St. Lawrence University

SLU Football Tackling Analytics

A Look at Level 2 Versus Level 3 Tackles

Joshua Larson ’25

LEAD-565-02: Tackling Analytics

Dr. Ivan Ramler

15 December 2025

SLU Football Tackling Analytics: A Look at Level 2 Versus Level 3 Tackles

Over a 13-week period, 543 tackle attempts were made and evaluated by members of the 2025 St. Lawrence University football team. To protect player privacy, individual names are not used in this report. Players are referenced only by their jersey numbers. I want to thank Dr. Ivan Ramler and Dr. LeAnn Holland for their guidance and support throughout this process. Their help made it possible for me to turn an idea into a structured project, and I'm genuinely grateful for the time and thought they invested along the way. This work represents my first real step into football analytics, and it felt especially meaningful to start by studying the program where I played and learned the game.

Abstract

Tackling is the backbone of football. In its early days, the game closely mimicked rugby: running, blocking, and tackling the ball carrier. Over the years, tackling has progressed from a violent, out-of-control collision to a process that involves speed, positioning, and technique, all while trying to maintain that violent nature while keeping yourself safe. In today's age, an efficient tackler must be able to use, train, and incorporate five performance indicators: track, prepare, connect, accelerate, and finish. Richie Gray of USA Football is classified as a "world-renowned and multi-sport tackling expert," where he thrives in teaching the methods of a proper tackle to ensure success, safety, and effectiveness for his players.

Effective tackling sits at the heart of good defensive football, but the way we talk about it is usually rooted in coaching language rather than evidence. Coaches know what a good tackle looks like, yet we rarely test those instincts with data. This study aims to bridge that gap by examining the relationship between contact height and technique and their impact on actual field outcomes. The foundation of this study draws from Robert Currin's existing tackling grading scheme and the variables he originally tracked. That framework was used as a starting point, with minor adjustments made to the grading scheme, as well as introducing additional variables to capture more detail and improve the quality of the data.

Guided by biomechanical principles emphasized in USA Football's shoulder tackling guidelines, tackles are grouped into three levels based on where and how contact is made. The focus

here is on Level 2 and Level 3 tackles. Level 2 tackles are designed to engage the ball carrier near the center of mass, creating balance, leverage, and control. Level 3 tackles involve higher contact points, where leverage is reduced, and the ball carrier's lower body remains more active. Rather than assuming that lower contact is better because coaches say it is, this study tests that belief directly. The goal is to see whether Level 2 tackles consistently outperform Level 3 tackles in measurable ways. The results are intended to provide coaches and players with clearer feedback, reinforce safer tackling habits, and support player development by providing evidence that connects technique to performance.

Data and Schematics

0.1 Operation

Balancing grading and variable tracking on top of the duties of a graduate assistant coach is quite a struggle. A grueling process that requires attention to detail and the replaying of the same clips repeatedly. Repetition is needed to remove bias in watching the clips. Devoting several minutes to a great tackle and a quick pass-by on a miss can heavily fluctuate the score of the tackle, especially when some aspects are better than others. Grading can not start until the film is uploaded. The process starts with uploading the game film from Hudl Sideline to Hudl, where each clip will be trimmed in length and tagged with offense, defense, and special teams. Base columns, which will be elaborated on further, are input, and then the tackle data can be input. Only defensive plays are graded due to the controlled situation of the play. Special teams leave too much open space, and yards gained from sudden change on turnovers do not go towards the team's totals. Only first-chance tackle attempts are graded, meaning a play with 10 missed tackles throughout the play will only have one grade, the initial attempt. Once all tackles for a game are graded, the data is exported into an Excel file and formatted for analysis. This structured dataset serves as the foundation for all subsequent statistical analysis and visualization used in this study.

0.2 Grading Scheme

The core of the tackle is made up of five different variables that rely on speed, position, and leverage. The categories are graded on a 0–4 scale, representing equal percentile increments. This scale was chosen intentionally. Expanding the grading to a 0–10 range would require finer distinctions between scores, such as separating a 2 from a 4 or a 5 from a 7, which introduces additional subjectivity. With a single grader evaluating each rep, the 0–4 scale provides enough resolution to capture meaningful differences in technique while minimizing unnecessary bias.

The first variable is **TRACK**, which accounts for tackle angle, field leverage, and closing speed. Positioning yourself to take a strong angle while limiting the ball carrier's ability to cut back is critical here. Equally important is the defender's awareness of where help is on the field and how that should influence the approach. Next up is **PREPARE**, where the focus is on the defender's base, pad level, head positioning, and closing footwork. Proper tackle preparation involves being in control when you go to initiate contact, which leads us to the next variable, **CONNECT**. Here, the positioning of the body on contact is evaluated, but most importantly, the score will be higher if the prior measures were taken to win at the point of contact. This means driving the ball carrier back or stopping them in their place. Once contact is made, and in order to truly win at the meeting point, the defender must **ACCELERATE**. This is where the defender maintains his momentum by running his feet and driving through his hips on contact. This is why lifts such as the hang and power clean are so beneficial for athletes because the biomechanics directly translate. If all of these scores are optimal enough, it should lead to a well-rounded **FINISH** score. This is a variable that can only be graded on made tackles. If a tackle was not finished, it was missed. A higher score will be determined by the effectiveness of the through motion on the way to the ground. The key here is to wrap and grab the cloth to limit extra yards gained by the ball carrier. Once all aspects are graded, a cumulative score out of 20 is calculated, and the tackle is given a technique score (**TECH_SCORE**). The purpose of calculating a cumulative technique score is to turn a complex, multi-step movement into a single, interpretable number without losing the detail of how the tackle actually happened.

0.3 Base and Tackle Variables

The core of the data would not be manipulable without the collection of the 20 additional variables. As stated earlier, the base columns are filled out during the initial watch of the game film, mainly for the coaches to evaluate the game. These variables are `PLAY #`, `ODK` (offense, defense, kick), `DN` (down), `DIST` (distance to first down), `YARD LN` (field position relative to the 50 yard line: -X on defending side, X on opposing side), `HASH` (left, middle, and right), `GN/LS` (gain/loss), `PERSONNEL` (offensive personnel), `PLAY TYPE` (run/pass), `PLAY RESULT`, and `DEF. PLAY CALL`.

The columns collected for each first chance attempt are `TACKLER` (jersey #), `TACKLE POS` (position), `TACKLE ENTRY` (front, side, and rear), `LEVEL` (3, 2, and 1), `ENGAGED` (being blocked during or right before the tackle attempt), `MAKE/MISS`, `YAFC` (yards allowed after first chance/contact), `ASSISTED` (other players aiding in bringing down the ball carrier), and `FIELD_TYPE` (box, alley, open-field, and sideline).

