Predicting NFL Contracts Based on Player Statistics

Jack Arbuckle, Ian Pezzella, Ben Scartz, David Sobek, and Reuben Dayal

Motivation/Problem Statement

This project looks to predict the salaries and contracts of current NFL players by looking at their previous statistics. This project used a linear regression model, lasso regression model, XGBoost model, and decision tree models. These models were then analyzed to determine which model can predict former NFL contracts with the highest accuracy. The clustering model was used to analyze what players have similar statistics and thus similar contracts.

The problem for this project is relevant to all football players and those who run NFL organizations. Being able to predict contracts and how contracts of former players have turned out is important for teams that are investing a lot of money into the players themselves. It is also important for current NFL players who are attempting to assess their worth. A player becoming a free agent can look at what other contracts players around the league with similar metrics received.

Although the data focuses on offensive skill players (Quarterbacks, Wide Receivers, Tight Ends) we hope that this model can help the majority of players and NFL executives in their contract negotiations.

Many websites and data companies such as Spotrac have come up with contract predictions based on market conditions and other factors. While many of these estimates are very detailed outputs on contract length and amount, they fail to provide exact details on how they obtained these values. They also fail to display how accurate their current and previous predictions have been.

It is clear from the research that this is a very interesting problem that many people have attempted to solve for not only NFL players but for all professional sports.

Problem Framing

While the issue of player value and an organization deciding what a player should get paid is not new, our analysis does have many real-world implications. Many players are often considered under or overvalued for a variety of reasons, which can lead to pressure from the

media and fanbases, salary cap issues, or potential front-office layoffs for making poor contract decisions. We are trying to understand which on-field attributes weigh the most in terms of contract evaluation. We are looking to see if certain game statistics matter more than others when determining a contract for skill positions. Do teams care more about receptions or receiving yards? Completions or passer rating? Total rushing yards or yards per carry? These questions are pondered by NFL executives and fans alike when looking at player contracts, and our goal is to accurately assess player value by seeing what on-field metrics carry the most weight within offensive NFL contracts.

Data Overview

To tackle this problem, we are using the package "nflfastr" to scrape play-by-play data from NFL games and Spotrac's website to extract NFL contract information. One of the challenges we faced in compiling our data was that we had to merge the data sets to get all of the information in one data frame. The play-by-play data included stats such as passing touchdowns, receiving yards after catch, games played, fantasy points, etc. The contract dataset includes stats such as signing bonus, guaranteed money, age, and length of contract. We are attempting to use various in-game statistics to predict the numbers for a given player's future contract. Depending on what position is being analyzed, different statistics will have to be used. Since we are looking into offensive skill positions for the scope of our project, we will focus on offensive statistics. For example, if the model is attempting to predict the contract of a quarterback, passing yards, passing touchdowns, QBR, and interceptions would be relevant categories that would be heavily correlated with the amount of money a QB is set to receive. Statistics like age, rushing yards, rushing touchdowns, games played, and yards after catch can predict future contracts for running backs. Receptions, receiving yards, yards after catch, and receiving touchdowns are potential predictors of wide receiver salaries. We took summary statistics to get an initial assessment of the data we are working with. One summary statistic that immediately stood out was the distribution of Average Annual Value (AAV). The median AAV for NFL players, according to this dataset, was $2.6 million. The mean AAV, however, was significantly higher, coming out to be about $6.6 million. This coupled with the fact that the maximum value here was $55 million indicates that this data is heavily skewed to the right. Another interesting summary statistic found in the data was fantasy points scored. Like the AAV variable, fantasy points seemed to be

heavily skewed to the right. The median fantasy points scored was 42.64, while the mean was 71.46. The large difference between the median and mean once again illustrates the skewed nature of the data. We will also need to scale the data because the units for the different variables all have different magnitudes. For example, rushing, passing, and receiving yards have a much wider range and higher max than stats such as fumbles and fumbles lost. The maximum number of yards is 1,653, while the maximum number of fumbles is 10. The data must be scaled accordingly, or else the stats with much larger values will be shown as better predictors. Also, there are some but not a whole lot of NA values that will need to be dealt with. These mainly occur for players who play a position that is not compatible with a certain statistic.

