Grant Starnes

Dr. Liz Wanless

TESM-S501: Advanced Sports Analytics

26 October 2025

Final Project Phase II

Preliminary Idea #1 for 2026 NFL Big Data Bowl: How quickly do defenders react while the ball is in the air?

Motivation

For this project step, my first preliminary idea for the 2026 NFL Big Data Bowl was regarding defenders and how quickly they close in on receivers when the ball has been thrown. For this idea, I called it defensive recovery speed, as there's more than likely to be a bit of separation between the defenders and receivers. My analytics question, as stated above, is how quickly do defenders react while the ball is in the air? This idea takes in the defender recovery speed, but then categorizes it by pass result and coverage type to make the analysis more applicable and usable. Pass results are categorized as complete, incomplete, or intercepted. Coverage types include man coverage and zone coverage. The defender recovery speed on its own wouldn't provide much actionable insight, but tying in the result of passes and what coverage the defense was in helps to make comparisons and truly understand what could be improved in game. Incomplete passes, for example, can result for a number of reasons like poor throws from the quarterback, but it's most likely because of the pressure the defense is applying

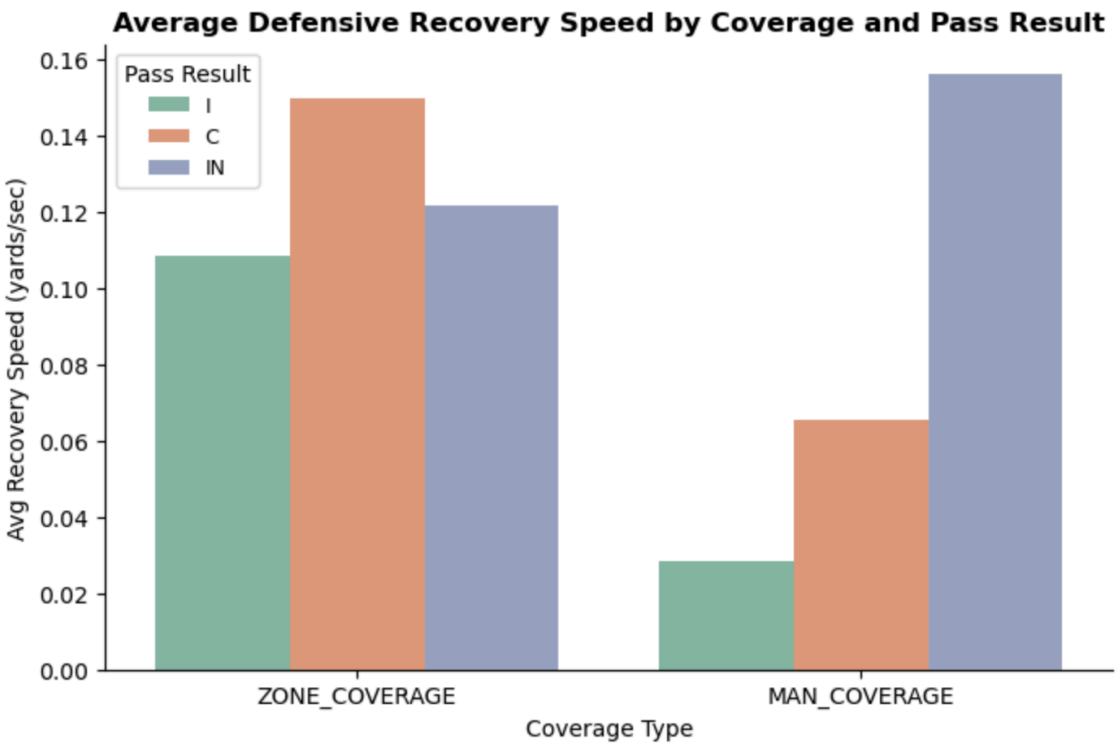
on the offense, or the receivers in their coverages, and how quickly the defender is able to recover and close in on the receiver. Analyzing recovery speed by defenders allows us to look at the closing ability by defenders and coverage pitfalls and successes.