The `LEVEL` variable is the premise of this study. The level of the tackle is determined by the section of the body where contact is initiated. A level 1 tackle is made below the knees, most commonly seen in ankle or "shoe-lace" tackles. While these tackles can be effective at getting a ball carrier to the ground, many level 1 tackles are the result of broken plays or explosive gains and are less informative for evaluating consistent tackling form or repeatable technique. As a result, level 1 tackles are excluded from the primary analysis. A level 2 tackle is made between the hips and knees and puts the defender in the ball carrier's center of mass. Ideally, it positions the defender to effectively run through the contact and bring the ball carrier down with force. Level 3 tackles are attempts made above the waist. It is an upper-body tackle where wrapping and grabbing cloth is crucial. This is the most common tackle attempt, especially in Division 3. The analysis is to show level 3 tackles are significantly less effective than level 2 tackles in limiting yards after first contact, producing made tackles, and reducing explosive gains.

0.4 Data Preparation

```python
import pandas as pd

# 1) Load
utica = pd.read_excel("../Python_data/UticaTackleData_2025-08-23.xlsx")
moville = pd.read_excel("../Python_data/MorrisvilleTackleData_2025-08-30.xlsx")
norwich = pd.read_excel('../Python_data/NorwichTackleData_2025-09-11.xlsx')
sjf = pd.read_excel('../Python_data/SJFTackleData_2025-09-17.xlsx')
alfred = pd.read_excel('../Python_data/AlfredTackleData_2025-09-20.xlsx')
union = pd.read_excel('../Python_data/UnionTackleData_2025-09-27.xlsx')
hilbert = pd.read_excel('../Python_data/HilbertTackleData_2025-10-04.xlsx')
ithaca = pd.read_excel('../Python_data/IthacaTackleData_2025-10-11.xlsx')
hobart = pd.read_excel('../Python_data/HobartTackleData_2025-10-25.xlsx')
rpi = pd.read_excel('../Python_data/RPITackleData_2025-11-01.xlsx')
buffstate = pd.read_excel('../Python_data/BuffStateTackleData_2025-11-08.xlsx')
rochester = pd.read_excel('../Python_data/RochesterTackleData_2025-11-15.xlsx')
# 2) Tag game and date
utica["GAME"],   utica["GAME_DATE"], utica['WEEK']   = "UT", "2025-08-23", "-1"
moville["GAME"], moville["GAME_DATE"], moville['WEEK'] = "MO", "2025-08-30", '0'
norwich["GAME"], norwich["GAME_DATE"], norwich['WEEK'] = "NR", "2025-09-05", '1'
sjf["GAME"], sjf["GAME_DATE"], sjf['WEEK'] = "SJF", "2025-09-13", '2'
alfred["GAME"], alfred["GAME_DATE"], alfred['WEEK'] = "ALF", "2025-09-20", '3'
union["GAME"], union["GAME_DATE"], union['WEEK'] = "UC", "2025-09-27", '4'
hilbert["GAME"], hilbert["GAME_DATE"], hilbert['WEEK'] = "HC", "2025-10-04", '5'
ithaca["GAME"], ithaca["GAME_DATE"], ithaca['WEEK'] = "IC", "2025-10-11", '6'
hobart["GAME"], hobart["GAME_DATE"], hobart['WEEK'] = "HOB", "2025-10-25", '7'
rpi["GAME"], rpi["GAME_DATE"], rpi['WEEK'] = "RPI", "2025-11-01", '8'
buffstate["GAME"], buffstate["GAME_DATE"], buffstate['WEEK'] = "BSU", "2025-11-08", '9'
rochester["GAME"], rochester["GAME_DATE"], rochester['WEEK'] = "UOR", "2025-11-15", '10'


# 3) Combine
tackle = pd.concat([utica, moville, norwich, sjf, alfred, union, hilbert, ithaca, hobart, rpi, buffstate, rochester], ignore_index=True)

tech_cols = ["TRACK", "PREPARE", "CONNECT", "ACCELERATE", "FINISH"]
tackle["TECH_SCORE"] = tackle[tech_cols].sum(axis=1)
tackle_clean = tackle.loc[tackle['PLAY RESULT'] != 'PENALTY']
tackle = tackle_clean
tackle
✓ [3] 183ms
```

All data analysis was conducted in a Python Jupyter Notebook. All analyses were conducted in Python. `pandas` and `numpy` were used for importing, cleaning, and organizing the tackling datasets. `matplotlib` and `seaborn` were used to create supporting visualizations. `plotnine` was used to generate ggplot-style figures with smoothing for trend interpretation. Statistical tests and confidence intervals were computed using `scipy` and `statsmodels`. Individual game datasets were imported using the `pd.read_excel()` function, with each file representing a single contest from the 2025 season. After loading the data, each game was tagged with a unique game identifier, game date, and week number to preserve chronological order and allow for game-level comparisons across the season. The datasets were then combined into a single master table using `pd.concat()`, creating one unified tackle-level dataset for analysis.

The technique score was calculated by summing the five technique components: **TRACK**, **PREPARE**, **CONNECT**, **ACCELERATE**, and **FINISH**. During initial exploration, several plays with extremely high yards after first contact values, often exceeding 20 yards, had penalties during the play that may have altered the first chance attempt. Tackle attempts where the player was in no position whatsoever to make the tackle were denoted as 'Level 0.' To avoid misrepresenting tackling effectiveness, these penalty and level 0 plays were removed from the dataset, bringing the total number of tackles down from 557 to 542. This filtered dataset, referred to as the initiated tackle dataset (`tackle_clean`), represents only plays where a legitimate tackle attempt occurred and serves as the foundation for all analyses.

Base Analyses

1.1 Tackle Level, MAKE/MISS, and YAFC

The dataset was first summarized by asking the simplest question, what was recorded and how often it occurred. After cleaning, there were 542 first chance tackle attempts in total, and the next step was to break those attempts down by tackle level.

LEVEL	count
3	342
2	145
1	55

LEVEL	MAKE/MISS	Count	Proportion	Mean_YAFC
1	MAKE	34	0.618182	2.000000
	MISS	21	0.381818	9.142857
2	MAKE	104	0.717241	1.221154
	MISS	41	0.282759	5.829268
3	MAKE	196	0.573099	1.673469
	MISS	146	0.426901	7.719178

Next, yards after first contact/chance (YAFC) were summarized across the full sample, which totaled 1853 yards, or about 3.84 per tackle on average. Then the outcomes were separated into made tackles versus misses to anchor everything in results. Overall, there were 334 makes and 208 misses, and the YAFC gap between them was huge, with makes averaging about 1.57 YAFC and misses about 7.49. Once that baseline was set, make percent and YAFC allowed were compared by level, both overall and split by make or miss, to see whether Level 2 consistently shows better control and limits extra yards compared with Level 3. Level 3 Tackles dominate the frequency aspect, while level 1 has the least. Level 2 has enough where we can still find significant statistical differences, which will be explored shortly

1.2 Tackle Entry

A basic component of the tackle besides the section of the body contact is initiated is the positioning of the tackle. Tackle entry explains if the tackle is being made on the front of the ball carrier, the sides, or from the rear. A front entry usually lets you square up and hit through the center of mass, which makes it easier to stop momentum and limit extra yards. A rear entry often turns into a chase and wrap situation where the runner is already moving away, so finishing depends on closing speed and grabbing cloth. Side entry is usually the messiest, because the runner can spin, stiff arm, or bounce off contact, and the defender is more likely to lose leverage and give up yards after contact.

