NFL Big Data Bowl 2025

Coaching Track

Jayden Cruz Berdecia, Colin Montie, Nicolas Thomas, Nathan Wright

Overview

Our task was to model offensive and defensive behaviors within NFL plays using provided player tracking and situational data, and we created a best plan of attack for NFL franchises based on the expected scenario of a given play

Our solution looked to analyze player movement and matchups, model offensive and defensive tendencies, and predict defensive reactions using game theory concepts.

We began with Data Handling, using the given datasets along with combining other data sources to expand on the capabilities of our research.

Background

- **Objective:** To create a comprehensive model for predicting defensive reactions in the NFL, including the expected number of rushers, coverage type, and post-snap adjustments based on pre-snap alignment, motion, and game context.
- **Motivation:**
  - **Tactical Adaptation:** Defenses continuously adapt to offensive strategies. Predicting these adaptations provides a significant advantage for teams, as it informs play calling and decision-making.
  - **Big Data Potential:** With advanced tracking and game data available, there is an unprecedented opportunity to model and predict defensive behaviors in real-time.
  - **Unanswered Questions:** Existing models often focus on static pre-snap formations or overly simplistic assumptions. Our project aims to address the complexities of defensive adjustments and player decisions, providing insights into dynamic, high-pressure environments.
  - **Impact for Teams:** By accurately predicting defensive responses, teams can better prepare for specific situations, optimizing offensive play calling and increasing their chances of success.

Data Handling

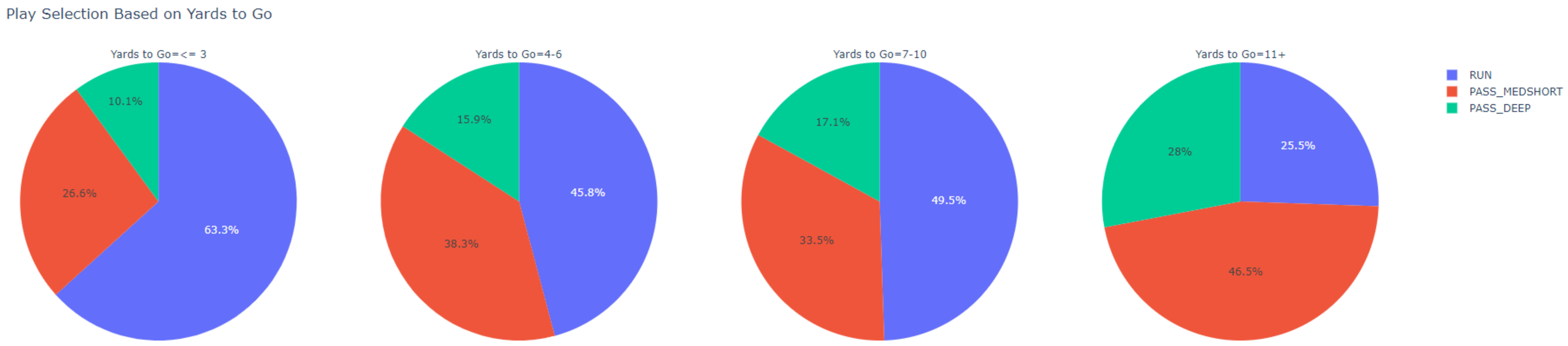
- We began with taking supplemental play-by-play data from nfl_data_py, game metadata provided by Big Data Bowl on Kaggle, and frame-by-frame tracking data.
- We focused primarily on filtering for key events, focusing on line_set, man_in_motion, and ball_snap events to reduce irrelevant data points while maintaining critical pre- and post-snap movements.

Methodology

- We derived metrics by calculating the play distances from the ball, line of scrimmage, and the closest offensive player to each defensive player. For all categorical variables, they were converted to binary to improve the usage and effectiveness of our datasets. These includes personnel, receiver alignments, motion, and down and distance.
- Supplemental Data integral for our research came in the form of merging filtered tracking data with the play-level context on the basic of gameId and playId. We also incorporated the success metrics of yardage gained and play outcomes into new data-frames to use for future modeling.
- Following this, the main modeling framework took shape, analyzing Offensive and Defensive tendencies.