Key Quarterback Metrics

- CPOE
- Passing EPA
- Rushing yards
- Rushing EPA
- Fantasy Points

Key Running Back Metrics

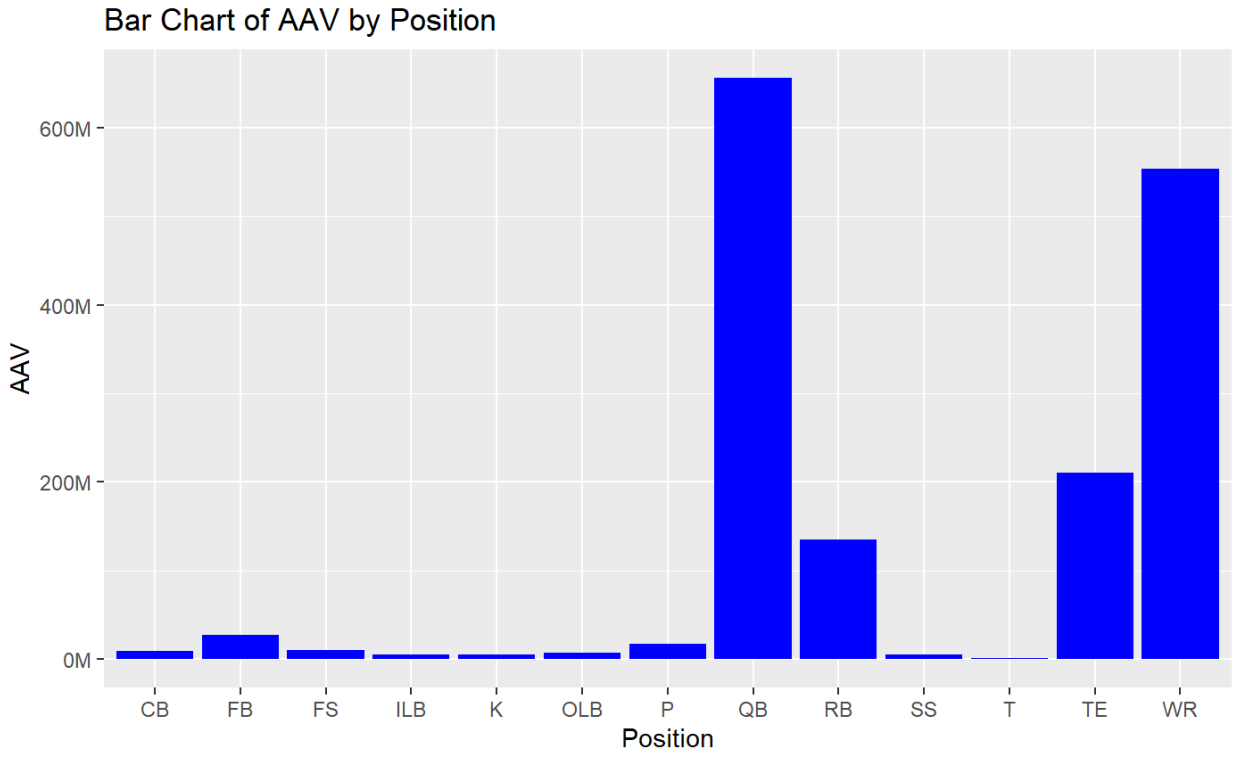
- YAC
- Rushing EPA
- Fumbles Lost
- Fantasy Points
- Touchdowns