TACKLE ENTRY	Count	Mean_YAFC	Mean_TECH_SCORE
FRONT	181	2.928177	9.176796
REAR	78	3.782051	9.282051
SIDE	283	4.438163	7.904594

TACKLE ENTRY	MAKE/MISS	proportion
FRONT	MAKE	0.629834
	MISS	0.370166
REAR	MAKE	0.756410
	MISS	0.243590
SIDE	MAKE	0.568905
	MISS	0.431095

Tackle entry was treated as an angle of approach variable to capture how leverage and control change at contact. In this dataset, `SIDE` entries occurred most often ($n = 283$) and were also the least efficient, producing the highest mean yards after first contact/chance of 4.44, along with the lowest average technique score, approximately 7.901. They also only result in successful tackles about 56.9 percent of the time. In contrast, `FRONT` entries produced lower mean YAFC of 2.93, as well as higher technique scores and made tackle rate, while `REAR` entries showed the highest make rate at 75.64%. Overall, entry angle appears to matter both for finishing the tackle and for limiting extra yards, with `SIDE` tackles representing the most difficult profile in the sample.

1.3 Top Tacklers

While the point of this work is to help the defense as a whole, it would be missing something if we did not also recognize individual performance. Knowing who your best first chance tacklers are is not just a nice stat, it can actually shape how you call a game in high-stakes moments. Think about a 3rd & 15 late in a drive. If you are playing the sticks, you are inviting the offense to throw underneath or even hand the ball off and hope you miss. In that situation, the entire call is built on one idea, rally and tackle, get off the field. So it matters who is most reliable in space and who consistently limits yards after contact. To be considered for this top 15 tacklers list, the individual must have at least 10 first chance attempts. For ethical and privacy reasons, athletes will not be identified by name in this paper. Instead, tacklers are referenced by jersey number.

TACKLER	TACKLER
15	67
52	44
31	42
20	35
3	33
21	32
48	29
35	28
28	26
47	24
75	20
27	17
94	14
56	12
55	11

TACKLER	TECH_SCORE
52	10.272727
3	10.181818
31	10.000000
47	9.666667
48	9.448276
94	9.357143
21	9.343750
28	9.192308
20	8.885714
15	8.835821
27	8.529412
55	6.363636
75	6.350000
35	5.428571
56	3.750000

TACKLER	MAKE/MISS
94	0.857143
97	0.777778
21	0.750000
47	0.750000
28	0.692308
52	0.681818
3	0.666667
24	0.666667
48	0.655172
31	0.642857
20	0.571429
15	0.567164
27	0.529412
75	0.500000
35	0.500000

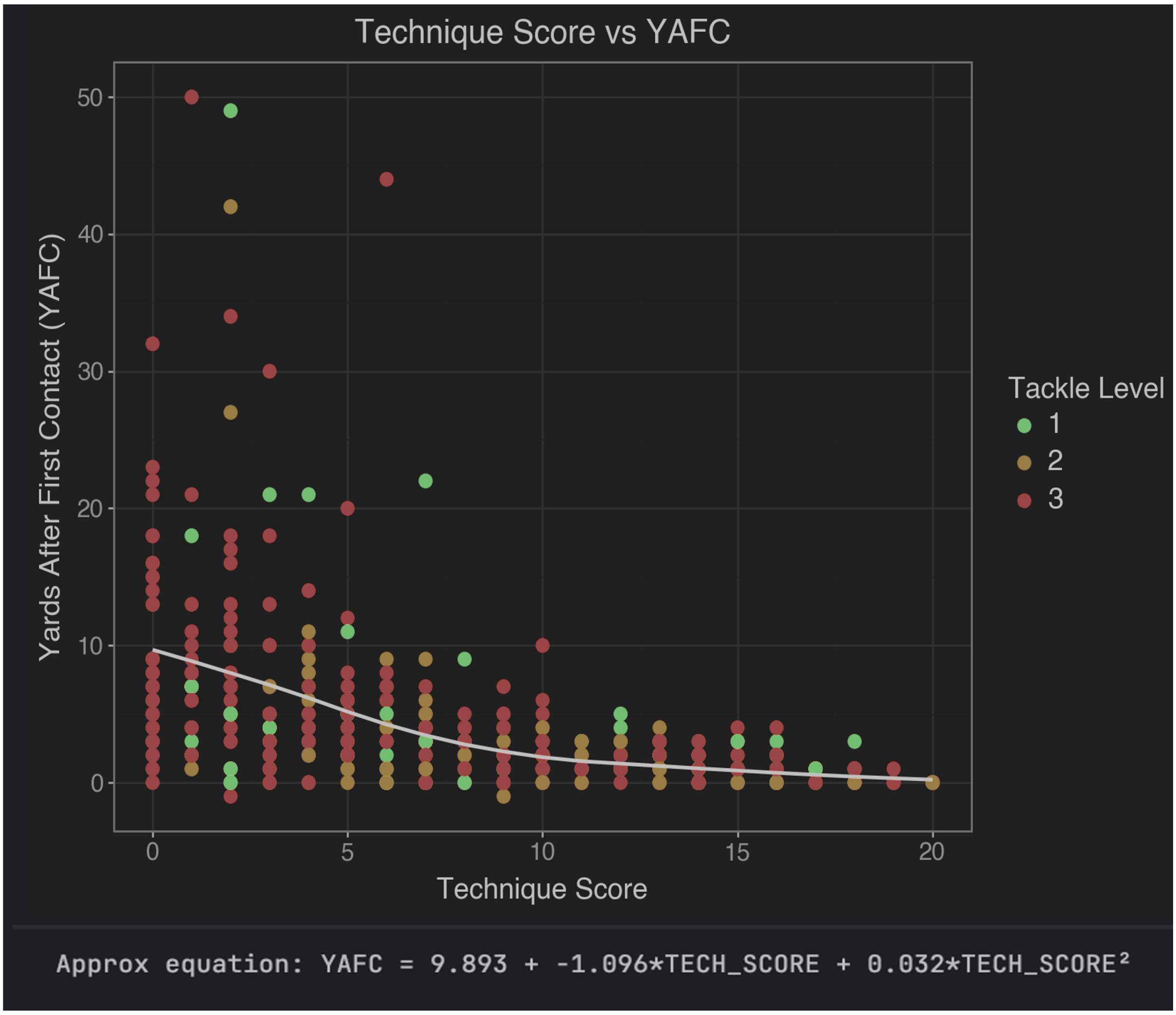
TACKLER	YAFC
94	2.285714
48	2.620690
21	2.812500
52	2.954545
55	3.363636
3	3.393939
20	3.457143
47	3.791667
15	3.880597
31	3.880952
27	3.941176
28	4.346154
75	5.250000
35	5.571429
56	5.916667