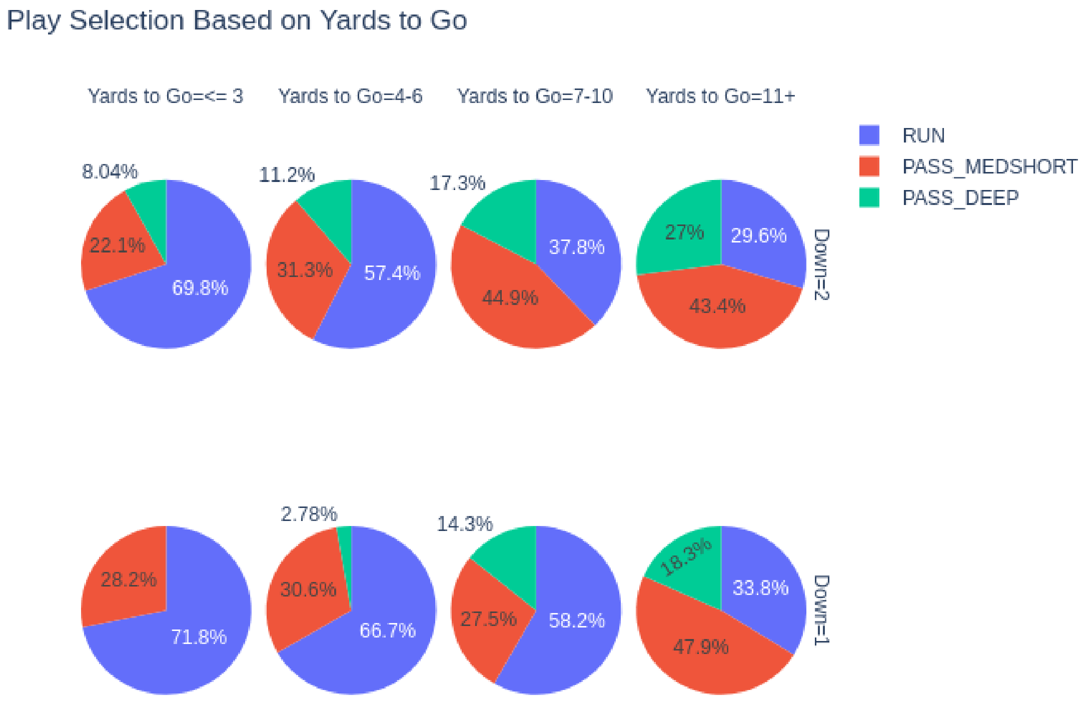
Play Type by Yards to Go

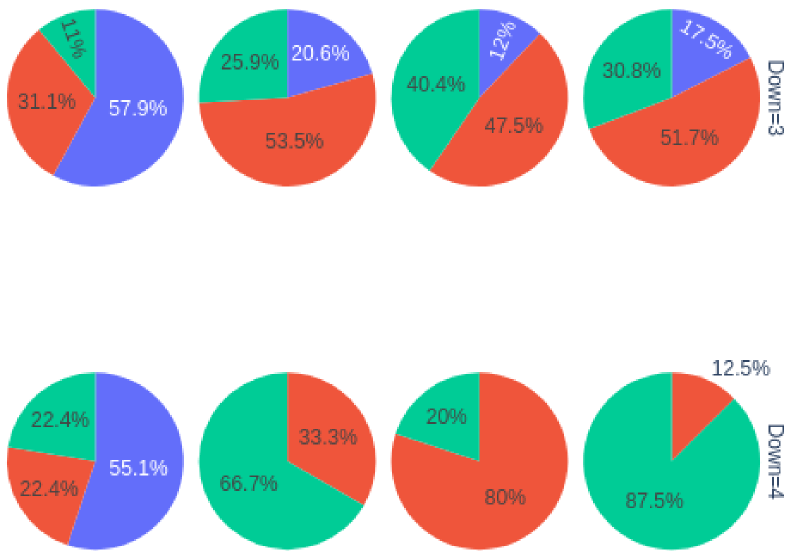
- The following pie charts display leaguewide offensive play-calling tendencies grouped by yards to go
- Plays are sorted into runs, short or medium passes, and deep passes
- Run plays are most common when there are fewer than 11 yards to go while short and medium passes are the most common for any longer yardage



