

Project proposal - NFL play by play data

Website: www.pro-football-reference.com/play-index/play_finder.cgi

Data Acquisition

09/23/2015

Group:

Alex Morris

Felipe Chamma

Felipe Formenti

Ryan Speed

The goal of this project is to collect data from the [website](#) mentioned above regarding every play of every game of the past seasons and present the most interesting insights that can be taken from it.

The main data table consists on the play by play data. Each row brings information such as: date of the game, offense and defense team, game quarter, number of yards, outcome of the play, expected points before and after the game, among other columns.

In the detail column there is a string from which we can extract many other important variables using regular expressions. Those variables include (but are not limited to): play type (pass or rush), status (complete or incomplete), location (left, right), whether there was a turnover, fumble, touchdown, and so on.

We are also going to extract another table with information on the game outcome. In that way, we may connect the game result with the game plays, which brings relevant insight on winning teams characteristics. Here are some examples of questions we expect to answer with that:

- On average, how many yards does a winning team achieve? What is the probability of a team with more than 500 yards lose a game?
- What's the touchdown to field goal ratios difference between winning and losing teams (especially inside the red zone)?
- Which team has the best yards-to-points conversion rate? What about yards-to-wins rate?
- What is the most common outcome of a rush play? What about a pass play? Which one brings more benefits and which one brings more risk?
- What's the conversion rate of onside kicks and fake field goals?

We hope that the answers to those questions may help people see games in the future in a more vibrant, passionate, data driven way, and also foster the competitiveness of the league.