Among players with meaningful volume, #15, a linebacker, recorded the most first chance tackle attempts with 67, averaging over 5 first chance attempts per game, including scrimmages. Next is #52, also a linebacker, with 44 and #31, a cornerback, with 42. Looking beyond volume, several players stood out for combining strong technique with efficient outcomes. #52 and #3 posted the highest average technique scores of 10.27 and 10.18, while still keeping average YAFC relatively low (about 3.0 and 3.4). On the efficiency side, #94, a defensive lineman, showed the best profile in this sample, even though his total attempts were lower than the top volume tacklers. His make rate of 85.7% and average YAFC of 2.29 were both team bests. #21, a safety, also paired a strong make rate of 75% with low average YAFC (2.81). In contrast, the tacklers with the highest average YAFC, such as #56 (DL), #35 (DE), and #75 (DL), were also the lowest in average technique score, which reinforces the broader pattern in the dataset that technique quality is closely tied to both finishing tackles and limiting yards after contact.

Visualizations

The goal of the visualizations in this section is to get a clean feel for what the data is actually doing before leaning too hard on formal tests. Tables give averages and percentages, but plots show the shape of the relationship, how spread out the outcomes are, and where the outliers live. That matters in a tackling dataset because one or two explosive plays can distort a mean, and different tackle situations can create clusters that you would never notice in a summary statistic. These figures are meant to show basic trends and give context for why certain modeling choices make sense later on, especially when we start separating performance by level, engagement, success, and technique.

2.1 Technique Score Versus YAFC

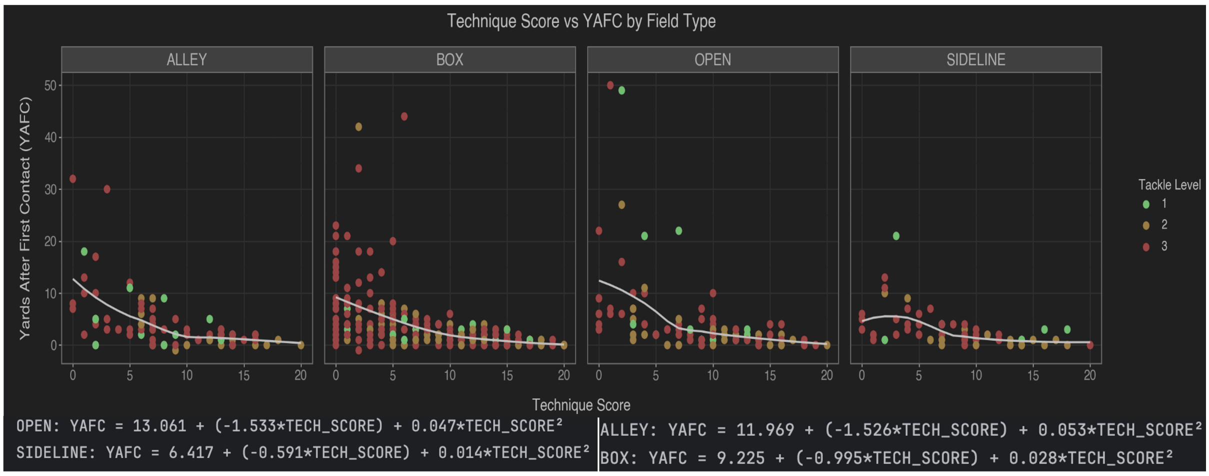


This scatterplot `TECH_SCORE` against yards after first contact and the relationship is pretty clear. As technique score increases, YAFC generally drops. The spread is wide at low technique scores, where you see the biggest chunk of explosive outliers, including several plays above 20 yards and a few approaching 50 yards. As `TECH_SCORE` climbs into the double digits, the point cloud tightens and most tackles cluster close to 0 to 3 YAFC, which is what you would expect if better technique is helping defenders finish cleaner and stop momentum sooner. No tackle exceeded a YAFC of 2 once `TECH_SCORE` reached 17.

To summarize the overall trend with a single equation, a second order polynomial was fit to approximate the smoothed curve:

$$\widehat{\texttt{YAFC}} = 9.893 - 1.096(\texttt{TECH\_SCORE}) + 0.032(\texttt{TECH\_SCORE})^2.$$

The quadratic form captures a strong early decrease in YAFC as technique improves, followed by diminishing returns as `TECH_SCORE` gets higher. The positive squared term means the curve flattens out at the high end rather than dropping forever, which matches what the scatter shows, since there are fewer observations at extreme technique scores and YAFC is already close to zero in that range. A higher order polynomial would start chasing noise and overfitting individual points, while a linear model would miss the clear curvature in the trend. The fitted equation was generated manually because `plotnine` does not print model equations directly.