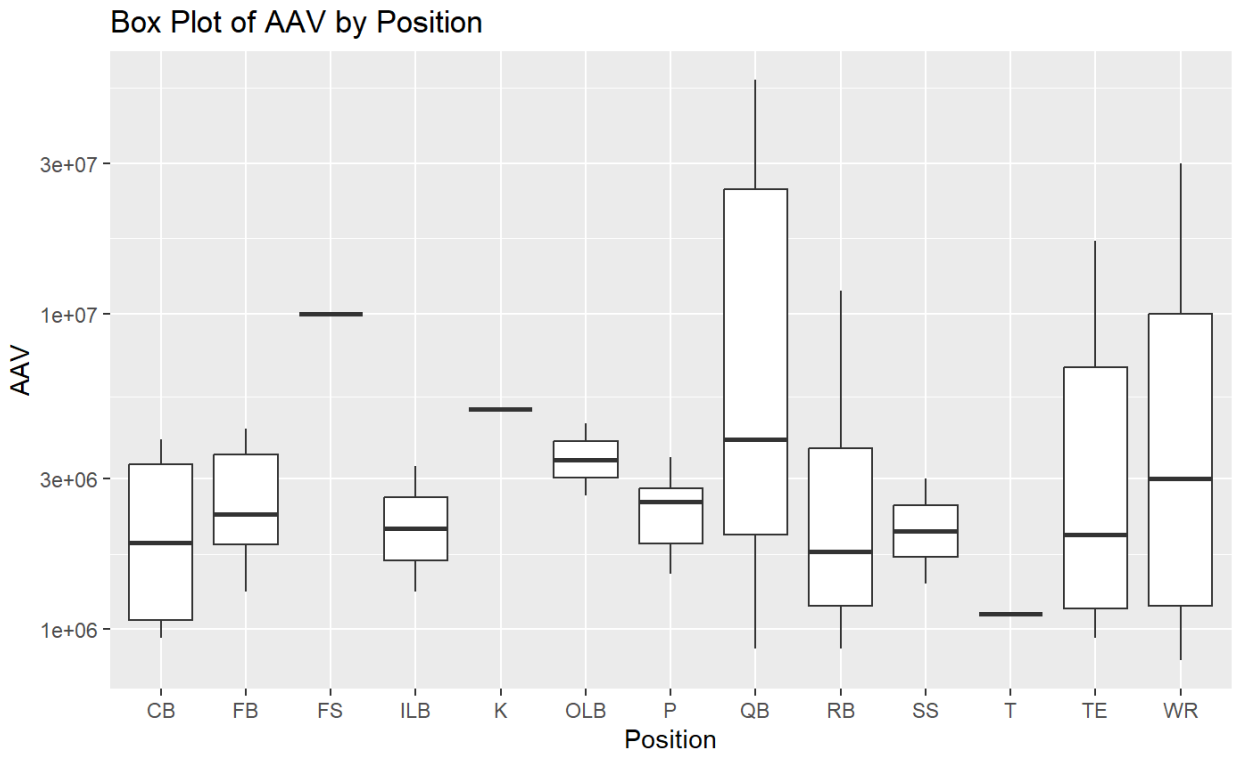
Key Wide Receiver Metrics

- Receiving Yards
- Yards after Catch
- Fantasy Points
- Touchdowns

Data Visualization

Two visualizations have been created for this introductory report. The first is a bar chart showing the average annual value of player contracts by position. From this chart, the positions with the largest annual contract value are all key offensive positions, Quarterback, Wide Receiver, Tight End, and Running Back. Since the project will focus solely on offensive skill positions which have the highest overall AAV, the models created should be able to help NFL players and executives understand what on-field statistics are good predictors of contract values. The second visualization is a box plot of the average annual contract value by position. We can see that along with having the largest average annual contract values, many of the offensive skill positions also have the largest spread of any of the positions, with their highest AAV being millions of dollars away from their lowest. This shows that there is a large amount of variability within these contracts and the overall performance of the players who receive them. Therefore it is important to make sure our models can predict all types of contracts, not just big or small ones.





Contribution

Jack: Problem Statement and Motivation and Secondary R man (20%)

David: Problem Framing and Data Visualizations (20%)

Ian: Problem Framing (20%)

Ben: Main R man (20%)

Reuben: Data Overview (20%)

Bibliography

NFL active player contracts. Spotrac.com. (n.d.). https://www.spotrac.com/nfl/contracts/

An R package to quickly obtain clean and tidy NFL play by Play Data. Dev status. (n.d.). https://www.nflfastr.com/

