Project 2 Proposal

Team Members
Jonathan Hsiao, Victoria Brendel, Michael Garfagnoli

Github Repo
https://github.com/UC-Berkeley-I-School/datasci200_project2_hsiao_brendel_garfagnoli

Primary Dataset
https://www.kaggle.com/datasets/ehallmar/nba-historical-stats-and-betting-data/

This dataset consists of three tables that we will be analyzing:

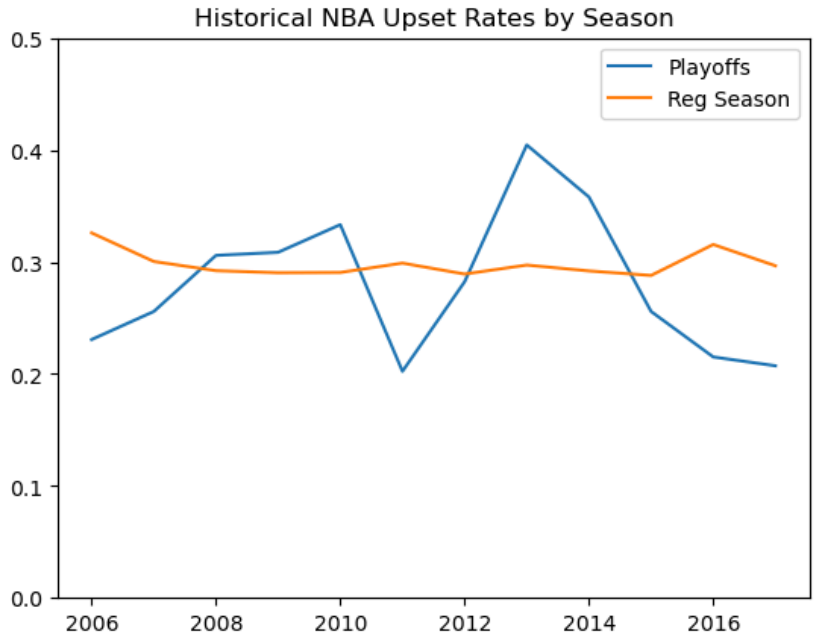
- **nba_betting_money_line**
  - Description:
    - Historical betting lines for NBA games
  - Important variables:
    - **price1** and **price2** represent the money line for each game, and we will use them to classify whether a game is an upset or not (depending on the winner)
  - Primary key:
    - **game_id**
  - Foreign keys:
    - **team_id**
    - **team_id_a**
- **nba_games_all**
  - Description:
    - Historical data on outcomes of NBA games
  - Important variables:
    - **wl** gives us the outcome of the game, which will be used to classify whether a game is an upset or not
    - **is_home** is a key feature indicating whether the team is the home team
    - **w_pct** gives us the team's record prior to the game, also a key feature
  - Primary keys:
    - **game_id**
  - Foreign keys:
    - **team_id**
    - **team_id_a**
- **nba_players_game_stats**
  - Description:
    - Historical data on players stats in each game
  - Important variables:

        - This table has several columns of player stats and attributes - these will be potentially important features to measuring individual player impact. Identifying which columns are ultimately the most important will be an outcome of our analysis.
    - Primary keys:
        - **player_id**
        - **game_id**
    - Foreign keys:
        - **team_id**

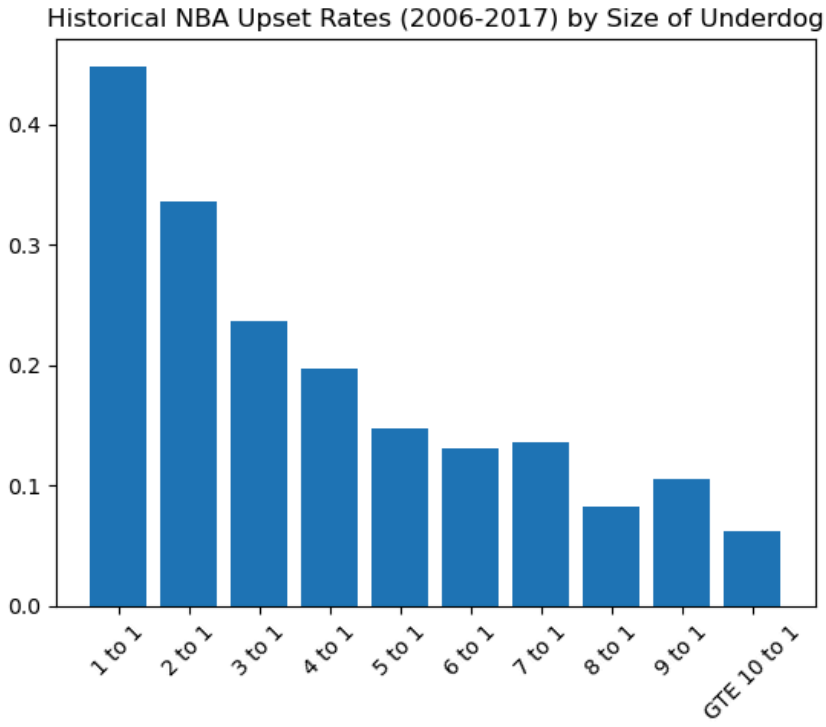
The tables can be joined together by matching the keys mentioned above.

Initial Exploration

Upset rates have been relatively flat around 30% over the period from 2006-2017.
Playoffs upset rates are more variable, likely due to smaller sample sizes.
What happened in 2013?



Upset rates decrease as the size of the underdog increases (as expected).
How does this translate to expected payouts if you had bet the underdog?



What we plan to cover in the final report

We aim to better understand what causes an upset in the NBA, where an upset is defined as a game where the winner was the underdog according to betting lines. We anticipate that could involve answering questions such as (but not limited to) the following:

- What teams have historically had the highest rate of winning upsets? Losing upsets?
- Have upsets decreased or increased over time?
- How do home and away team dynamics factor into upsets (both in the initial line and the final result)?
- How does time since last game factor into upsets (short vs long breaks)?
- Are any specific players especially valuable for achieving an upset? Detrimental?
- Does team experience (players with many YOE) help in achieving or preventing upsets?
- How does the chance of winning an upset differ by how large a favorite one team is (according to initial lines)?
- Do certain positions (Guard, Forward, Center) have a larger impact on upsets than others?
- How do playoffs differ from regular season when it comes to upsets?

We will organize our report by