Analysis of NFL Player Valuation

Isaiah Bryant, Edan Eingal, Kevin Li, Steven Liao

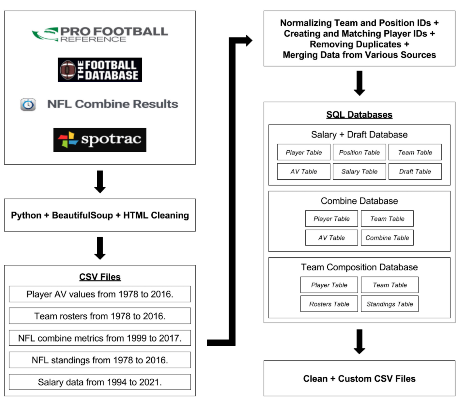
OVERVIEW

The central objective of our project is to create a means by which to evaluate the "true value" of a player in the NFL. Our objectives surrounding the determination of true value include:

- Approximating an appropriate salary for a player given their performance.
- Determining which players are over or underpaid.
- Predicting a draft prospect's performance based on combine metrics.
- Elucidating a draft pick's fair dollar value.
- Predicting a player's contribution to a team's playoff performance in a trade.

A key statistic in all of our models is the **Approximate Value** (AV) of a player. AV is a metric developed by *Pro Football Reference* that can be used as a proxy to describe the performance of a player at a given position over a period of one or more seasons.

DATA COLLECTION



- We **collected** AV data from 1978 to 2016 from *Pro Football Reference*, roster data from 1978 to 2016 from *Pro Football Reference* and *The Football Database*, combine data from 1999 to 2017 from *Pro Football Reference* and *NFL Combine Results*, standings data from 1978 to 2016 from *Pro Football Reference*, and salary data from 1994 to 2021 from *Spotrac*.
- We **cleaned** and **loaded** the data into **SQL databases**. We decided that it would be best to store data used in separate parts of our project in different databases.
- We **created** custom-made CSV files with cleaned data using **SQL queries**.

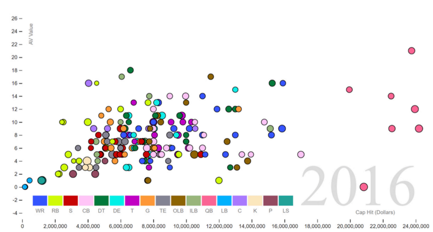
CHALLENGES

- Merging roster data and combine data from various sources and removing duplicates.
- Finding salary data from before 2011.
- Assigning the correct player IDs to players from different datasets.
- Deciding on a standard set of position and team IDs.
- Tuning the team composition model and attempting to avoid overfitting.

SALARY ANALYSIS

Our goal with salary analysis was to value players and draft picks in relation to how much they are paid. In other words, given the NFL's salary cap restraint, the value of a player is determined not just by performance, but also by how much money must be spent to acquire that player.

AV vs. Salary (top 10% in each position based on AV value)

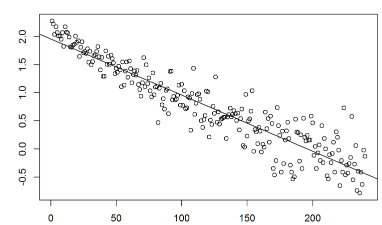


We began by creating a 'fair salary' metric which measures the salary a player deserves to make based on their performance for a given year. The formula for the fair salary of a player is:

$$fair\_salary(player\_av) = \frac{player\_av}{total\_av} * salary\_cap * number\_of\_teams$$

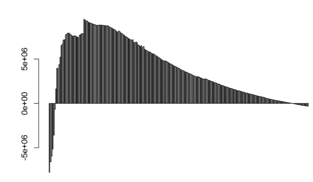
Next, we wanted to value draft picks. Since NFL rookie contracts are standardized and four years long, we looked at historical data for the performance over the first four years of a player's career.

Pick Number vs. Average AV per Season over Rookie Contract (Log)



Fitting an **exponential regression** to these data, we can approximate a player's AV contributions over their rookie contract given their draft pick number. From here, we approximated the fair contract for each rookie using our fair salary metric. The true value of a draft pick can be construed as the difference between the actual and fair salaries of a draft pick.

