**Get Strategy Together**

* Visualize Data
  + Look for historical hole in data
  + Who has vegas been wrong about in 2019 thus far? Why is that?
  + What is a more stable bet, amongst Spread, Moneyline, and Total
    - How do the ups and downs of a game (score-wise) help predict the totals?
* Simple regressions to get a sense of what is driving what
* Think about ML
  + Maybe start with PCA to narrow down variables contributing to win? Will want data on teams prior to that game they play to get a sense of what everyone else is thinking?
* What (lagged) level of time makes the most sense to look into as we’re doing this?
  + Is it best to look at team performance in last 3 games? 4 games? What is vegas doing? Where are they off..?
* Something worth confirming: Is Pinnacle generally pretty close to PGW on a given night? Want to make sure of this or else using Pinnacle data isn’t the right move (should use another company)
* What is the confidence interval of spread? When does it make sense to jump in?
* Does Vegas predictions get better over time within the same
  + Visualize
  + Run Regression (controlling for year) to see if there is a decrease in incorrect predictions for Vegas from beginning of season to end.

**Betting on things that are inherently more predictive**

* Totals seem like a good place .. function of offensive/defensive ability of both teams without worrying about last 2 minutes
* Random Forest on TRUE/FALSE of whether bet will hit.
* **Model to predict final spread/total given spread/total in x minute of x quarter (live betting)**
  + Q’s:
    - When’s the best time to take a live bet?
  + Input:
    - Original spread/total + weighting of who is betting on what
    - Look at players who are doing well vs not. Who is playing minutes? Does the team have ability to turn it around? Will they regress?
    - Have some specific team/year fixed effect to control for certain iterations of teams that are really good/bad that yr
  + Nearest neighbors might be good. Find a game this year where the team came back in a similar situation (e.g., they were down 10 points in 1st quarter but Paul Pierce wasn’t playing well and had similar minutes (might be too specific)
  + Random Forest to choose which variables are important and not overweight
  + This goes well with the visualization of the margins for wins and losses for each game
* **Model to assess “good” or “bad” Moneyline odds (pregame). Would be especially helpful if it could give a probability that those live odds would change mid-game given that team and how they handle margin through the games so we could hedge.**
  + Idea here is to take advantage of movements in spread mid-game to hedge bet pregame
  + A really cool output would be for it to tell you where the odds would need to get to jump on and ensure some level of profit. Also would be cool if it assessed risk
  + This is good in that it takes little basketball knowledge to set up this model.. you’re just playing the odds (and looking at some historical data to predict the chance the team playing will score x amount)
  + Maybe the output could be the % chance a team has a chance to hit a certain margin in the second quarter.
  + Example walkthrough:
    - *Let’s say we want to assess the night’s slate of games. We see that the Celtics are +250 against the Knicks at home. We know that the Celtics often go up early, and the Knicks often go down early. We run the model and gives a prediction of what the margins might be after each minute (or what the possible variance of the point spread will be in the game).*
    - *We conclude via some model that the Knicks have a good chance to be down 10 (or maybe +1 (-110) in live spread) in the second quarter of the game.*
    - We decide to put in $100 on Celtics pregame ($250 payout). Then wait for 2nd quarter to put $110 on Knicks in second quarter ($100 payout) .
    - If the Celtics win I make $150. If the Celtics lose I make $0. You can switch the $ around to balance things out a bit more.

**Helpful Links**

<https://towardsdatascience.com/which-nba-players-are-most-similar-machine-learning-provides-the-answers-r-project-b903f9b2fe1f>

<https://towardsdatascience.com/making-big-bucks-with-a-data-driven-sports-betting-strategy-6c21a6869171>