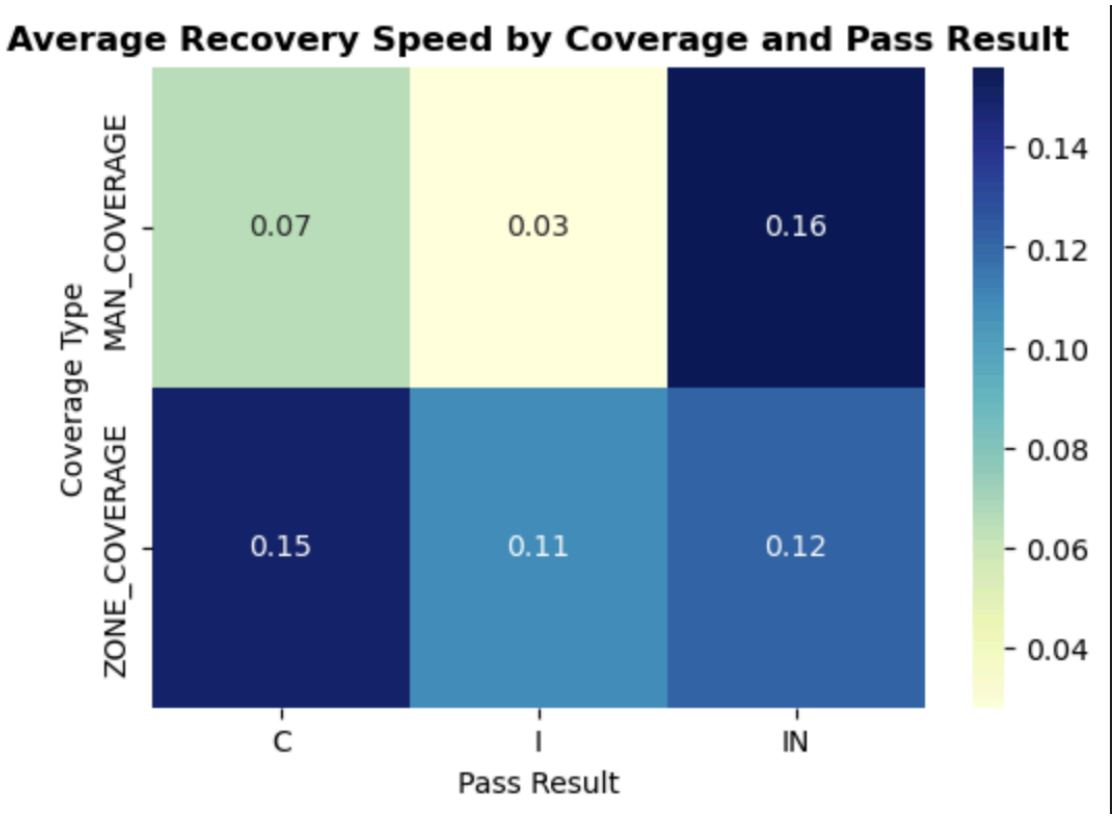
As for the data sources used, I used tracking data, such as the receiver and defender x and y coordinates on the field for each frame. This is helpful in calculating the distances between a defender and their receiver over the course of a play. Next, I used play-by-play data to get data for areas like pass results (complete, incomplete, interception) as well as play_ids and game_ids to make sure the data matches up correctly. Lastly, a very critical aspect of the data I included was team_coverage_man_zone to determine what coverage the defense was in for each play, man or zone coverage. These data sources are important for my question of defensive recovery speed because it tracks the player movement throughout each play to determine their speed, the outcome of each play to help understand how tighter coverages cause more incompletions, and the type of coverage to show how it may vary between either man or zone.

The potential value of this to coaches, front-office staff, or player development would be to help these key stakeholders monitor a defenders' current closing ability and reaction time and how to improve upon it if needed, determine which coverage, man or zone, suggests better defender recovery speeds which could be applied to certain situations in game, and provide insight to defensive back coaches or strength and speed trainers in order to improve these recovery speeds and reaction times.

Methods

As for my process, I read in the input, output, and supplementary files to my jupyter notebook in Visual Studio Code. After that was complete and the data was present, I started with cleaning and filtering the tracking data for the players I wanted to observe, such as offensive players and then the offensive players' role, with the desired role being a targeted receiver. Since we're mainly concerned with defenders here, I then filtered for defensive players. After that, I was able to pair receivers and defenders on attributes like game_id, play_id, and frame_id to maintain the pairings over the course of each play. Then, for each receiver-defender pair, I calculated the distance between them using the equation of the square root of the receiver x - defender x squared plus the receiver y - defender y squared. This helped determine the changes in distance over the course of each frame, with negative outcomes portraying a defender closing the distance. I then averaged these distances for each play to determine the average recovery speed. For my final steps, I then applied this average recovery speed to pass results and defensive coverages to get a bit of a more applicable and interpretable analysis.