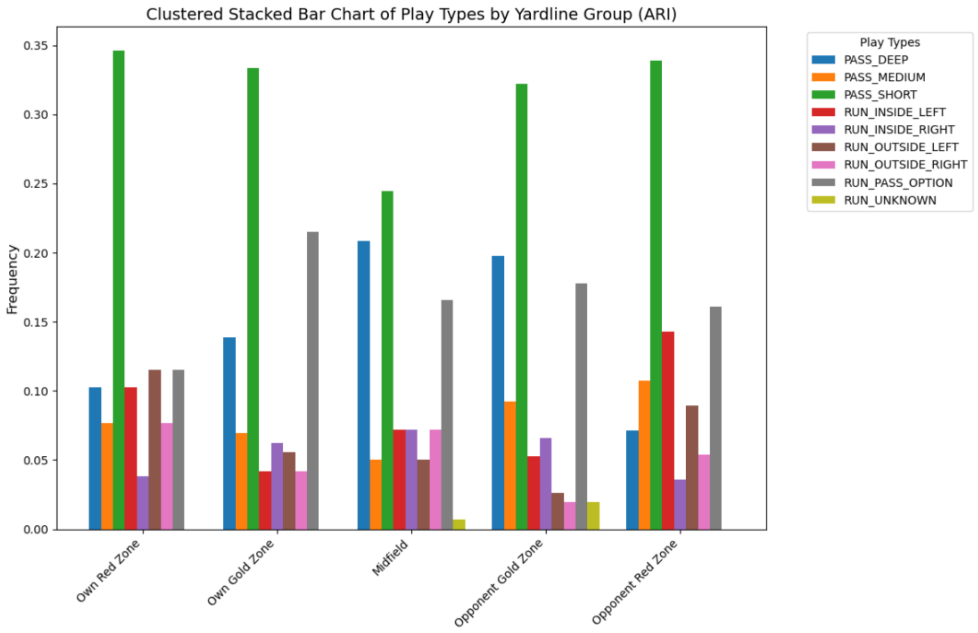
Play Type by Yards to Go (based on downs)

- These pie charts once again depict leaguewide play-calling by yardage to go, however they are also grouped by down
- An interesting takeaway from these is the lack of deep passes on 1st and short situations
- The idea of taking a shot on 2nd and short can be seen with 8% of plays on 2nd and <=3 being deep passes and 11% on 2nd and 4-6
- This idea has not carried over to 1st downs despite having an extra down to work with, leading to teams potentially missing out on big plays





