Setting the Odds

https://www.actionnetwork.com/education/how-do-betting-lines-work-vegas

- Odds set based on account win totals, future odds
- Odds adjusted based on home advantage, weather, injuries, specific head to head matches

Lopez, Matthews, Baumer

https://github.com/yp0328/sportbet/blob/main/How%20often%20does%20the%20best%20team%20win_A%20unified%20approach%20to%20understanding%20randomness%20in%20North%20American%20sport.pdf

- Due to narrower distributions of team strengths and smaller home advantages, a typical contest in the NHL or MLB is much closer to a coin-flip than one in the NBA or NFL. The NBA demonstrates both the largest dispersion of talent and the largest home advantage→ **NBA (sport of choice in betting)**
- NBA home advantage: teams averaging a 62.0% chance of winning v. like-caliber opponent

Sport (q)	t_q	n_{games}	$\bar{p}_{games}$	n_{bets}	$\bar{p}_{bets}$	Coverage
MLB	30	26728	0.541	26710	0.548	0.999
NBA	30	13290	0.595	13245	0.615	0.997
NFL	32	2560	0.563	2542	0.589	0.993
NHL	30	13020	0.548	12990	0.565	0.998

TABLE 1

Summary of cross-sport data. t_q is the number of unique teams in each sport q. n_{games} records the number of actual games played, while n_{bets} records the number of those games for which we have a betting line. $\bar{p}_{games}$ is the mean observed probability of a win for the home team, while $\bar{p}_{bets}$ is the mean implied probability of a home win based on the betting line. Note that we have near total coverage (betting odds for almost every game) across all four major sports.

- NBA: 82 games a season (NFL-16, MLB-162, NHL-82)

provide justification on which betting to use (fantasy vs. over/under)
draft a more detailed background/introduction with proper citations

Background
Daily Fantasy Basketball

https://www.connoryoung.com/resources/AML_FinalProject_Report.pdf
https://wilsonholland.medium.com/hacking-the-nba-maximizing-dfs-lineups-with-machine-learning-4ce9728712c9

- Players gain points for scoring baskets, getting rebounds, steals, and blocks. They lose points for committing turnovers.
- A new lineup is created for each slate of games (for whichever NBA games are scheduled for that day)
- Limits: total salary cap (each player has a salary) and every position must be filled
- Fan Duel
    - Total salary cap: $6000

- Positions: 2 Point Guards, 2 Shooting Guards, 2 Small Forwards, 2 Power Forwards, 1 Center
- NBA Piggy Bank Shot: Tournament size of around 280,000 with a 5 cent buy-in and the top 30% get paid out. People can enter up to 150 lineups so it is important to note that 280,000 unique people are not playing at once.
- Draft Kings
  - Total salary cap: $50000
  - Largest platform with most external data and research → focus on **DraftKings**
  - Goal: ***determine and predict fantasy scores for all players so that the most efficient lineup can be constructed in terms of fantasy value per salary unit***

Assumptions to question & points to consider
- Team defense
  - In most models, team defense does not matter, which assumes that the player is averaging the same amount of points regardless of the opponent they play. Different models can adjust this by fluctuating a player's predicted points based on previous performance on similar opponent teams.
- Player's averages
  - Player averages are determined by x number of past games.
- Injuries
  - Determine if injuries will remove or suppress a player from the dataset.
- Good players get priced out
  - Dependent on a winning streak of good performances which causes an inflation in salary
- Cheaper utility players
  - Players who show up every game, average around the same points every game, and cost around the same price every game → limits ceiling but creates stability and does not affect the model too much

Models
- Various approaches for the problem: predicting the fantasy score directly, training models on solely a player's personal history versus the entire league's history, lineup position-based models, team-based models, etc.
- Feature engineering
  - Team: The current team a player is on
  - Opponent: The opponent a player is facing
  - Home or Away: Whether the game is home or away
  - Position: What position the player plays
  - Rolling averages of each individual player stat over 3, 5, and 10 game windows (i.e. points scored over last 3 games, points scored over last 5 games, points scored over last 10 games, rebounds over last 3 games, rebounds over last 5

games, rebounds over last 10 games, etc.). There are a total of 24 individual player stats we are measuring, resulting in 72 additional features.
- Odds: The probability of the player's team winning this game, according to Vegas betting odds
- Train set: first four seasons, test set: last season
- "While seven random forests performed well on the dev set compared to baseline, the default forests had high variance resulting in overfitting, as indicated by the large decrease in R2 scores from train to dev."