Average Defensive Recovery Speed by Coverage and Pass Result
Pass Result
I
C
IN
Avg Recovery Speed (yards/sec)
0.16
0.14
0.12
0.10
0.08
0.06
0.04
0.02
0.00
ZONE_COVERAGE
MAN_COVERAGE
Coverage Type


Average Recovery Speed by Coverage and Pass Result
Coverage Type
MAN_COVERAGE
ZONE_COVERAGE
0.07
0.03
0.16
0.15
0.11
0.12
C
I
IN
Pass Result
0.14
0.12
0.10
0.08
0.06
0.04

Results

Focusing on my early findings, the results I got seem to suggest that the defenders involved in incomplete pass plays have a higher recovery speed getting back to break up a pass to the receiver, which helps support the idea and correlation that faster defender recoveries tend to lead to more pass incompletions. Also, defenses in man coverage seem to have faster defender recovery speeds compared to defenses in zone coverage due to the fact that the defenders are keyed in on one receiver and can react off of what they do, rather than sit in a zone where defenders are more static.

As for patterns or relationships observed, as touched on previously, for faster defender recovery speeds there are more incompletions as the pass result, supporting a negative relationship between the two. Next, zone coverages are more about where the defender is positioned compared to man coverage where the defender is covering one receiver and far more active.

The results I received relate back to my motivation in the sense that the results support initial football intuition and what is expected generally where tighter coverage and faster defender recovery lead to more incomplete passes. Since the results are behaving as expected, they could be used as stated earlier to improve player recovery speeds and reaction times and help inform the coaching staff with which coverages work better regarding recovery speeds and how they can apply that to in-game situations.

Discussion

Moving on to the limitations of this idea, pairing defenders to receivers could be a bit shaky in this model as the pairing is a bit ambiguous. If I stick with this model, I would most likely need to improve the tracking to the closest defender or base it on coverage assignments if possible so the results are far more accurate. I'm not sure if this is a limitation or not, but something else I could implement and analyze is how the speed attribute varies from the acceleration attribute in the data. One could end up being more accurate and provide better insights than the other, and it wouldn't hurt to try.

As for feasibility for a full Big Data Bowl submission, I think this idea is feasible and mainly meets the goals for this year's submission criteria of player movement while the ball is in the air. I would refine, like stated previously, the receiver-defender pairing to use the closest defender at all times so my output and visuals are as accurate as they can be. I don't think I'd ditch this idea altogether because it can be pretty insightful, but it definitely could use some work.

Some collaborations that might be helpful for the project would be reaching out to Geoffre Sherman since he already took time to talk to the class about the project and kind of break it down. He could help me think outside the box and tie in new attributes aside from pass results and defensive coverages, and give me a push in the right direction. Another collaboration

could be working together with peers to see what they think and if they did something similar and see if they have ideas to improve my idea.

Preliminary Idea #2 for 2026 NFL Big Data Bowl: How does separation affect pass results?

Motivation

The focus of my second idea for the 2026 NFL Big Data Bowl is the average separation between receivers and defenders while the ball is in the air. My analytics question, as seen above, is how does separation affect pass results? In football, separation is extremely important due to the fact that receivers with more space are far easier to throw to from the quarterback's point of view and are far more likely to make a catch. The defender's job is to close the gap and break up the pass in any way they can.