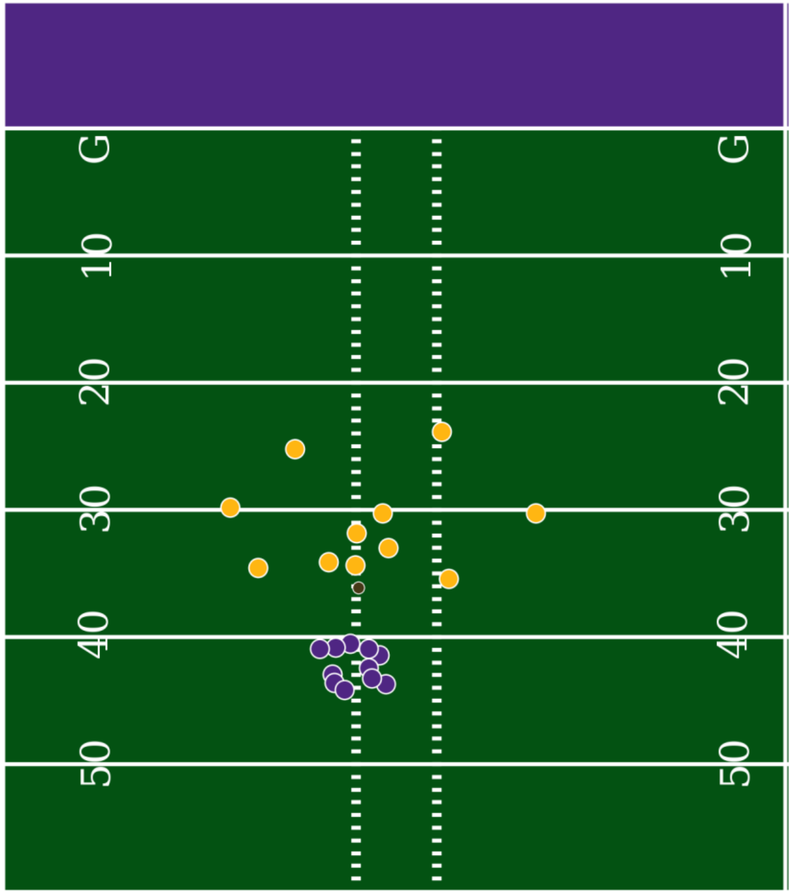
Offensive Modeling



- We began our offensive modeling by first sorting for the tendencies of each team, sorting plays by: Run Inside Left/Right, Run Outside Left/Right, Run-Pass Option, and passes(Short, Middle, Deep) for each NFL team, as well as finding an NFL average and comparing each team to it.
- Further cleaning was implemented, as some plays with unknown types (like QB kneels) were sorted out to remove variance and false results.
- We then created expected play type percentages based on how the offense is aligned pre-snap and at the time of the snap. Further research was conducted into the offensive team's tendencies in that situation, and the unique variables separating that play from how the offense usually runs, like motion.
- Play tendencies for each team were sorted into large graphics to relay their typical actions, providing a scouting report for how teams operate in given situations. Some variables we tested were: Offensive Personnel, probability of a Field Goal or Touchdown, RB Yards Per Game (derived through a custom function), Expected Points, Field Segment, Offense Formation, Time Remaining (by quarter and game), Distance, Down, Yards To Go, Drive Yards, and Yardline Number.

Offensive Play Prediction

- This example of the offensive play prediction model comes from a 36-yard touchdown pass from Kirk Cousins to Justin Jefferson
- This play occurred on a first down from the opposing 36-yard line with 45 seconds remaining in the half
- When the Vikings initially come set, they are lined up in a 3x1 singleback formation
- They then motion the tight end to the left side of the field to create a 2x2 alignment
- A second man went in motion to the left just before the snap
- Given these factors along with the other variables taken into consideration, our model correctly predicted this play to be a **deep pass**