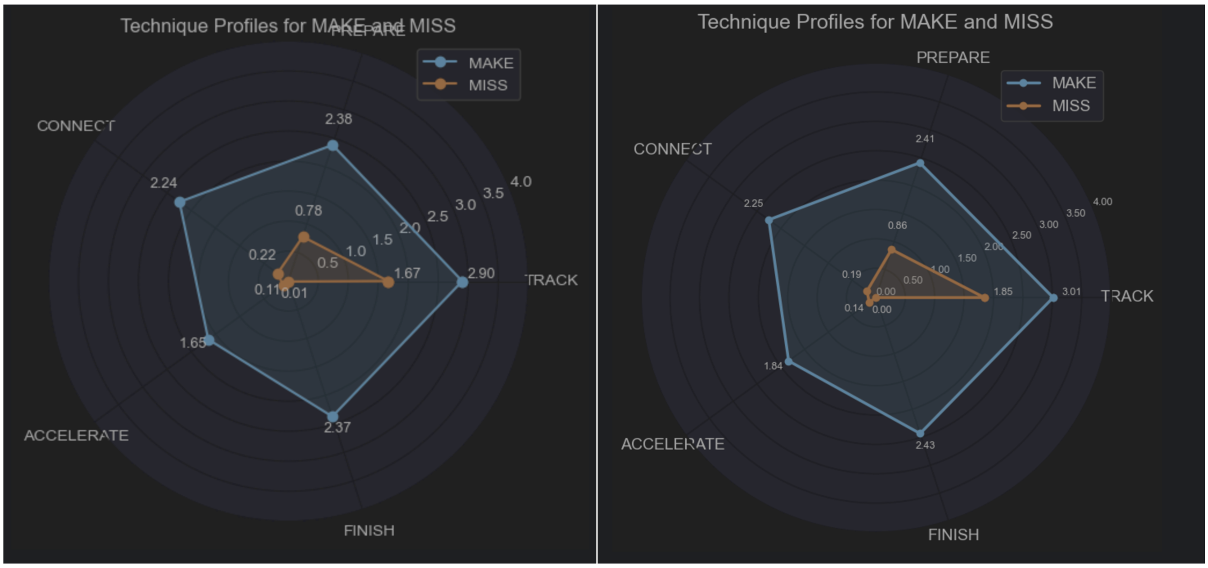


When the `TECH_SCORE` versus `YAFC` relationship is split by `FIELD_TYPE`, the same overall pattern holds, better technique generally means fewer yards after contact, but the context changes how steep that improvement is and where the outliers show up. The `BOX` panel is the noisiest and it is dominated by Level 3 tackles, which is not surprising since most box tackles happen in tighter space with more traffic and more upper body contact. What stood out to me was the composition in the other panels. `OPEN` and `SIDELINE` show a noticeably higher concentration of Level 2 tackles, which fits the idea that when defenders have space to track, close, and square up, they are more likely to get contact at the center of mass instead of reaching high. In terms of the fitted curves, `OPEN` and `ALLEY` show the strongest drop in predicted `YAFC` as technique improves, meaning technique matters most when there is space and the ball carrier has room to create extra yards. `SIDELINE` is flatter and starts lower, which makes sense since the boundary naturally helps cap explosive outcomes. Overall, breaking the plot by field type makes it clear that technique is important everywhere, but the payoff from improving technique is not identical across contexts, and the box is where leverage and crowding make clean tackling hardest to maintain.

2.2 Averages in Technique Profiles

These visuals are meant to give a quick snapshot of what our technique looks like on average, and how that changes depending on outcome and tackle level.

2.2.1 Radar Charts



The radar charts make a simple point that is easy to miss in a table. Made tackles and missed tackles do not just differ by one technique category, they differ by the overall shape of the profile. The left chart shows averages on each of the profiles for made and missed tackles spanning from the first two scrimmages to Week 6. Made tackles show a balanced pattern across `TRACK`, `PREPARE`, `CONNECT`, and `FINISH`, while missed tackles are heavily front loaded.

Misses still show some `TRACK` and a little `PREPARE`, but they fall off sharply once the tackle reaches the contact and completion stages, especially `CONNECT` and `ACCELERATE`. The idea for the `FINISH` profile is that a successful tackle needs a nonzero `FINISH` score. That data was corrected on the final cleaning of the data. On the right is the chart with the last five games are added. The overall premise stays the same, but the made tackle profile improves slightly and becomes a bit more consistent across categories. The biggest visible gain on the make side is `ACCELERATE`, which increases noticeably, and there is also a small uptick in `TRACK` and `FINISH`. Missed tackle profiles also rise slightly in `TRACK` and `PREPARE`, which suggests that more misses later in the season are not always coming from poor approach, but from breakdowns at or after contact.

2.2.2 Bar Charts

These bar charts break the technique profiles out by tackle level, and they help put some numbers behind what we tend to feel on film. Just like the radar charts, these are separated by different windows. The top graphs show the first two scrimmages through Week 6, and the bottom includes all 12 events. In both time windows, Level 2 is consistently the cleanest profile overall, especially in the fundamentals that actually set up and finish a tackle. From the early sample, Level 2 leads Level 3 in `TRACK` (2.77 vs 2.30), `PREPARE` (2.29 vs 1.54), `CONNECT` (1.82 vs 1.33), and `FINISH` (1.70 vs 1.41). In the later sample, that separation still holds, and Level 2 improves in every category, most noticeably in `TRACK` (up to 3.03), `ACCELERATE` (up to 1.33), and `FINISH` (up to 1.82). The interesting shift is in `ACCELERATE`. Early on, it was one of the few areas where Level 3 could stay close, even though it is a category where running your feet through contact should matter. By the end of the season, `ACCELERATE` stops being a category where Level 3 can hang around. It clearly becomes strongest for Level 2 in the full sample, which suggests we got better at running through contact on our better leverage tackles as the season went on. Level 3 remains the lowest in `PREPARE` and `CONNECT` across both windows, which lines up with the idea that higher contact leads to more reaching, less leverage, and more reliance on wrapping and grabbing cloth to get the ball carrier down.

Comprehensive Analyses

3.1 Two-Proportion Z-Tests

Many of the comparisons in this study involve a binary outcome, such as whether a tackle attempt results in a `MAKE` or a `MISS`. When the outcome of interest is binary, such as whether a tackle is made or missed, a natural way to compare groups is through proportions. A two-proportion z-test evaluates whether the difference between two sample proportions is larger than what would be expected from random sampling variation alone. The core assumption is that each observation can be treated as a Bernoulli trial, meaning it has only two possible outcomes, and that within each group the trials are independent with a common underlying success probability.

Let Group A and Group B represent two subsets of tackle attempts being compared. In this study, Group A and Group B correspond to Level 2 and Level 3 tackles, either in the full dataset or within a restricted subset such as engaged or not engaged tackles. Let p_A be the true proportion of successes in Group A and let p_B be the true proportion of successes in Group B.

The hypotheses are

$$H_0 : p_A = p_B$$

$$H_A : p_A \neq p_B.$$

Let x_A and x_B be the observed number of successes, and n_A and n_B be the total number of trials, in Group A and Group B respectively. The sample proportions are

$$\hat{p}_A = \frac{x_A}{n_A}, \qquad \hat{p}_B = \frac{x_B}{n_B}.$$

Under H_0, the pooled estimate of the common success proportion is

$$\hat{p} = \frac{x_A + x_B}{n_A + n_B}.$$

The two-proportion z-test statistic is

$$z = \frac{\hat{p}_A - \hat{p}_B}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_A} + \frac{1}{n_B}\right)}}.$$

For a two-sided test, the p-value is computed as

$$p = 2\left(1 - \Phi\left(|z|\right)\right),$$

where $\Phi(|z|)$ is the standard normal cumulative distribution function. The test output is reported using the z-statistic and its associated p-value, and statistical significance is evaluated relative to a chosen level α. In addition to hypothesis testing, a 95% confidence interval can be reported for the difference in success proportions, $p_A - p_B$. A common large-sample form is

$$(\hat{p}_A - \hat{p}_B) \pm z_{0.975}\sqrt{\frac{\hat{p}_A(1-\hat{p}_A)}{n_A} + \frac{\hat{p}_B(1-\hat{p}_B)}{n_B}},$$

where $z_{0.975}$ is the 97.5th percentile of the standard normal distribution. This interval provides a range of plausible values for the true difference in success rates and is interpreted relative to zero. If the interval does not include 0, it indicates a statistically significant difference at the 0.05 level for a two-sided comparison. Luckily, Python has tools do all of this with simple inputs and outputs, which saves the headache of tedious math.