Summary Statistics

```
                  age             years           value
 Min.   :21.00   Min.   :1.000   Min.   :   800000
 1st Qu.:26.00   1st Qu.:1.000   1st Qu.:  1232500
 Median :27.00   Median :1.000   Median :  4000000
 Mean   :27.68   Mean   :1.862   Mean   : 20394981
 3rd Qu.:29.00   3rd Qu.:2.500   3rd Qu.: 12217000
 Max.   :39.00   Max.   :5.000   Max.   :275000000


      aav             sign_bonus            gtd
 Min.   :  800000   Min.   :   20495   Min.   :    20495
 1st Qu.: 1232500   1st Qu.:  400000   1st Qu.:  1000000
 Median : 2600000   Median : 1880000   Median :  3500000
 Mean   : 6644817   Mean   : 5991972   Mean   : 12578886
 3rd Qu.: 6300000   3rd Qu.: 6000000   3rd Qu.: 10980000
 Max.   :55000000   Max.   :72500000   Max.   :146510000
                    NA's   :68         NA's   :50
    prac_gtd             games         completions
 Min.   :    20495   Min.   : 1.000   Min.   :  0.00
```

```
1st Qu.:  1000000   1st Qu.: 4.000   1st Qu.:  0.00
Median :  3525000   Median :11.000   Median :  0.00
Mean   : 16609230   Mean   : 9.943   Mean   : 27.73
3rd Qu.: 13000000   3rd Qu.:16.000   3rd Qu.:  0.00
Max.   :219010000   Max.   :21.000   Max.   :502.00
NA's   :49
```

```
   attempts      passing_yards     passing_tds     interceptions
Min.   :  0.00   Min.   :   0.0   Min.   : 0.000   Min.   : 0.000
1st Qu.:  0.00   1st Qu.:   0.0   1st Qu.: 0.000   1st Qu.: 0.000
Median :  0.00   Median :   0.0   Median : 0.000   Median : 0.000
Mean   : 42.62   Mean   : 310.1   Mean   : 1.955   Mean   : 0.915
3rd Qu.:  0.00   3rd Qu.:   0.0   3rd Qu.: 0.000   3rd Qu.: 0.000
Max.   :742.00   Max.   :6074.0   Max.   :50.000   Max.   :20.000
```

```
    sacks          sack_yards      sack_fumbles
Min.   : 0.000   Min.   :  0.00   Min.   : 0.0000
1st Qu.: 0.000   1st Qu.:  0.00   1st Qu.: 0.0000
Median : 0.000   Median :  0.00   Median : 0.0000
Mean   : 2.968   Mean   : 19.92   Mean   : 0.4291
3rd Qu.: 0.000   3rd Qu.:  0.00   3rd Qu.: 0.0000
Max.   :52.000   Max.   :373.00   Max.   :10.0000
```

```
sack_fumbles_lost passing_air_yards passing_yards_after_catch
Min.   :0.0000    Min.   :  -6      Min.   :   0.0
1st Qu.:0.0000    1st Qu.:   0      1st Qu.:   0.0
```

```
Median :0.0000    Median :   0      Median :   0.0
Mean   :0.1822    Mean   : 339      Mean   : 143.5
3rd Qu.:0.0000    3rd Qu.:   0      3rd Qu.:   0.0
Max.   :5.0000    Max.   :6296      Max.   :2667.0


passing_first_downs  passing_epa        passing_2pt_conversions
Min.   :  0.00      Min.   :-50.34170   Min.   :0.00000
1st Qu.:  0.00      1st Qu.: -4.59880   1st Qu.:0.00000
Median :  0.00      Median :  0.05115   Median :0.00000
Mean   : 14.66      Mean   :  8.91029   Mean   :0.08097
3rd Qu.:  0.00      3rd Qu.:  8.57880   3rd Qu.:0.00000
Max.   :291.00      Max.   :165.58236   Max.   :4.00000
                    NA's   :184
     pacr            dakota            carries
Min.   :0.0000   Min.   :-0.16155   Min.   :  0.00
1st Qu.:0.7148   1st Qu.: 0.02564   1st Qu.:  0.00
Median :0.8937   Median : 0.07418   Median :  2.00
Mean   :0.8514   Mean   : 0.06267   Mean   : 26.61
3rd Qu.:1.0496   3rd Qu.: 0.10392   3rd Qu.: 15.00
Max.   :2.2000   Max.   : 0.27280   Max.   :340.00
NA's   :189      NA's   :205
rushing_yards     rushing_tds     rushing_fumbles
Min.   :  -6.0   Min.   : 0.000   Min.   : 0.0000
