SIS Football Analytics Challenge

Entry for Zach Feldman, Michael Egle, and Anthony Reinhard

The Problem At Hand

SIS posed a series of questions regarding value of defensive line positions. Which one is the most valuable? What's the talent distribution like? How can it change in different situations?

Data for players on the defensive line is both limited in its history and rarely made available to the public. The three of us were very excited to dive in and learn more about the potential impact different positions can make in different situations.

Defining the Positions

DL Technique Correlation Matrix

% of Time "Player Technique" has a Teammate at "Teammate Technique"

Teammate Technique \ Player Technique	0	1	2i	2	3	4i	4	5	7	6	9
Opposite 9		17%	30%	26%	20%	30%	25%	15%	21%	25%	45%
Opposite 6		16%	15%	8%	6%	5%	8%	10%	16%	16%	9%
Opposite 7		12%	10%	8%	12%	6%	4%	10%	23%	19%	8%
Opposite 5		8%	5%	9%	15%	6%	8%	22%	12%	13%	7%
Opposite 4		3%	4%	6%	6%	10%	25%	5%	3%	6%	7%
Opposite 4i		8%	14%	11%	8%	27%	16%	6%	7%	7%	14%
Opposite 3		75%	68%	40%	27%	28%	36%	54%	48%	28%	33%
Opposite 2		3%	4%	22%	5%	5%	5%	4%	5%	5%	6%
Opposite 2i		2%	7%	10%	22%	16%	7%	6%	13%	23%	16%
Opposite 1		4%	3%	11%	37%	14%	9%	15%	25%	37%	14%
Same 0	0%	0%	0%	2%	13%	18%	25%	12%	6%	5%	9%
Same 1	0%	0%	0%	1%	3%	16%	33%	44%	27%	12%	12%
Same 2i	0%	0%	0%	0%	0%	4%	5%	14%	17%	14%	16%
Same 2	1%	0%	0%	0%	0%	0%	3%	4%	6%	6%	6%
Same 3	35%	6%	1%	0%	0%	0%	1%	20%	41%	60%	36%
Same 4i	14%	9%	4%	1%	0%	0%	0%	1%	4%	6%	15%
Same 4	12%	12%	3%	3%	0%	0%	0%	0%	1%	2%	6%
Same 5	9%	24%	11%	9%	5%	0%	0%	0%	0%	0%	2%
Same 7	4%	13%	13%	10%	10%	3%	1%	0%	0%	0%	0%
Same 6	3%	5%	9%	10%	13%	4%	2%	0%	0%	0%	0%
Same 9	14%	14%	29%	26%	21%	31%	22%	4%	0%	0%	0%

Our first goal was to better understand the responsibilities of the players based on technique and how we could group players together. We started with only down-linemen and made this matrix to help visualize what other techniques - split by side of the ball - might be lined up on a play given the player’s technique and side.

Positional Cutoffs

We ultimately decided on the following classification system:

