**Problem Statement 1: PAC Spending**

Problem Statement: Can political advertising be predicted for each election cycle, including how much will be spent on adds and the proportion of advertising that will be attack or opposition advertising.

Pitch: Over the past two decades, political advertising spending has increased and advertising has become much more divisive as our two political parties have become more ideologically opposed. Because of this, I’m interested in looking at how money is spent on attack or support ads.

Specific questions include:

* Are opposition ads spent comparatively across Presidential, Senate and House ads?
* Do incumbents face more opposition advertising?
* And for model, can we predict the amount of money spent on opposition or support ads over the coming election cycles?

The data I’m intending to use is from the Federal Election Commission. The FEC tracks independent expenditures, spending on ads that specifically mention a candidate in a positive or negative light. The money for these ads comes from Political Action Committees that are independent of the candidate and campaign, and the PACs cannot coordinate with the candidate. The data set, “Independent Political Ad Spending (2004-2016)”, is on [Kaggle](https://www.kaggle.com/fec/independent-political-ad-spending).

Potential questions or challenges about the data include if there is enough data to develop an accurate model. With only 12 years of data and election only held every other year, there are only seven cycles of election year data with which to develop a prediction model. There is a lot of information about each expenditure, meaning it will require a lot of testing to identify which features impact future spending.

**Problem Statement 2: Is offense or defense more important to team success in baseball**

Problem Statement: Baseball teams don’t have the resources to obtain great pitchers and hitters. Is having a good pitching staff or a strong hitting team a better predictor of team success.

Pitch: For over a hundred years, management and front offices of major league baseball have tried to determine the best way to consistently win games. They’ve built rosters using hundreds of different methodologies to fill out their roster given their constrained resources. The overarching questions is should they strive to build their team to focus on pitching or hitting.

Specific questions include:

* Does a strong hitting team or strong pitching staff a better predictor of team success?
* Do front offices have better success spending money on pitchers or hitters? That is, is there a better alignment between salary on pitchers and a successful pitching staff or salary on hitters and a good hitting team?
* What is the optimal split in salary between hitters and pitchers?

The Baseball Databank from [Kaggle](https://www.kaggle.com/open-source-sports/baseball-databank) is a compilation of historical baseball data. As a team sport, baseball is relatively easy to pare down to individual contributions, specifically when it comes to pitching and hitting. Defense is a bit more difficult to quantify. We can use this data to evaluate winning percentage and whether hitting or pitching features are better predictors of that success.

Potential questions or challenges about the data is whether there is too much data and how to distill the multiple features within pitching and hitting down to identify what are the drivers of team success. A lot of testing will be needed to identify which features impact winning. The Baseball Databank comes in a multitude of files and tying them together requires additional work to complete the analysis.