3.1.1 Difference in Level 2 Versus Level 3 MAKE Percentage

Let p_2 denote the true probability that a Level 2 tackle attempt is made, and let p_3 denote the true probability that a Level 3 tackle attempt is made. To test whether tackle make probability differs by contact level, the hypotheses are defined as

$$H_0 : p_2 = p_3 \qquad H_A : p_2 \neq p_3.$$

Let x_2 and x_3 be the number of made tackles and let n_2 and n_3 be the total tackle attempts for Level 2 and Level 3, respectively. The sample make rates are $\hat{p}_2 = x_2/n_2$ and $\hat{p}_3 = x_3/n_3$.

```
     Level  Made Tackles  Total Tackles  Make %
0  Level 2           104            145  0.7172
1  Level 3           196            342  0.5731
Proportion Difference = 0.1441
95% CI = [0.0540, 0.2343]
Z-stat = 2.9907
P-value = 0.00278353
```

In the observed data, the make rates were 71.72% for Level 2 and 57.31% for Level 3; the confidence interval, test statistic, and p-value corresponding to this comparison are reported in the results section. The estimated difference in make rate is $\hat{p}2 - \hat{p}3 = 0.1441$, which implies Level 2 has a 14.41 percentage point higher make rate than Level 3 in the sample. A 95% confidence interval for the true difference $p_2 - p_3$ is $[0.0540, , 0.2343]$, meaning the data suggest Level 2 is likely between 5.40% and 23.43% higher than Level 3. The two proportion z test produced $z = 2.9907$ with $p = 0.00278$, and since $p < 0.05$, we reject H_0. Therefore, the evidence indicates that, with 95% confidence, the true make probability for Level 2 tackles is higher than for Level 3 tackles in this dataset.

3.1.2 Difference in Level 2 Versus Level 3 MAKE Percentage while ENGAGED

Let p_2 denote the true probability that an *engaged* Level 2 tackle attempt is made, and let p_3 denote the true probability that an *engaged* Level 3 tackle attempt is made. To test whether make probability differs by contact level within engaged tackles, the hypotheses are

$$H_0 : p_2 = p_3 \qquad H_A : p_2 \neq p_3.$$

Let x_2 and x_3 be the number of made tackles and let n_2 and n_3 be the total tackle attempts for engaged Level 2 and engaged Level 3, respectively. The sample make rates are $\hat{p}_2 = x_2/n_2$ and $\hat{p}_3 = x_3/n_3$.

```
Two-proportion z-test: MAKE% (Engaged only)
             Group  Made Tackles  Total Tackles  Make %
0  Engaged Level 2            15             28  0.5357
1  Engaged Level 3            58            115  0.5043
Difference (L2 minus L3) = 0.0314
95% CI for difference = [-0.1747, 0.2375]
Z-stat = 0.2977
P-value = 0.765894
```

In the observed data, the engaged make rates were 53.57% (15/28) for Level 2 and 50.43% (58/115) for Level 3. The estimated difference in make rate is $\hat{p}_2 - \hat{p}_3 = 0.0314$, which corresponds to a 3.14 percentage point higher make rate for engaged Level 2 tackles in the sample. However, the 95% confidence interval for the true difference $p_2 - p_3$ is $[-0.1747,\ 0.2375]$, which includes 0, indicating that the data are consistent with no true difference. The two-proportion z-test produced $z = 0.2977$ with $p = 0.7659$, and since $p > 0.05$, we fail to reject H_0. Therefore, within engaged tackles, there is not sufficient statistical evidence in this dataset to conclude that the true make probability differs between Level 2 and Level 3.

3.1.3 Difference in Level 2 Versus Level 3 MAKE Percentage while NOT ENGAGED

Let p_2 denote the true probability that a *not engaged* Level 2 tackle attempt is made, and let p_3 denote the true probability that a *not engaged* Level 3 tackle attempt is made. To test whether make probability differs by contact level within not engaged tackles, the hypotheses are

$$H_0 : p_2 = p_3 \qquad H_A : p_2 \neq p_3.$$

Let x_2 and x_3 be the number of made tackles and let n_2 and n_3 be the total tackle attempts for not engaged Level 2 and not engaged Level 3, respectively. The sample make rates are $\hat{p}_2 = x_2/n_2$ and $\hat{p}_3 = x_3/n_3$.

```
Two-proportion z-test: MAKE% (Not engaged only)
                 Group  Made Tackles  Total Tackles  Make %
0  Not Engaged Level 2            89            117  0.7607
1  Not Engaged Level 3           138            227  0.6079
Difference (L2 minus L3) = 0.1528
95% CI for difference = [0.0527, 0.2528]
Z-stat = 2.8332
P-value = 0.0046089
```

In the observed data, the not engaged make rates were 76.07% (89/117) for Level 2 and 60.79% (138/227) for Level 3. The estimated difference in make rate is $\hat{p}_2 - \hat{p}_3 = 0.1528$, which corresponds to a 15.28 percentage point higher make rate for not engaged Level 2 tackles in the sample. A 95% confidence interval for the true difference $p_2 - p_3$ is $[0.0527, 0.2528]$, meaning the data suggest Level 2 is likely between 5.27% and 25.28% higher than Level 3 within not engaged tackles. The two-proportion z-test produced $z = 2.8332$ with $p = 0.00461$, and since $p < 0.05$, we reject H_0. Therefore, within not engaged tackles, the evidence shows that, with 95% confidence, the true make probability for Level 2 tackles is higher than for Level 3 tackles.

3.2 Two-Sample T-Tests

Other comparisons in this study focus on outcomes that are numeric rather than binary. For example, yards after first contact is measured on a continuous scale, and the technique score is treated as a quantitative summary of tackling execution. When the goal is to compare the average value of a continuous variable between two groups, a two-sample t-test provides a standard framework. Because the spread of the data and the number of observations can differ between groups,a two-sample t-test is used, since it does not require equal variances.

Let Group 2 represent Level 2 tackles and Group 3 represent Level 3 tackles. Let μ_2 be the true mean of the outcome in Group 2 and let μ_3 be the true mean of the outcome in Group 3. The hypotheses are

$$H_0 : \mu_2 = \mu_3$$

$$H_A : \mu_2 \neq \mu_3.$$

Let $\bar{x}_2$ and $\bar{x}_3$ be the sample means for the two groups, with sample variances s_2^2 and s_3^2, based on sample sizes n_2 and n_3. Welch's test statistic compares the observed difference in sample means to its estimated standard error:

$$t = \frac{\bar{x}_2 - \bar{x}_3}{\sqrt{\frac{s_2^2}{n_2} + \frac{s_3^2}{n_3}}}.$$

Because the variances are not assumed equal, the reference distribution is a t-distribution with degrees of freedom estimated using the Welch–Satterthwaite approximation:

$$\nu \approx \frac{\left(\frac{s_2^2}{n_2} + \frac{s_3^2}{n_3}\right)^2}{\frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2-1} + \frac{\left(\frac{s_3^2}{n_3}\right)^2}{n_3-1}}.$$

For a two-sided comparison, the p-value is computed as

$$p = 2\left(1 - F_{t,\nu}\left(|t|\right)\right),$$

where $F_{t,\nu}(|t|)$ is the cumulative distribution function of a t-distribution with ν degrees of freedom. The test output is reported using the t-statistic and its associated p-value, and statistical significance is evaluated relative to a chosen level α.

