

Syracuse University's

Football Analytics Blitz

Powered by SIS



Evaluating Defensive Lineman Through Advanced Analytics

Case Written by Mitch Berezney and Owen Trainor (Syracuse University '23)

Case Prepared by Zak Koeppel (Syracuse University '21), Ben Ayers (Syracuse University '22),

Dr. Jeremy Losak (Syracuse University), and Matt Manocherian (Sports Info Solutions)

Third Annual Event - February 24th, 2023

Introduction

If there is a defensive equivalent to having a superstar quarterback, it has to be having an unblockable defensive lineman. As teams continue to improve their blocking schemes to try and negate these disruptors, evaluating their true impact on the game has become an increasingly tough challenge. Using advanced measures from Sports Info Solutions as well as salary data from nflverse and OverTheCap, your goal will be to explore what makes a good defensive lineman and build your optimal defensive line.

This year's Super Bowl is not just a clash of two high-powered offenses led by excellent young quarterbacks. It features the two teams at the top of the league in sacks: the Eagles in first with 70 and the Chiefs in second with 55. The Eagles not only set the single-season sack record, but they were the first team in league history to have four players with 10 or more sacks. It is clear that building a strong defensive line has been a key to success in recent years.

Your evaluations should be determined by utilizing the data provided. A more in-depth breakdown of what is contained within these data is provided later in this prompt. The requirements of this are detailed below as well.

Your analysis should address the following:

- **What are the traits that make a good defensive lineman?**
- **How does a team's defensive scheme (3-4, 4-3) change your analysis?**
- **What are the traits that make a good defensive line overall?**
- **Under salary constraints and based on your analysis, create an optimal defensive line unit.**

Data

These data contain everything necessary to satisfy the requirements of this prompt, however you are also welcomed to supplement your analysis using data from any **public** source. Make sure to acknowledge these sources during your presentation. Any raw data given to you for the competition are **NOT TO BE SHARED OR PUBLISHED** on any public platform and are provided solely for the sake of this competition. The raw data **MUST** be deleted after the completion of the competition, although you are allowed to share/publish code or presentation materials.

The dataset used for the challenge combines both play-level information ("*pbp.csv*") and player-level information ("*total_points.csv*") from the 2022 season into two files. Let's talk about the Play-by-Play file first.

The play-level section includes basic play-by-play plus some extra data points like the_Expected Points Added (EPA) of the play. There are also some interesting details about the play that might be relevant to your analysis, with a few data points specific to each of pass or run plays.

The player-level section includes the identity of all the up-front defenders on the play, their positions as named on the roster, their alignment on that play, and several stats they might have accumulated.

A player is included if he (a) was in a 3-point stance, (b) lined up standing on the edge on the line of scrimmage, or (c) usually lines up as a DL, even if he might have been up or off the ball on this play.

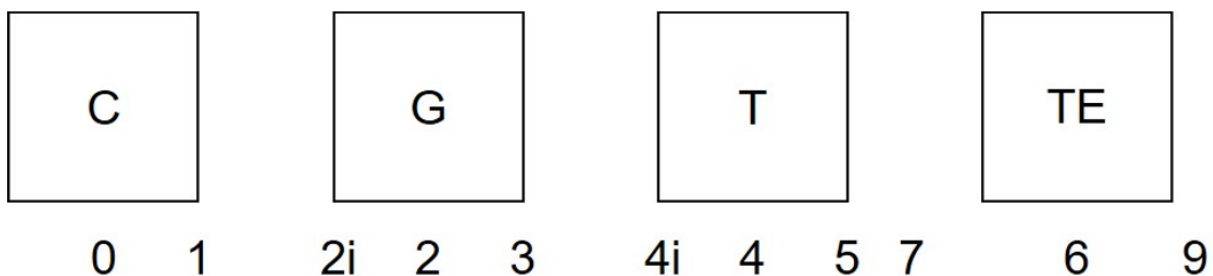
Some of the stats included in the file are given at both a play level and a player level. For example, *InterceptionOnPlay* will tell you if anyone intercepted the pass, and *Interception* will tell you if the specific player referenced in that row intercepted the pass.

Most of the data we're releasing for this challenge is self-explanatory to anyone who has played around with football data before. The defensive alignment info is probably the biggest exception.

For starters, we have what we're calling *RosterPosition* and *OnFieldPosition*. The former is just what we have the player labeled as on the roster. The latter is his position on the given play. In this context, that basically means "did you have your hand on the ground?" If "yes", then you're a defensive lineman. If "no," then you're a linebacker.

The one piece that requires a little more football know-how is the technique (i.e. alignment) of the defenders on each play. In the file it's called *TechniqueName*.

The technique of a defender is encoded using a (mostly) numeric system where your alignment is measured by which offensive player you line up against and on which side of that player you line up. See this image:



Looking at this image, you can see that when people refer to "3-technique" or "5-technique" they're talking about lining up just outside of either the guard or tackle. And the same structure is used for either side of the center, so you might have multiple players with the same technique on a given play, just on different sides of the center. The player's side of the ball is encoded with *SideOfBall*, which is from the defense's perspective.

There's also another data point that isn't quite alignment-related but does convey specific information about what a player was doing on a given play. The *IsRushing* column tells you whether the given player was rushing the passer on designed pass plays. That column will always be zero on designed run plays.

Event Types: Unsurprisingly, there are run plays and pass plays in the data set. The *EventType* column tells you whether the play was a pass or a run—not by design, just in result. So, a scramble would be counted as a run play for this purpose.