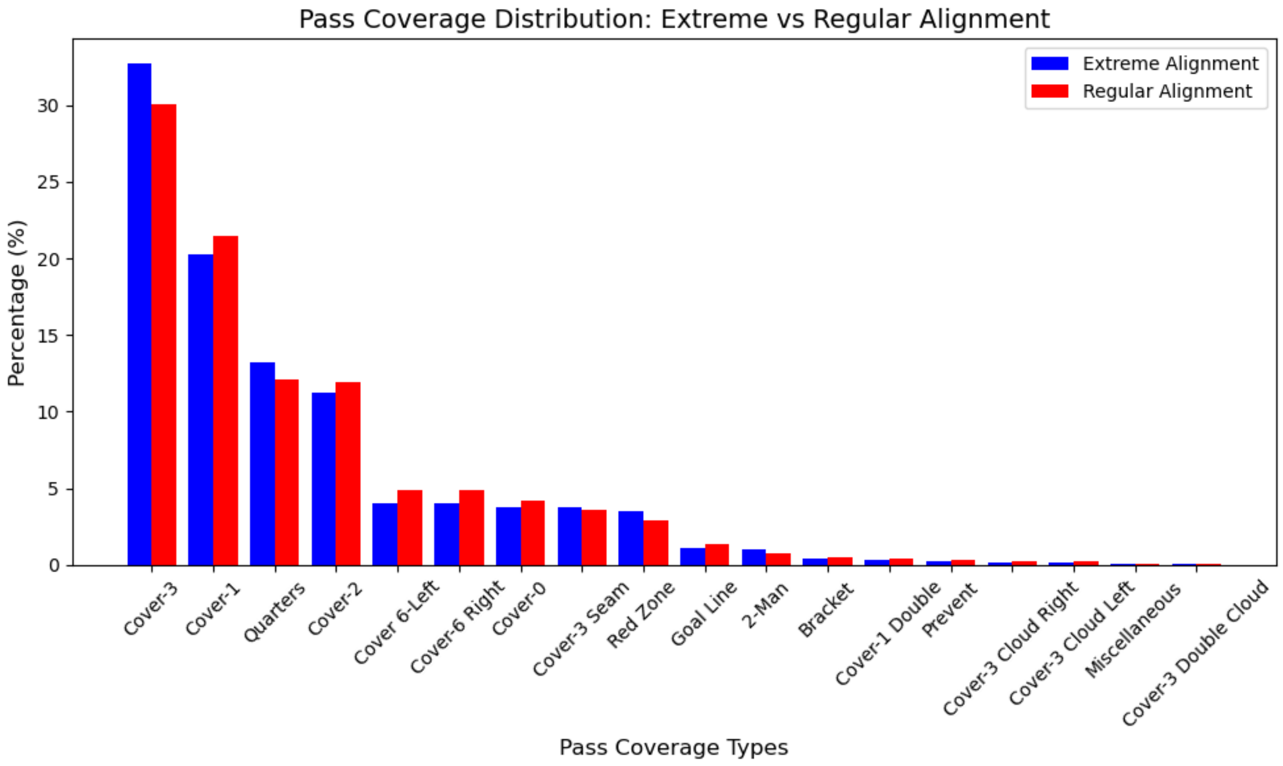
Defensive Modeling

Coverage	Precision	Recall	F1-score	Support
2-Man	1.000000	1.000000	1.000000	2608
Bracket	1.000000	1.000000	1.000000	968
Cover 6-Left	1.000000	0.944865	0.971651	10447
Cover-0	0.999324	1.000000	0.999662	8873
Cover-1	0.867812	0.763583	0.812368	49015
Cover-1 Double	1.000000	1.000000	1.000000	762
Cover-2	0.956246	0.697107	0.806368	27683
Cover-3	0.729967	0.952746	0.826609	79062
Cover-3 Cloud Left	1.000000	1.000000	1.000000	426
Cover-3 Cloud Right	1.000000	1.000000	1.000000	462
Cover-3 Double Cloud	1.000000	1.000000	1.000000	134
Cover-3 Seam	0.998705	0.969611	0.983943	9543
Cover-6 Right	0.997220	0.933758	0.964446	10371
Goal Line	1.000000	1.000000	1.000000	2305
Misc.	1.000000	1.000000	1.000000	141
Prevent	1.000000	1.000000	1.000000	523
Quarters	0.935546	0.735548	0.823579	31692
Red Zone	0.997851	1.000000	0.998924	7892
Accuracy	0.852813	0.852813	0.852813	0.852813

- Defensive Modeling began with deducing the pre-snap alignment based on the X and Y value of the offensive player, along with their Speed, Acceleration, Distance Traveled, Orientation and Direction of Motion.
- Focusing on tendencies, we looked at the Defensive Team and gathered what they would likely call based on the Offense is expected to run, analyzing number of rushers, Defensive coverage scheme, and Man/Zone coverage.
- Through our main variables of Pass Coverage and Man/Zone coverage, we created a scouting report to provide NFL franchises on how certain Defenses will react when presented with particular formations, alignments, and personnel. Play tendencies for each team were furthermore sorted into large graphics to relay their typical actions, providing a scouting report for how teams operate in given situations.
- Important frames were taken from the main resulting datasets and merged with offensive motion to find the defensive tendencies dataset. Following this, the dataset was sorted to find the tendencies from different coverage types, along with finding the closest coverage defender. After testing, random forest models were created and analyzed to predict the likelihood of what the defensive coverage would be, based on specification from player alignment and motion.

