

Final Project Proposal

Michelle Liu, David Jablonski, Shane Conner

Instructor: John Alexis Guerra Gómez

W241-3: Experiments and Causality

Domain

The premise of Fantasy Football is where participants manage a team comprising on NFL stars and compete amongst themselves in head-to-head league matchups each week of the season. A team accumulates points using the actual weekly performances of their selected NFL players they opted to "start" in that particular week.

League players with relatively little knowledge of current NFL players can be at a significant disadvantage when selecting players with their friends who may know a great deal about the NFL. Fortunately, before the NFL season even starts, experts and analysts around the country spend time anticipating player performance.

Concept

Our project proposes to leverage the vast amount of real-time information generated by experts into a format that users can easily synthesize and ultimately equip them with insights to execute an educated decision. To do so, we will aggregate, transform, and visualize expert projections in a concise, yet comprehensive manner.

Data

Typically, expert data on players comes in two primary varieties - rankings and projections. Similar to a security in the stock market, evidence supports that the most accurate projection of a player's performance is the the consensus projection. In the stock market, very few individuals outperform the "market" (e.g., S&P 500) over the long run because it is an aggregate of individual valuation across assets and as a result, mitigating individual biases. This is sometimes referred to as the "wisdom of the crowd".

By the same token, we want to extract real-time expert projections aggregated from multiple sources (ESPN, CBS, NFL.com, and FantasyPros, Accuscore, Fantasy Football Nerd, FantasySharks, FFtoday, Footballguys, FOX Sports, WalterFootball, and Yahoo). Expert analyst Isaac Petersen does just that [by creating a dataset of](#) aggregated expert analysis weight based on historical accuracy.

Unfortunately, we haven't found a way to extract this information without logging in so we're considering two options at this point. The first is manually uploading their weekly projections to an outside source our data can access. The second option is to develop our own script which extracts data from the various sources. However, this poses a challenge in runtime for our script and development time of that script. In consideration of that, we're leaning towards the first option to concentrate on the visualization.

Users

At the same time, our project *will* scrape data directly from ESPN to determine league information specific to a user. The users in this case are intended to be any Fantasy Football participant and to best service them we need to tailor our visualizations around their league.

One measure we're interested in and often used in sports analyst is "Value Over Replacement" (VOR) which standardizes the data to convey player value. This is done by taking the difference between a player's projections and their position's average points. To determine that average, we first need to know the player base (rostered players). As a result, we're interested in their existing team's roster, the field of unavailable players, and the field of available players. To accommodate this, our project will have an input field requesting a user's league ID and team abbreviation.

Tasks

Using this data, we want to construct a visualization that informs a user about what players to start and the value of players. To determine what players to start, one visualization we are considering is a simple graph which shows a player's projected points (ordinal, continuous data) on the X-axis and projected position rank (ordinal, discrete data) on the Y-axis. Since the player's projected points are continuous, a potential modification is to add the 95% confidence interval using the mean and standard deviation of each player's expert projections. The reasoning is that a user can gauge player's with a high floor (safer) or a high ceiling (riskier). For example, if the user is facing an opponent projected to outscore them, it might make more sense to use players with more variation (risk) to equip their starting team with the best odds to better their opponents' higher score.

This information helps determine who to start, but we also want to convey player value. We believe one solution to do that is to allow user to switch the variables in our visualization. The one limitation is that the X-Axis will be limited to continuous variables (projected points, value over replacement, etc) to maintain confidence intervals and Y-Axis to discrete variables.

Lastly, we want to color code the categorical data such as position (QB, RB, WR, TE, K, DST) and availability status (available, unavailable, user's team). As a default, player position will be color coded using changes in hue and player availability status in changes of saturation. Although, we also want to offer the user the functionality to filter or change the visualization of categorical data. The primary objective of our project is to construct a tool which allows the user to discover their own insights, fuel curiosity, and promote additional enjoyment to their Fantasy Football experience.