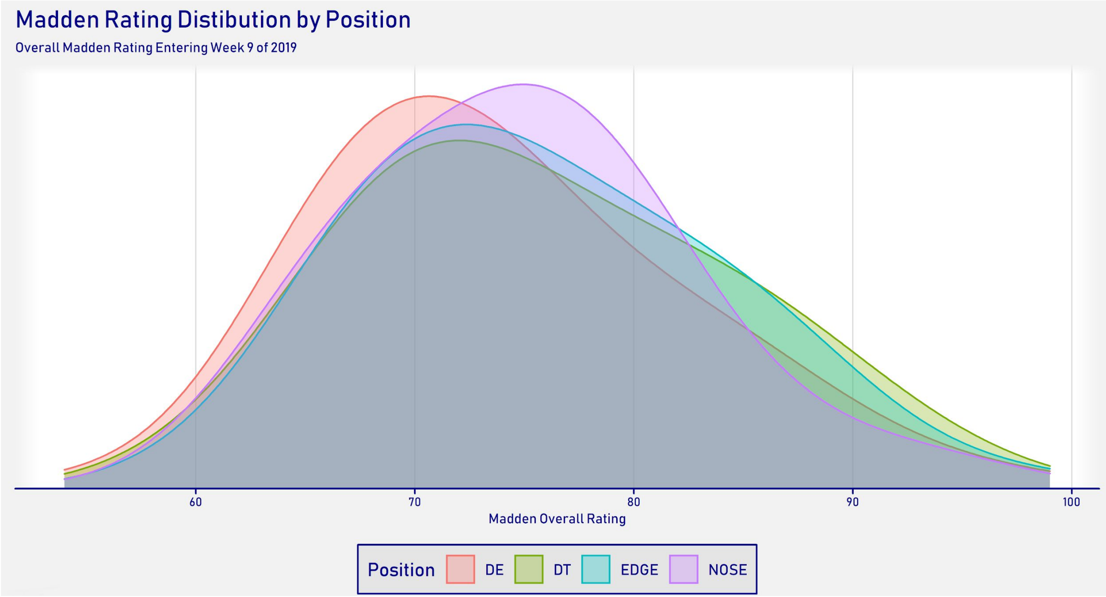
- EDGE - Player without a hand in the dirt (Technique Name = “Outside”)
- NOSE - Player at the 0 or 1 technique that does not have a teammate inside a three technique on either side.
- DT - Player inside a four technique who is not a NOSE
- DE - Any player with a hand in the dirt who does not meet the previous criteria

The Most Valuable Positions - Prior

Now that we had established some positional groups, we thought about which of these positions are most valuable. Here are the rankings we settled on:

1. EDGE - Can affect the passing game as a pass rusher, but could also drop back into coverage
2. NOSE - This player takes up a lot of space usually as the middle defender in a 3-4 scheme.
3. DE - Usually strong pass rushers, but may not provide as much support in the run game
4. DT - Not responsible for as much space and generally are not as strong pass rushers

Talent Distribution - Prior



We thought it would be interesting to view the talent distribution in terms of Madden Ratings. The NOSE position is the most unique as it shows a high floor, but low ceiling. This tracks with our prior that these players need to be good enough to take up a lot of space, but aren't necessarily the players who impact the game the most.

Situational Differences - Prior

While brainstorming how positions could gain or lose value based on specific situations we came up with rushing the passer and short yardage/goalline situations. Our belief is that EDGE and DE will be better at rushing the passer and Nose will gain the most value in short yardage situations.

Models

Individual EPA Impact: Since not all players contribute the same to a given play, we wanted to quantify a player's contribution. We used a generalized additive model to determine the impact a player has on a play's EPA depending on what they contributed to the play. Two separate models were used for this:

- Passing Model: includes pressure, sack, interception, forced/recovered fumble, pass breakups.
- Rushing Model: includes unforced/forced/recovered fumble, solo/assisted tackle, whether the runner used the designed gap, and if the defender filled the gap the runner used.

Models

Next we used a Mixed Effects Model to determine Individual EPA based on Position as a way to determine which is most valuable in terms of Individual EPA. The formula is simply IndividualEPA ~ (1|Position). To the right is the model's results. As you can see, there are slight yet distinct variations among the positions.

```
Linear mixed model fit by REML ['lmerMod']
Formula: IndividualEPA ~ (1 | UpdatedPosition)
   Data: sis_df

REML criterion at convergence: 25385.4

Scaled residuals:
     Min       1Q   Median       3Q      Max
-25.9344   0.1639   0.2006   0.2323   0.3195

Random effects:
 Groups          Name        Variance  Std.Dev.
 UpdatedPosition (Intercept) 3.819e-05 0.00618
 Residual                    8.248e-02 0.28720
Number of obs: 74026, groups:  UpdatedPosition, 4

Fixed effects:
             Estimate Std. Error t value
(Intercept) -0.030348   0.003296  -9.208
> ranef(sis_model)
$UpdatedPosition
      (Intercept)
DE   -0.003357180
DT    0.003993758
EDGE -0.006367219
NOSE  0.005730640
```