Player Alignment

- Player alignment through either a wide formation or motion is critical to evaluating the different coverage responsibilities that may happen on a play.
- To model this, all of the plays were sorted to find the values where alignment is significantly over the average, in our case it was the top 80% that became extreme values cases.
- If alignment for any of the following was over the threshold, it became critical pieces of our data: X and Y value of the offensive player, along with their Speed, Acceleration, Distance Traveled, Orientation and Direction of Motion.



Model Evaluation Alignment

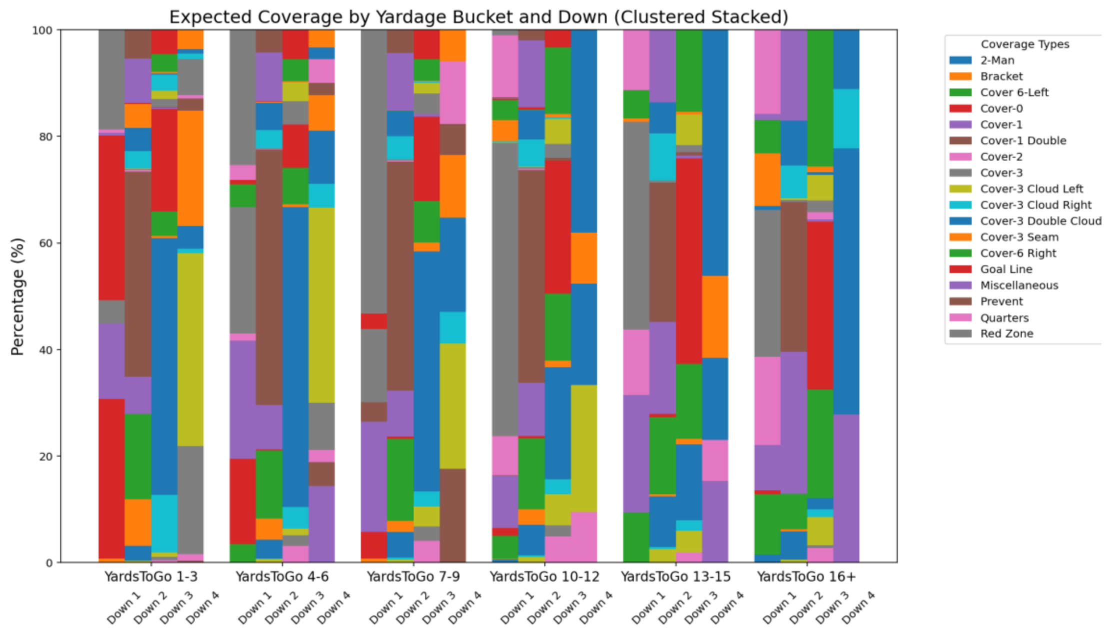
- Models were created within our research to determine how useful and accurate player alignment features were for determining the coverage on the play. It also implemented down, yards to go and motion in the calculation, and focused on the variables of "pff_passCoverage" (Coverage Types) and "pff_manZone" (Man or Zone coverage).
- The models were then separates to test the coverage variables by themselves, along with using different model calculations with and without motion to determine their value for coverage responsibilities.
- Through the models, they all yielded 84% or better accuracy at predicting coverage, particularly with the Man vs Zone coverage responsibilities. Motion also improved the model for both variables.