Pick Number vs. Pick Value in Dollars

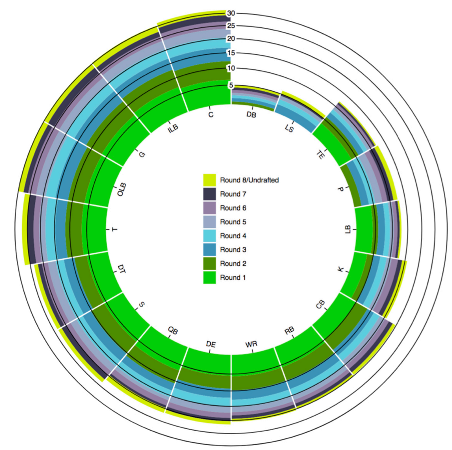


DRAFT ANALYSIS

Our analysis of the draft begins with the **1980** class, and groups all players in a given position and round together (with the exception of Round 8 and undrafted players). The results clearly show:

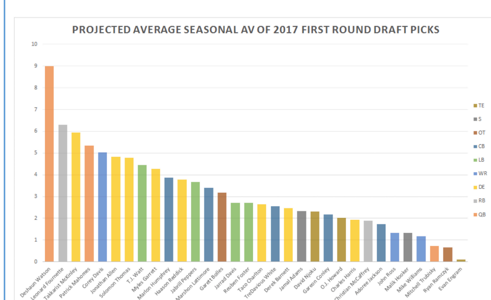
- A **decreasing** average AV value as draft rounds increase.
- A **higher** contribution to AV from certain positions such as quarterbacks (QB).
- A relatively **steady** average AV value for kickers (K) indicating that teams may wish to draft these positions in later rounds.

Average AV By Round Drafted



DRAFT COMBINE NEAREST NEIGHBORS

We performed a **k nearest neighbors** search using the combine and AV data that we collected in order to determine the projected seasonal AV of all players selected in the first round of the 2017 NFL draft. We based this analysis purely on **combine metrics**, choosing to ignore other measurables such as college statistics, personality, and injury history that NFL teams often utilize to evaluate draft prospects. The results below were obtained using a k-value of 3 and by limiting comparison to neighbors within the same position.

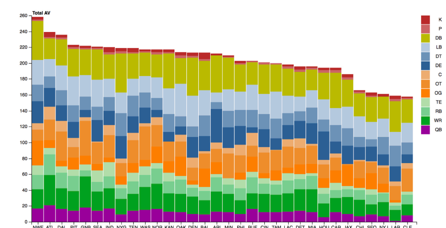


TEAM COMPOSITION

Our team composition analysis was done with rosters since **1978** (the year the NFL switched to a 16 game regular season). The visualization below displays the sum of the AVs of each position for every team in the 2016 season.

- The total AV of each team is **highly correlated** with its success in the 2016 season and playoffs.
- **Balance** in team composition is important; the best-performing teams receive significant contributions from all positions.

Total AV By Team and Position (2016)



COMPOSITION ENGINE

Model Selection - We wanted to select a model that would perform well for our classification task and allow transparency about which components of a team contributed to playoff success. We decided on using sklearn's *RandomForestClassifier* with 2500 trees, balanced class weights and use of out-of-bag error set to 'True'.

Feature Engineering - We trained our model on summary statistics with respect to each position. These included the mean, sum, and max AV by position as well as the number of players for a given position.

Validation - We validated our models for each playoff milestone using stratified 10-fold cross validation.

Results - For all playoff milestones the maximum AV and sum of AVs for DBs and QBs were found to be the most important components of a successful team. Among Super Bowl champions, the sum of DB AVs was the most important feature by a margin of 160% over a team's total QB AV. It is trailed in importance by the max DB AV and the sum of LB AVs, suggesting that **defense does indeed win championships**.

Trading Tony Romo to the Houston Texans:

Wildcard Team Prediction: **True** (Pr=0.5524)
Divisional Team Prediction: **False** (Pr=0.5696)
Conference Team Prediction: **False** (Pr=0.8892)
Super Bowl Team Prediction: **False** (Pr=0.9712)
Championship Team Prediction: **False** (Pr=0.982)