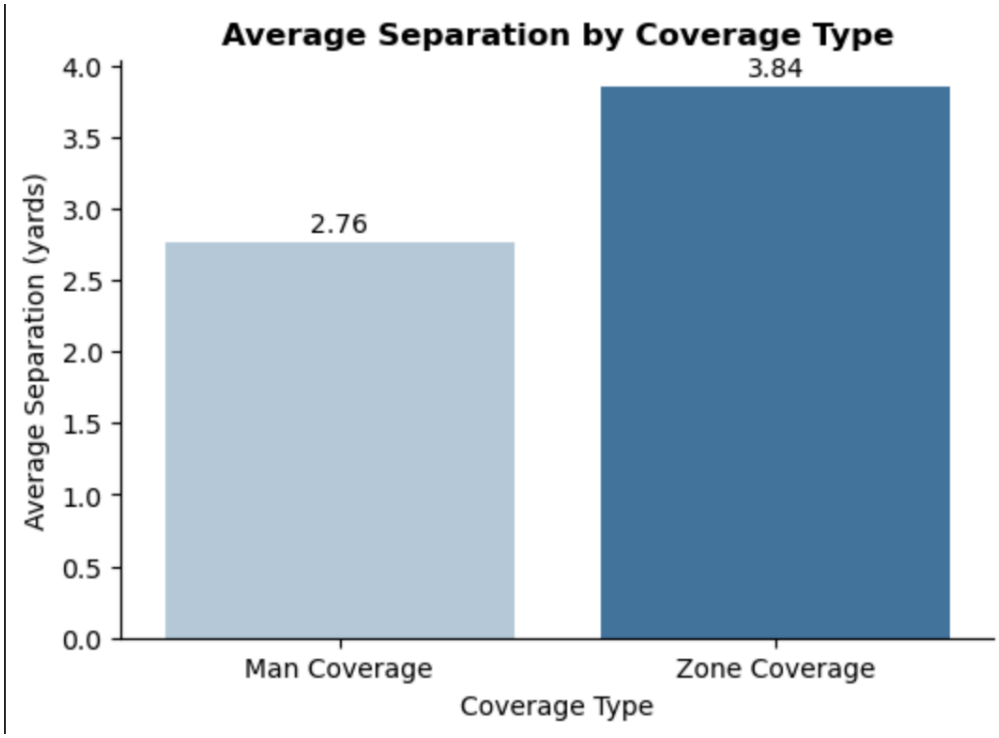
As for the data sources I used, tracking data was very important here since I used the receiver and defender x and y coordinates on the field as well as the frame positions to help calculate the distance between a receiver and the closest defender. Play-by-play data was also important in order to match data on game_id and play_id and also determine things like pass results. Lastly, we look at the different coverage types, man or zone, so I used data from the supplementary file to look at the attribute team_coverage_man_zone. These data sources are important for my question of separation and pass results because it tracks the player movement throughout each play to determine their positioning on the field, the outcome of each play to

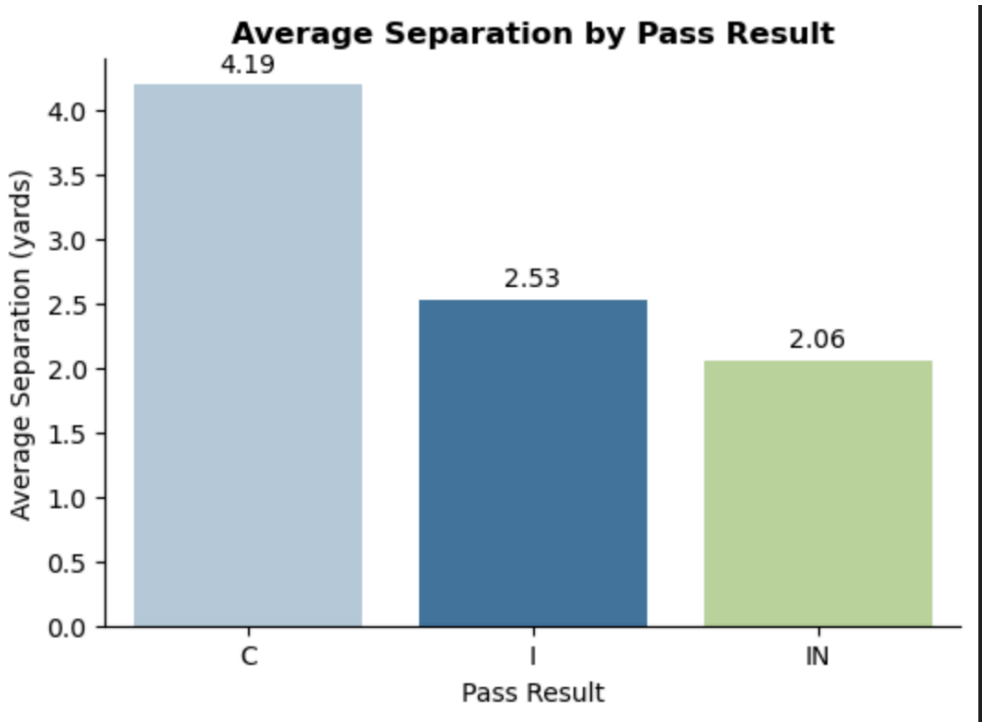
help understand if the play result was either a completion, incompletion, or interception, and the type of coverage to show how how receiver separation can vary between man and zone coverages.