In addition to hypothesis testing, a 95% confidence interval can be reported for the mean difference $\mu_2 - \mu_3$. A standard form is

$$(\bar{x}_2 - \bar{x}_3) \pm t_{0.975,\nu}\sqrt{\frac{s_2^2}{n_2} + \frac{s_3^2}{n_3}},$$

where $t_{0.975,\nu}$ is the 97.5th percentile of the t-distribution with ν degrees of freedom. This interval gives a plausible range for the true difference in means and is interpreted relative to 0. If the interval does not include 0, it indicates a statistically significant difference at the 0.05 level for a two-sided comparison.

In this study, the same two-sample framework is applied to two separate mean comparisons between Level 2 and Level 3: the mean YAFC on made tackles, and the mean overall technique score. Python computes the test statistic, p-value, and confidence interval directly from the sample inputs.

3.2.1 Difference in Level 3 Versus Level 2 Average YAFC

Let μ_2 represent the true mean yards after first contact (YAFC) on made Level 2 tackles, and let μ_3 represent the true mean YAFC on made Level 3 tackles. To evaluate whether average YAFC differs by contact level when the tackle is successfully completed, the hypotheses are

$$H_0 : \mu_2 = \mu_3 \qquad H_A : \mu_2 \neq \mu_3.$$

Let $\bar{Y}_2$ and $\bar{Y}_3$ denote the sample mean YAFC values for made Level 2 and made Level 3 tackles, based on sample sizes n_2 and n_3, respectively.

```
Level 2 made tackles: n = 104, mean YAFC = 1.221
Level 3 made tackles: n = 196, mean YAFC = 1.673
Mean difference (L2 minus L3) = -0.452
95% CI for mean difference = [-0.787, -0.118]
Welch t-test: t = -2.664, p-value = 0.00816969
```

Looking at the observed data, Level 2 made tackles had $n_2 = 104$ with a mean YAFC of $\bar{Y}_2 = 1.221$, while Level 3 made tackles had $n_3 = 196$ with a mean YAFC of $\bar{Y}_3 = 1.673$. The estimated mean difference is $\bar{Y}_2 - \bar{Y}_3 = -0.452$, indicating that Level 2 made tackles allowed 0.452 fewer yards after first contact on average in this sample. A 95% confidence interval for the true mean difference $\mu_2 - \mu_3$ is $[-0.787, -0.118]$, which suggests the true difference is likely negative and falls between 0.118 and 0.787 yards in favor of Level 2. The t-test produced $t = -2.664$ with $p = 0.00817$, and since $p < 0.05$, we reject H_0. Hence, the results provide evidence, with 95% confidence, that mean YAFC on made tackles differs between Level 2 and Level 3, with Level 2 allowing fewer yards after contact

3.2.2 Difference in Level 3 Versus Level 2 Technique Score

Let μ_2 denote the true mean overall technique score for Level 2 tackles and let μ_3 denote the true mean `TECH_SCORE` for Level 3 tackles. To test whether average technique score differs by tackle level, the hypotheses are

$$H_0 : \mu_2 = \mu_3 \qquad H_A : \mu_2 \neq \mu_3.$$

Let $\bar{T}_2$ and $\bar{T}_3$ represent the sample mean `TECH_SCORE` values for Level 2 and Level 3 tackles, based on sample sizes n_2 and n_3.

```
Level 2 mean TECH_SCORE = 10.469 (n=145)
Level 3 mean TECH_SCORE = 7.699 (n=342)
Mean difference (L2 minus L3) = 2.770
95% CI for mean difference = [1.759, 3.781]
Welch t-test: t = 5.392, p-value = 1.38511e-07
```

The data shows that Level 2 tackles had $n_2 = 145$ with a mean `TECH_SCORE` of $\bar{T}_2 = 10.469$, while Level 3 tackles had $n_3 = 342$ with a mean `TECH_SCORE` of $\bar{T}_3 = 7.699$. The estimated mean difference is $\bar{T}_2 - \bar{T}_3 = 2.770$, meaning Level 2 tackles scored 2.770% higher on average than Level 3 tackles in this sample. A 95% confidence interval for the true mean difference $\mu_2 - \mu_3$ is $[1.759, 3.781]$, suggesting the true Level 2 advantage in average technique score likely falls between 1.759% and 3.781%. The t-test produced $t = 5.392$ with $p = 1.38511 \times 10^{-7}$, and since $p < 0.05$, we reject H_0. In conclusion, the evidence indicates that, with 95% confidence, the mean technique score differs between Level 2 and Level 3 tackles, with Level 2 higher on average.

3.3 Chi-Square Test of Independence for Tackle Level and Explosive Plays

In the dataset, an explosive play after contact is denoted as gaining at least 3 yards after first chance/contact. To evaluate whether the likelihood of allowing more than 3 yards after first contact depends on tackle level, a chi-square test of independence was used. The outcome variable was converted into a binary indicator, YAFC_{gt3}, where $\text{YAFC}_{gt3} = 1$ if a tackle allowed more than 3 yards after first contact and $\text{YAFC}_{gt3} = 0$ otherwise. A contingency table was then formed by cross-tabulating tackle *Level* with YAFC_{gt3}, producing observed cell counts O_{ij} for each combination of level i and outcome category j.

The hypotheses for the chi-square test of independence are

$$H_0 : \text{Tackle level and YAFC}_{gt3} \text{ are independent}$$

$$H_A : \text{Tackle level and YAFC}_{gt3} \text{ are not independent.}$$

Under H_0, the expected count in each cell is computed from the product of the corresponding row and column totals divided by the grand total,

$$E_{ij} = \frac{(\text{row total})_i(\text{column total})_j}{N}.$$

The chi-square test statistic is

$$\chi^2 = \sum_i \sum_j \frac{(O_{ij} - E_{ij})^2}{E_{ij}},$$

which measures how far the observed counts deviate from what would be expected if level and outcome were unrelated. The degrees of freedom are

$$\text{df} = (r - 1)(c - 1),$$

where r is the number of levels included in the table and c is the number of outcome categories. The p-value is computed from the chi-square distribution with the stated degrees of freedom. The test output is reported using the χ^2 statistic, degrees of freedom, and p-value, and is evaluated relative to a chosen significance level α. Because the chi-square test is global, it indicates whether any association exists across the levels included, but it does not by itself quantify pairwise differences between specific levels. For that reason, a secondary proportion-based comparison can be reported between two specific levels by estimating the difference in their YAFC_{gt3} proportions and constructing a 95% confidence interval for the difference. If the confidence interval for excludes 0, it supports a meaningful difference between those two levels at the 0.05 level for a two-sided comparison.

