Group name: acjr-project
Group members: Antony Awad, Cooper Powell, Johann Hatzius, Robert Gillespie

Project description and goals:

We aim to create a system that allows football coaches to determine the best play for an inputted scenario given historical NFL play-by-play data. In sports like baseball, data on teams' previous games has been used to determine outfield positioning and other strategic moves. However, a similar use of data to predict optimal outcomes has not been used in American football. Thus, our goal is to create a project that will be able to calculate the optimal play to reach a first down given the offense's specific situation and provide this information to the coach.

To describe the system in more detail, we envision that a user inputs data relevant to the offense's field placement and game environment. These variables will consist of the down (1-4), how many "yards to go" in order to reach a first down, the location on the field (where on their own half of the field or where on their opponent's half of the field), and the quarter (1-4). We are also considering using the defense's rating against the run and the pass as inputs, as well. Using this information, the system will then output one of 12 possible play options as the optimal one: run (left, middle, or right) or pass (left-short, left-middle, left-deep, middle-short, middle-middle, middle-deep, right-short, right-middle, or right-deep). With this play call, it will also output charts/graphs displaying the success rates of each of the 12 options based on averages of previous NFL plays that were in similar situations.

To achieve these goals, we plan to scrape the data of each play of every NFL game from the past 10 seasons. Using this information, we will determine the success rates of each play call in a given scenario (did the play result in a first down?) and use the previous success rates as a basis for our decisions for the optimal play call. To sort this data and make conclusions, we plan on grouping field positions into 10-yard partitions and dividing "yards to go" into short (3 or fewer yards), medium (4-7 yards), and long (8 or more yards). Using these groups, we will categorize the past NFL plays into their respective situations (e.g. first down between the team's own 20 and 30 yard line in the third quarter). When the system is given the input of a situation, it will determine the percent success rate (how often a first down was achieved?) of each kind of play in that situation from the past NFL data and output the best play choice. We will say that the best play call is the one that worked the most frequently in the past.

Data sources:

- Pro Football Reference (pro-football-reference.com) - A database of thousands of NFL games with a detailed play-by-play log. This will likely be the main database we will need to crawl and scrape, as most of the vast amount of data we need is contained within the play-by-play logs of this database.
- NFL.com - A website with detailed accounts of team stats, including team tendencies to run or pass the ball and when. We will probably use this as a secondary source of

information to augment and refine our analysis of the play-by-play. Potential sources of further refinement would be taking into account the opponent's defensive rating against both the run and the pass, as well as potentially even the team's passer ratings in particular situations.

Tasks and timeline:
- *Week 4*
  - Project written proposal and presentation
  - Plan framework by breaking down project into the methods we need to implement
- *Week 5*
  - Implement functions to crawl through the databases
  - Place retrieved data into database with columns categorizing plays into field position category and "yards to go" category
- *Week 6:*
  - Implement functions to retrieve past NFL play results in the same situation as the input
  - Check In
- *Week 7:*
  - Implement functions to determine success rates for each kind of play using retrieved data from past NFL plays in the same situation
  - Implement functions to return optimal play call from success rates of each play option
- *Week 8:*
  - Begin working on the terminal interface and specific functions for interpreting user response and providing the desired outcome as an output
  - Check In
- *Week 9:*
  - Complete work on terminal interface and input/output functions
  - Final polishing
  - Submit Friday