Situational Changes - Rushing the Passer

- Getting sacks is the most valuable part of a pass play for the defense outside of turnovers, which are effectively random
- It is widely accepted in football analytics that pressures are a better stat than sacks for analysis due to the random nature of turning a pressure into a sack
- This played out in our analysis after first creating a similar model for getting sack, but the model having very poor performance
- The only variable that stuck out was if the player got a pressure on the play
- That led us to modeling only for generating a pressure

Model - Generating a Pressure

Generating Pressure: A logistic mixed model to determine the factors most important for an individual player to generate a pressure on a given pass play. The formula for the model is Pressure ~ IsRushing + pass_down + ovr_rating + (1|pos_grp), where

- Pressure indicates if the player got a pressure
- IsRushing indicates if the player was rushing the passer
- pass_down indicates if the down and distance was a typical passing down
- ovr_rating as the scaled madden rating for the player
- (1|pos_grp) a random intercept for the position of the player

Model Summary

From our model output we can see that the most obvious impact to generating a pressure is actually rushing the passer with a change in log-odds of 2.03.

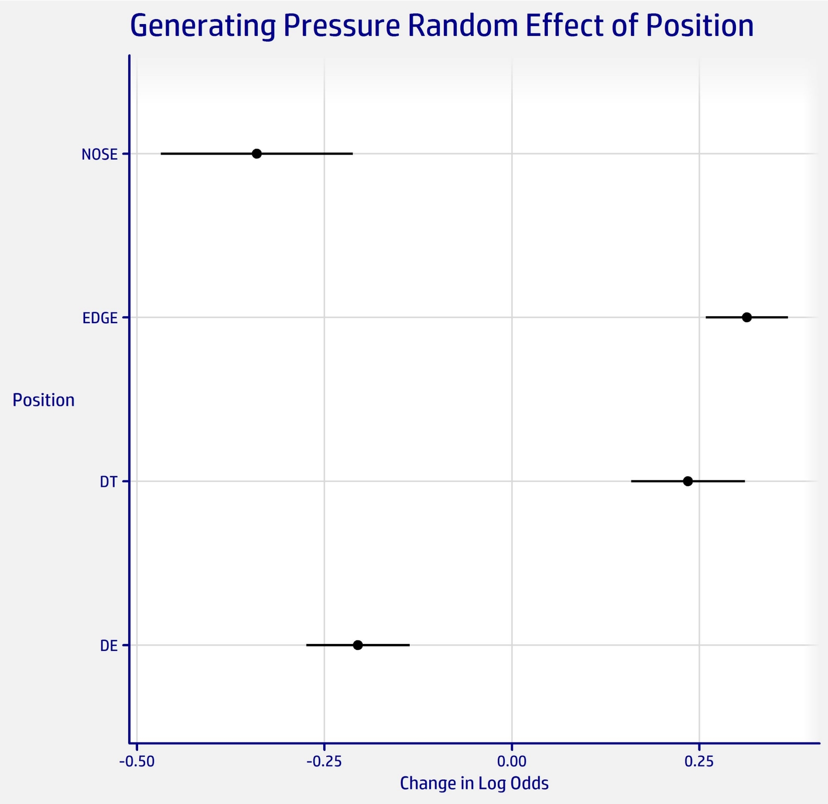
After that, whether or not it is a passing down had a very low effect relative to other factors. The ovr_rating estimate will be multiplied by the actual ovr_rating which has range [0,1], a median of .466, and mean of .501.

```
Generalized linear mixed model fit by maximum likelihood (Laplace Approximation) ['glmerMod']
 Family: binomial  ( logit )
Formula: Pressure ~ IsRushing + pass_down + ovr_rating + (1 | pos_grp)
   Data: .

     AIC      BIC   logLik deviance df.resid
 21737.6  21780.5 -10863.8  21727.6    38908

Scaled residuals:
    Min      1Q  Median      3Q     Max
-0.4538 -0.3391 -0.2893 -0.2421 12.1976

Random effects:
 Groups  Name        Variance Std.Dev.
 pos_grp (Intercept) 0.07998  0.2828
Number of obs: 38913, groups:  pos_grp, 4

Fixed effects:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -5.01853    0.21741 -23.083  < 2e-16 ***
IsRushing    2.03170    0.15734  12.913  < 2e-16 ***
pass_down    0.18849    0.03755   5.020 5.18e-07 ***
ovr_rating   0.90485    0.09924   9.118  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:
           (Intr) IsRshn pss_dw
IsRushing  -0.708
pass_down  -0.084 -0.008
ovr_rating -0.220 -0.027 -0.039
```