3.3.1 Test

Let p_2 denote the true proportion of Level 2 tackles that result in $\text{YAFC}_{gt3} = 1$ (meaning YAFC > 3), and let p_3 denote the true proportion of Level 3 tackles that result in $\text{YAFC}_{gt3} = 1$. The goal is to test whether the probability of allowing more than 3 yards after first contact differs by tackle level. The hypotheses can be written as

$$H_0 : p_2 = p_3 \qquad H_A : p_2 \neq p_3.$$

```
YAFC_gt3    0    1
LEVEL
1          38   17
2         117   28
3         216  126
Chi-square = 14.504, p = 0.0007
Level 2 proportion (YAFC>3) = 0.1931 (28/145)
Level 3 proportion (YAFC>3) = 0.3684 (126/342)
Proportion Difference = -0.1753
95% CI for difference = [-0.2574, -0.0932]
```

The chi-square test returned $\chi^2 = 14.504$ with $p = 0.0007$, and since $p < 0.05$, we reject H_0 and conclude that tackle level and YAFC_{gt3} are statistically associated. To describe the size and direction of the difference between Level 2 and Level 3 specifically, the sample proportions were also computed. In the observed data, Level 2 had $\hat{p}_2 = 0.1931$ (28/145) and Level 3 had

$\hat{p}_3 = 0.3684$ (126/342). The estimated proportion difference is $\hat{p}_2 - \hat{p}_3 = -0.1753$, indicating that Level 3 tackles produced a higher rate of YAFC > 3 by 17.53% in this sample. A 95% confidence interval for the true difference $p_2 - p_3$ is $[-0.2574, -0.0932]$, suggesting the true Level 2 minus Level 3 difference is likely between -25.74 and -9.32%. Because this interval does not include 0, the data support a meaningful difference between Level 2 and Level 3 in the probability of allowing more than 3 yards after first contact.

Discussion and Limitations

A big strength of this project is that it's about something coaches talk about every day: tackling, it's technique, and how can the team improve on it. The premise of this study was to take tackle data and put it into a structure that can be measured and tested. For the data that was collect, a lot was accomplished. At the same time, there are a few limitations that matter when interpreting the results, and they point directly to what should be improved next. One limitation is that tackle difficulty was not fully controlled for beyond the variables already tracked. In particular, tackle angle and the ball carrier's path were not coded in a specific way. An inside out tackle where the defender can run the feet on contact is a very different situation than a downhill tackle in a tight space, or a pursuit tackle with the runner bending back across the grain. Without an angle category, some of the variation in make rate and YAFC is probably being driven by geometry rather than contact level alone. This make investigating probabilities hard. For example, setting up a binomial distributiion to predict tackles wouldn't be fair because a box, level 2 side tackle is much harder than a level 3 side line tackle. Those probabilites are not even.

A second limitation is sample size, especially once the data get split into smaller groups. The full dataset is solid for a team level view, but when the analysis is broken down into engaged versus not engaged, or Level 2 versus Level 3 within those groups, some categories shrink quickly. That matters because statistical tests become less stable and confidence intervals widen. It also increases the chance that a few unusual plays, like one or two long runs after contact, can move the averages more than they should. The grading process had its own hiccups as well. To start, It was incredibly time intensive. Even with a clear rubric, grading each tackle takes real time and focus, and fatigue can affect consistency. Related to that is a source of bias: the dataset was graded by one person. Even if the rubric stays the same, a single evaluator may unintentionally drift over time or interpret borderline plays differently depending on game context, camera angle, or expectations. That does not mean the grading is wrong, but it does mean reliability cannot be verified without a second grader or repeat grading checks. This was the reason why the grading wasn't out of 10 to avoid those borderline plays.

Another practical limitation is video quality and camera angle. Some tackles are clear from end zone or sideline film, but others are partially blocked or happen off screen, especially near the boundary or during quick hitting plays. When the view is limited, the contact point and the exact moment of first contact can be harder to identify, which can lead to small errors in level classification or YAFC assignment. This makes tackles by defensive lineman in giants piles particularly hard.There is also the issue of independence. The statistical tests used treat tackles like separate observations, but football is not fully independent play to play. The same defender appears repeatedly, the same opponent tendencies show up within a game, and certain play calls create similar tackle situations. That clustering can make results look more certain than they really are if it is not accounted for, especially in player specific summaries.

Finally, the data focus on first chance tackle attempts, which is the right coaching lens, but it can miss context around help defenders and pursuit, and also gives no credit to the players that made the tackles after a miss. A tackle might be scored as a make with low YAFC partly because a second defender arrived quickly and closed space, or because the ball carrier had already slowed. That does not invalidate the metric, but it means tackle outcomes are not only the tackler's technique, they are also the environment around the tackle.

Conclusion

With all analyses completed, the broad conclusion is that Level 2 tackles were more effective and more efficient than Level 3 tackles in this dataset. The clearest support for that statement comes from the set of hypothesis tests that evaluate both tackle success and tackle control. Across the full sample, Level 2 produced a higher make rate, and in not engaged situations that advantage remained clear. On the efficiency side, made Level 2 tackles allowed fewer yards after contact on average, and Level 2 tackles also carried a higher average technique score. The chi-square test added a different angle to the same story by showing that tackle level is associated with whether a tackle allows more than three yards after first contact, which is a simple and coach friendly threshold for describing explosive plays after contact.

What matters most from a statistical certainty standpoint is that the comprehensive testing picture is consistent. When the results are taken together, five of the six hypothesis tests supported the same direction of effect, Level 2 outperforming Level 3 in either make probability, limiting yards after contact, or technique quality. That consistency across multiple metrics and multiple testing approaches strengthens the conclusion beyond any single number. The one comparison that did not show a difference was within engaged tackles, where the sample size was smaller and the confidence interval was wide. Even with that exception, the overall evidence points toward a practical coaching takeaway: emphasizing Level 2 contact, especially in situations where the defender is not engaged, is likely to improve both tackle completion and the ability to limit yards after contact.

References

Shoulder tackling & the 5 fights progression. Shoulder Tackling — Football Skills & Drills — USA Football. (n.d.). https://usafootball.com/coaches-organizations/shoulder-tackling

Wikimedia Foundation. (2025, December 12). Chi-Squared Test. Wikipedia. https://en.wikipedia.org/wiki/Chi-squared_test

Wikimedia Foundation. (2025, December 12). Two proportion z-test. Wikipedia. https://en.wikipedia.org/wiki/T proportion_Z-test

Wikimedia Foundation. (2025, December 12). Welch's t test. Wikipedia. https://en.wikipedia.org/wiki/Welch%2 test

Code Reference

All Code will be uploaded to my personal GitHub page at www.GitHub.com/jlars88. A submission of my .ipynb will also be available.