Personnel: We've included some columns that tell you how many running backs (RB), tight ends (TE), defensive linemen (DL), and linebackers (LB) were on the field for the play. This helps understand the goals of the offense and defense on a given play, and what you can expect from individuals who were on the field for a given play.

Run Plays: We have included *RunDirection* and *UsedDesignedGap* to help you analyze run plays based on where the play was designed to go and whether the offense succeeded in running that direction.

The run directions are gap-based using the A-B-C-D naming convention (moving from inside to outside). A run to the left B gap, for example, was intended to go between the guard and the tackle on the left side.

If a run was intended to go between the right guard and the center and the rusher bounced the run outside the tackles, *RunDirection* would be "Right A Gap" and *UsedDesignedGap* would be set to 0.

We've also included Blocking Scheme information, which is particularly useful for distinguishing between zone and gap blocking on runs. In short, zone blocking involves offensive linemen blocking an area, while gap blocking involves linemen blocking specific people.

Pass Plays: In addition to basic information like whether the pass was completed or intercepted, we have also included the air yards on the throw (*ThrowDepth*). At both the play level and player level, we've included information about *Pressure* (hits, hurries, knockdowns, sacks) and *PassBreakup* (defensed, batted, deflected, or intercepted passes).

We have also included timing data for a few different events, in seconds. We have the time until the throw, the time until the sack, the time until the first pressure, and the time until the scramble (for each of these that are relevant for the play). And we've identified which defensive player the snap-to-pressure and snap-to-sack times were computed for, which can be useful for analyzing individual players.

Penalties: While we understand that the value of defensive players can be affected by their ability to draw offensive penalties (or commit penalties themselves), we decided that we would remove all plays with an accepted penalty from the data. There is enough gray area in how one should approach analyzing plays with penalties that it was decided to remove them from the picture.

We've also included a **Total Points** file, with one row per player per team from the 2022 season. **Total Points** is SIS's attempt at estimating the value of each player's contributions in total. You can think of it as the number of points on the scoreboard the player was worth. For the purpose of this project, use Total Points as a way of valuing players, teams, and positions in general. Visit <https://www.sportsinfosolutions.com/2020/12/01/a-primer-on-total-points/> for more information.

For data regarding a player's salary, teams must use "*active_contracts.csv*" and an individual player's APY (Average Per Year). This is the number you will use in determining your defensive line under salary constraints.

As mentioned above, you are not required to only use the data provided, you are highly encouraged to get data from other sources (which would enhance the quality of your analysis). Other potential sources for teams to consider using include pro-football-reference.com, which contains a variety of different player and team stats, nflfastR, which contains detailed play-by-play data, or any other publicly accessible data.

Case Requirements

This case is designed to elicit creative approaches to the prompt. Listed here are the main topic points your presentation should address, but you are encouraged to add additional layers of analysis.

- **Determining traits or attributes that make a good defensive lineman:**
 - Do these attributes differ between lineman in 3-4 vs. 4-3 defenses?
 - How does a run-stuff specialist compare to an elite edge rusher?
 - What traits differ between Defensive Tackles, Defensive Ends, and Edge Rushers?
- **Identifying player personnel and attributes that make a good defensive line overall:**
 - What types of players work well together?
 - What are characteristics of a strong defensive front?
- **Creating your optimal defensive line for the 2023 season within the set salary cap limit of \$40 million:**
 - If you choose a 4-3 defensive formation, you will select **two** defensive tackles, **two** defensive ends, and **two** depth players (one of each position)
 - If you choose a 3-4 defensive formation, you will select **one** defensive tackle, **two** defensive ends, **two** linebackers as edge rushers, and **one** depth player (for position of your choice)
 - Teams must use between **80-100%** of their fixed budget (\$40 million)
 - Teams may select up to **THREE (3)** players on their first contract
 - The upcoming rookie class is **ineligible**

Judges will be focused on your approaches to these points but will also be judging you on the following criteria:

- **Overall Process:**
 - Clear presentation and understanding of the problem being addressed
 - Depth of statistical tools utilized to obtain results
 - Methodological soundness
 - Strength and reasonableness of conclusions and recommendations
- **Creativity:** Elements of personal insight beyond the requirements stated in the case
- **Clarity and Communication:** The organization and succinctness of your presentation.

Being convincing with your results is important in impacting decision-making. Strong visualizations are expected. Presenters should be dynamic in their delivery.

It should be understood that the judges are especially knowledgeable of the game of football and are equally familiar with the questions associated with the prompt. Teams should format their presentations with this in mind to avoid including irrelevant and excessive background information that would cut into your time for presenting your analysis. The presentations are 25 minutes in total, which is comprised of 20 minutes of presentation and 5 minutes for judge Q&A.

This competition is made for undergraduate students to compete against other undergraduate students. As such, teams are prohibited from acquiring assistance from any faculty, family, or friends and students not registered on their team. In addition, teams should not contact anyone associated with an NFL team, as well as any other individual for advice on this case.

Teams are allowed to use the internet to help answer the case questions and develop their analyses but should include personal insights to their presentation. While there is analysis that could be related to this topic online, an over-reliance on outside work will be detrimental to your presentation.

Each team must submit their final presentation to footballanalyticsblitz@gmail.com by 9:00 PM EST on Thursday February 23rd at the latest. Late submissions will be penalized during the judging process.

If you have any questions about this case, please feel free to contact Mitch Berezney at msberezn@syr.edu.