Random Effects of Position

We can see from the random intercepts from position that there is a significant difference between positions in the change in log-odds for generating a pressure.

EDGE and DT separate themselves from DE and NOSE, even after accounting for the skill of the player (as defined by Madden rating).



Question 1: Most Valuable D-Line Position?

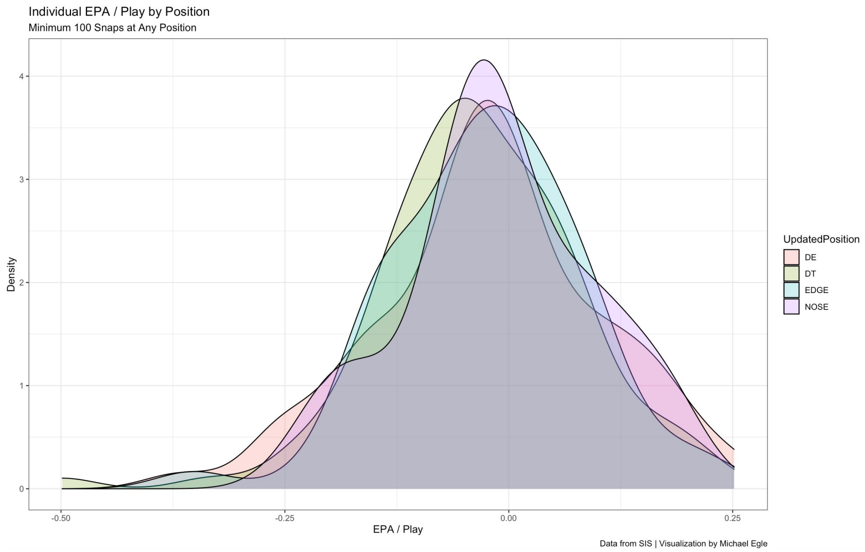
Based on our analysis, we determined that the most valuable defensive line positions (per our definitions) are as follows:

1. EDGE
2. DE
3. DT
4. NOSE

Question 2: Talent Distribution Among Positions?

In some contrast to the distribution of Madden ratings shown in our priors, we have a new distribution of EPA / Play for any player at any position they lined up at for at least 100 snaps.

Note: Unlike the earlier Madden ratings distribution, more talented players in terms of EPA / Play would appear further to the left.



Question 3: Different Situations?

From our two prior situations, only one ended up making a difference in how to view the positions. Short yardage situations didn't change the ordering of positions, it only tightened up the range between them.

We saw a more substantial difference in generating pressure on the quarterback. Through the random effects from the logistic mixed model, we can see that there is a gap between EDGE/DT and DE/Nose.

While EDGE remains the top position, we see that DT jumps DE when we focus on generating pressure in the passing game.

Limitations and Future Analysis

One of the biggest challenges to overcome in this was the small sample size of the data. It's difficult to draw groundbreaking conclusions from only half a season's worth of data that only looks at defensive lineman. In the future, we believe it would be worth a look into whether our conclusions are the same in past seasons.

Furthermore, it would be interesting to do similar analysis with NFL tracking data of all players. Knowing how the linebackers, defensive backs, and offense are aligned and move throughout the play could add a lot more context to how the defensive lineman performed.