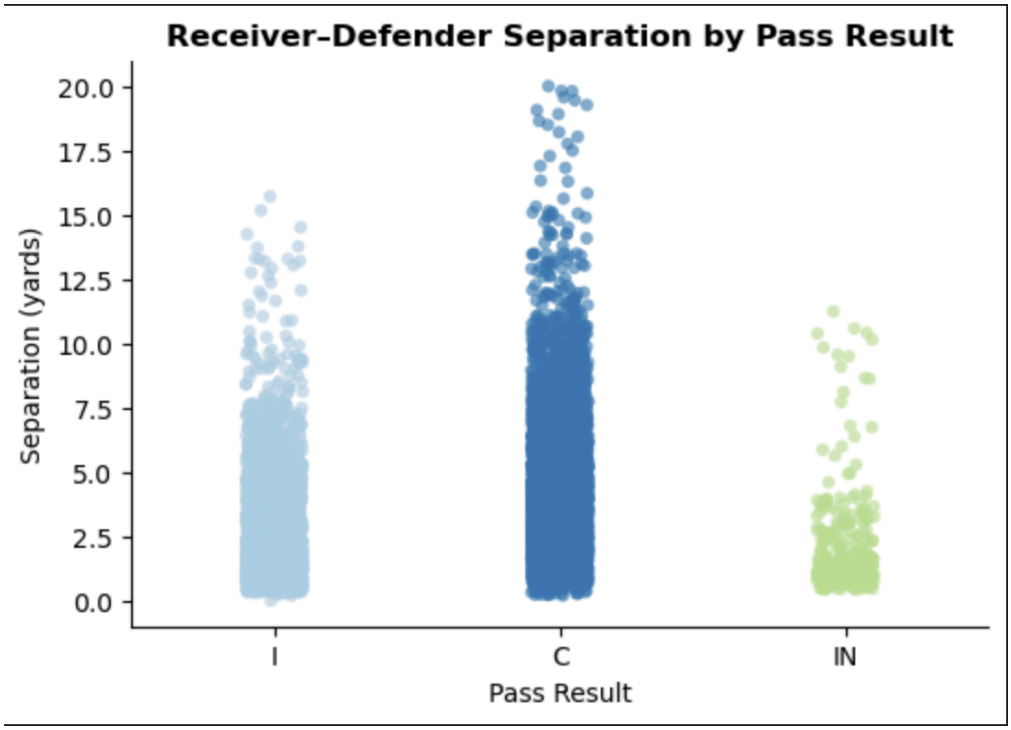
As for potential value to coaches, front-office staff, and player development, coaches can look at how efficient defenders are in man and zone coverages and what the separation looks like, and can call plays based on that. Similarly, offensive coordinators can use this to determine favorable matchups, for example, where the receiver can very well get open and create lots of separation against the opposing defender. For player development, these coaches can really emphasize the importance of receiver routes and positioning in order to capitalize and generate more separation.

Methods

For my process, since the data was already read in for the first idea, I didn't have to do that again and could start this idea right away. That said, I first determined receiver and defender trajectories. Next, I did a pairwise distance calculation for all receiver-defender pairs in each frame. Moving on, for each receiver in each frame, I only kept the defender with the least distance between a defender and the receiver. Next, I tied in coverage types, man and zone, as well as the pass results, complete, incomplete, or intercepted. Then, I created a metric that determines good coverage, which is based on the average separation for both man and zone coverages.


Average Separation by Coverage Type
Average Separation (yards)
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
2.76
3.84
Man Coverage
Zone Coverage
Coverage Type


Average Separation by Pass Result
Average Separation (yards)
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
4.19
2.53
2.06
C
I
IN
Pass Result



Results

As for patterns and relationships observed, average separation is smaller regarding completed passes, which indicates that tighter coverage doesn't always prevent a completion from happening, but larger separation on incomplete passes may show defenders reacting faster, closing the separation more quickly, somewhat going hand in hand with my first idea. Also, man coverage tends to have a bit of a smaller average separation than zone coverage, showing that defenders are staying closer to their paired receiver, and supporting intuition as this is what is expected when comparing man and zone coverages.

Discussion

As for limitations, the closest defender may not always be guarding that specific receiver, as things can happen during the play based on the offensive scheme. Also, frame distances may not capture things like anticipation or attempts at interceptions. Lastly, separation can vary by different routes being run and the length of the passes, which this metric doesn't really account for as of now.

Moving on to feasibility, this idea is more straightforward to scale across all weeks of the season. Although this is true, I could refine this idea by assigning a primary defender in some way, as well as take in vertical and horizontal separation for further insights.

As for collaboration, I could reach out to coaches, such as high school coaches in the area to validate if my thinking is valid or not. Again, I could reach out to Geoffe Sherman to get his insights as he is rather familiar with this kind of data and analysis.