1st Qu.:   0.0   1st Qu.: 0.000   1st Qu.: 0.0000
Median :   3.0   Median : 0.000   Median : 0.0000
```

```
Mean   : 115.7   Mean   : 1.053   Mean   : 0.4049
3rd Qu.:  56.5   3rd Qu.: 1.000   3rd Qu.: 0.0000
Max.   :1653.0   Max.   :18.000   Max.   :10.0000

rushing_fumbles_lost rushing_first_downs  rushing_epa
Min.   :0.0000       Min.   : 0.000      Min.   :-49.5066
1st Qu.:0.0000       1st Qu.: 0.000      1st Qu.: -3.6902
Median :0.0000       Median : 0.000      Median : -0.4712
Mean   :0.1336       Mean   : 6.968      Mean   : -0.8736
3rd Qu.:0.0000       3rd Qu.: 4.000      3rd Qu.:  1.3994
Max.   :3.0000       Max.   :93.000      Max.   : 54.1359
                                         NA's   :94
rushing_2pt_conversions   receptions        targets
Min.   :0.00000         Min.   :  0.00   Min.   :  0.00
1st Qu.:0.00000         1st Qu.:  0.00   1st Qu.:  1.00
Median :0.00000         Median : 11.00   Median : 16.00
Mean   :0.03239         Mean   : 23.34   Mean   : 34.24
3rd Qu.:0.00000         3rd Qu.: 38.00   3rd Qu.: 58.50
Max.   :3.00000         Max.   :178.00   Max.   :233.00

receiving_yards  receiving_tds   receiving_fumbles
Min.   : -10.0   Min.   : 0.00   Min.   :0.0000
1st Qu.:   0.0   1st Qu.: 0.00   1st Qu.:0.0000
Median :  99.0   Median : 0.00   Median :0.0000
Mean   : 255.7   Mean   : 1.66   Mean   :0.2591
```

```
3rd Qu.: 416.0   3rd Qu.: 2.00   3rd Qu.:0.0000
Max.   :2425.0   Max.   :22.00   Max.   :4.0000

receiving_fumbles_lost receiving_air_yards
Min.   :0.000          Min.   : -63.0
1st Qu.:0.000          1st Qu.:   0.0
Median :0.000          Median :  38.0
Mean   :0.166          Mean   : 266.1
3rd Qu.:0.000          3rd Qu.: 440.0
Max.   :2.000          Max.   :2017.0

receiving_yards_after_catch receiving_first_downs receiving_epa
Min.   :  -2.0              Min.   :  0.0         Min.   :-15.7796
1st Qu.:   0.0              1st Qu.:  0.0         1st Qu.: -0.9957
Median :  55.0              Median :  4.0         Median :  3.5393
Mean   : 120.6              Mean   : 12.4         Mean   :  8.9250
3rd Qu.: 178.0              3rd Qu.: 18.5         3rd Qu.: 13.2688
Max.   :1052.0              Max.   :108.0         Max.   :138.2328
                                                  NA's   :60
receiving_2pt_conversions      racr            target_share
Min.   :0.00000           Min.   :-322.0000   Min.   :0.02372
1st Qu.:0.00000           1st Qu.:   0.6082   1st Qu.:0.05572
Median :0.00000           Median :   0.8985   Median :0.08920
Mean   :0.07287           Mean   :  -0.0349   Mean   :0.10896
3rd Qu.:0.00000           3rd Qu.:   1.5109   3rd Qu.:0.14692
```

```
Max.   :2.00000           Max.   :  79.5000   Max.   :0.32254
                          NA's   :60          NA's   :60
air_yards_share         wopr         special_teams_tds
Min.   :-0.04372   Min.   :0.01779   Min.   :0.00000
1st Qu.: 0.01305   1st Qu.:0.09457   1st Qu.:0.00000
Median : 0.05424   Median :0.17520   Median :0.00000
Mean   : 0.10467   Mean   :0.23671   Mean   :0.04049
3rd Qu.: 0.17831   3rd Qu.:0.34746   3rd Qu.:0.00000
Max.   : 0.46675   Max.   :0.75251   Max.   :2.00000
NA's   :60         NA's   :60
fantasy_points   fantasy_points_ppr
Min.   : -2.88   Min.   : -2.88
1st Qu.:  6.00   1st Qu.: 10.00
Median : 42.64   Median : 57.68
Mean   : 71.46   Mean   : 94.80
3rd Qu.: 97.08   3rd Qu.:148.60
Max.   :457.50   Max.   :554.80
```