Model Accuracy: 0.9256946612550734

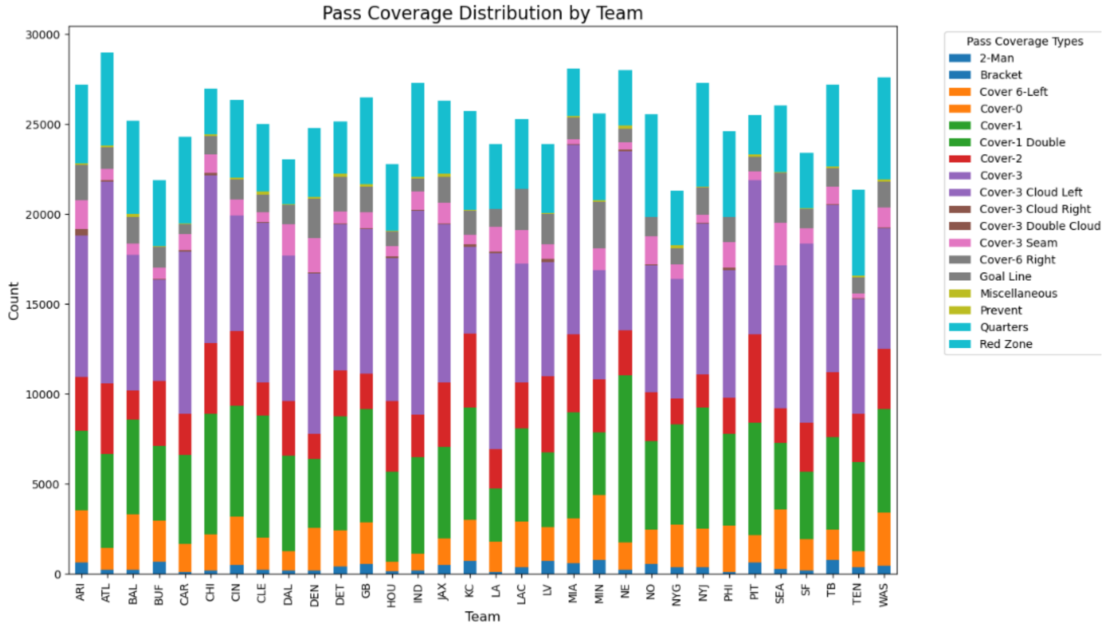
	precision	recall	f1-score	support
Man	0.97	0.74	0.84	13015
Other	0.98	0.93	0.96	2786
Zone	0.91	0.99	0.95	35447
accuracy			0.93	51248
macro avg	0.95	0.89	0.92	51248
weighted avg	0.93	0.93	0.92	51248

Model Accuracy: 0.8644630034342804

	precision	recall	f1-score	support
2-Man	0.97	0.79	0.87	570
Bracket	1.00	0.88	0.94	236
Cover 6-Left	0.95	0.71	0.82	2020
Cover-0	0.97	0.88	0.92	1996
Cover-1	0.89	0.87	0.88	10263
Cover-1 Double	0.99	0.85	0.92	186
Cover-2	0.90	0.80	0.85	6131
Cover-3	0.78	0.97	0.86	16204
Cover-3 Cloud Left	0.99	0.70	0.82	97
Cover-3 Cloud Right	0.96	0.66	0.78	125
Cover-3 Double Cloud	1.00	0.74	0.85	38
Cover-3 Seam	0.94	0.72	0.82	1967
Cover-6 Right	0.97	0.71	0.82	2128
Goal Line	0.99	0.92	0.95	454
Miscellaneous	0.94	0.85	0.89	34
Prevent	0.99	0.94	0.96	157
Quarters	0.91	0.78	0.84	6737
Red Zone	0.97	0.97	0.97	1905
accuracy			0.86	51248
macro avg	0.95	0.82	0.88	51248
weighted avg	0.88	0.86	0.86	51248



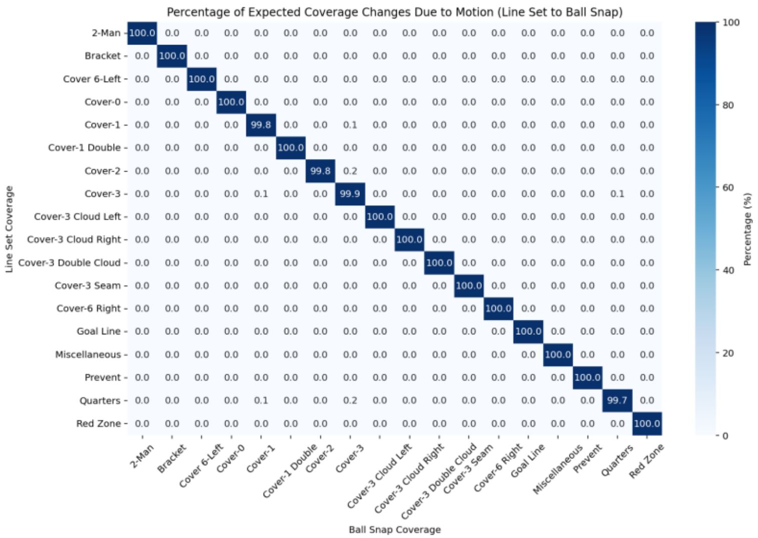
Expected Coverage Scheme by Down and Yardage Bucket

