: Time remaining on the shot clock as a continuous variable.
    - Dribbles: The number of dribbles before the shot.
    - Player ID: To capture player-specific shooting abilities.
  + Model: Logistic regression is selected for its interpretability, showing how each factor contributes to shot success.
  + If necessary, we will also explore more complex models, such as Random Forest and XGBoost, to capture non-linear relationships.

Research Question 2:

How do players like Steph Curry, known for making high-difficulty shots, compare to league-average players regarding shot quality?

* Methodology:
  + Extend the logistic regression model by incorporating player-specific terms using a multilevel (hierarchical) model. This model will account for player-level differences in shot-making ability.
  + Model:
    - A multilevel logistic regression model where player ID is treated as a random effect. This allows us to adjust predictions for individual players and estimate how "good" or "bad" a shot is relative to league norms.
    - This will help quantify whether shots considered "low quality" for average players are still good for elite shooters.

Research Question 3:

What is the optimal distance from a defender that players should have on a catch-and-shoot? Does this distance vary significantly among different players?

Research Question 4:

How much separation does a player need in isolation to increase shot quality, and what other factors contribute to shot success in isolation plays?

Research Question 5:

How does the remaining time on the shot clock influence shot quality, and does it differ based on play type (e.g., pick-and-roll vs. transition)?

Research Question 6:

How does the probability of success and quality of shot vary based on the type of shot? Are dunks more probable and of “higher quality” than 3-pointers?

Research Question 7:

When do players attempt shots of higher/lower quality? Do players take more risks in the playoffs rather than the regular season?

## General Methodology:

1. Data Exploration
2. Feature Engineering:
   * Defender proximity: Categorical or continuous distance metrics.
   * Shot creation method: Categorical feature to distinguish between catch-and-shoot, isolation, and other shot types.
   * Player-specific information: Include player IDs and positions to account for shooting ability.
   * Time on shot clock: Group shot clock time into ranges (e.g., "early," "mid," "late").
   * Dribble and touch time: Create features from the number of dribbles and possession time before the shot.
3. Model Selection:
   * K Means to cluster shots of similar quality.
   * Logistic Regression to output a probability of success associated with every shot.
   * Possibly use Pytorch to create a nn model using ReLU and Softmax activation functions to output a probability of success.
4. Model Evaluation:
   * Models will be evaluated based on accuracy, precision, and recall for classification tasks and cross-validation to ensure generalizability across seasons.

## Expected Outcomes:

* A predictive model for shot success based on various situational factors.
* Insights into how shot quality varies across different pay types, and game situations.
* Optimal conditions for shot types like catch-and-shoot and isolation, providing practical recommendations for teams and players.