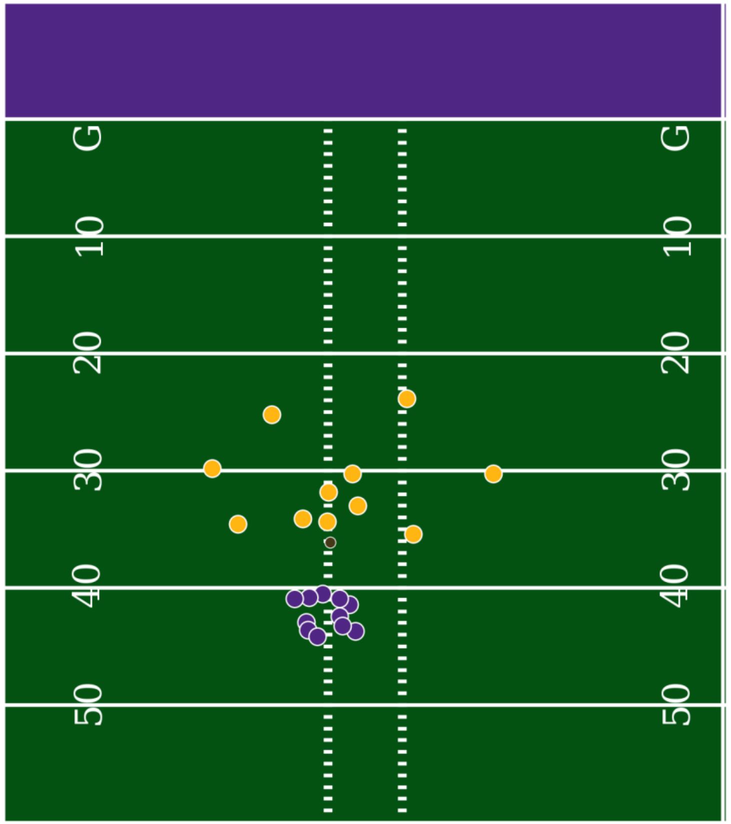


Coverage Distribution Tendencies by Team

Defensive Play Prediction

- Using the same play as before, we can see that the defense is lined up in a 2-safety look when the offense gets set
- However, one safety rotates down towards the line of scrimmage after the offensive shift, creating a single safety look
- Given the expectation of a deep pass as well as the other situational factors taken into consideration, our model correctly predicted the coverage to be **Cover-3**
- What we found through our analysis is that a defensive team's pre-snap shifting has very little impact over the coverage they are expected to run





Finalizing Research/Future Expansion

- Throughout the project, we conducted Exploratory Data Analysis to validate feature correlations through modeling NFL Offensive and Defensive tendencies.
- Many different features and variables were explored, with the Offensive side focusing more on the play-call and how it was determined by outside factors, along with the Defensive side focusing on player alignment and how teams would adjust in terms of coverages.
- Future investigation into the tendencies of both the Offense and Defense can be greatly expanded upon with more years of data accounted for and analyzed. As more and more games are played, our dataset will only become larger.
- A machine learning model, taking user input and showing the percentages that it happens on a particular play would also be down the line in potential innovations to the project. This would computer real-time user situations to determine what the best course of action would be for a particular team given the live situation.

Appendix

Link to Google Colab Notebook with code:

https://colab.research.google.com/drive/1auSNHL-PVsUHyNcSrBVrbFG_mC06bvRm?usp=sharing