Clutch Up: Analyzing Passing Tendencies in the Two-Minute Drill

Exploratory Data Analysis

Importing and Merging the Data

First, the data sets are loaded into the notebook environment using Python's `pandas` library to create three different DataFrames, `games`, `offense`, and `plays`, after reading the CSV files. The `offense` and `plays` DataFrames are then merged on observations matching the `game_id` and `play_id` columns. This new data set contains 231,371 entries and 88 columns highlighting various aspects of player and play-level information. Therefore, `player_play_id` is the primary key representing the unique identifier for each player during each play in each game.

Cleaning the Data

Before conducting any analysis, it's important to check and fix any data issues, which can include duplicates, missing values, or incorrect data types. First, many of the variables are modified with the appropriate data type. Some integer columns with binary values are converted to Boolean. The Boolean columns `off_is_touchdown` and `off_has_man_in_motion` have missing values, which are assumed to have False values. There are also 64 duplicated entries appearing in the primary key column, so these entries are removed.

Next, each `play_id` should count exactly 11 values representing the 11 offensive players on the field. Since there is only one play in the entire data set without 11 men on the field, it won't hurt to remove this entry as well. About 6% of the remaining plays have been nullified by penalties, and these entries are also removed since the data does not contain enough information, like penalty type, to draw conclusions about their effect on the play's outcome.

Creating a New Dataset

While `df` represents all passing plays without penalties throughout the season, a new data set is created to represent the following situation: passing plays in the fourth quarter with two minutes or less in which the offense is trailing by more than three but no more than eight points. Focusing on this specific scenario minimizes bias compared to analyzing all two-minute drill situations for a couple reasons.

First, the analysis targets a set of plays that are more homogeneous in terms of offensive objectives and defensive expectations. This narrowed game context ensures that the offense's goal is unambiguous – they must score a touchdown to tie or win the game. Analyzing situations where a team might be playing for a field goal, running out the clock, or

facing a large deficit increases the variability in the offense's intent which can introduce bias. Furthermore, defenses are more likely to play the pass and prevent big plays specifically, standardizing their approach and minimizing the variability in defensive play calling.

Second, all teams in this scenario are operating under similar time constraints and urgency levels. This contrasts with broader two-minute drills where the urgency can vary widely based on the score differential. Teams down by a significant margin may behave differently than teams in a close game, introducing bias in the analysis of their play calling and execution. The stakes (i.e. the chance to tie or win the game) are comparable across all instances in this scenario, allowing for a more consistent analysis of late-game, high-pressure tendencies. This comparability is less certain in a broader set of two-minute drills where the importance of the game situation can vary.

Ultimately, the homogeneity stemming from this scenario reduces the impact of various biases that can arise from differing game contexts. Hence, the `tm_needTD` DataFrame containing 363 plays (78 different drives) is created to reflect these conditions.

A successful drive for the offense in the defined scenario is one resulting in a touchdown – any other outcome will almost certainly result in a loss. Out of 78 total drives, there were 18 successes: 16 passing touchdowns, a seven-yard QB scramble, and a one-yard QB sneak (inferred by the "Jumbo" formation), equating to an approximate 23% success rate. Seventeen drives (~22%) ended with an interception, and the rest are assumed to end with a turnover on downs, fumble, or game clock expiration.

Leaguewide Passing Tendencies of the Two-Minute Offense

The basic strategy of the two-minute offense is to move the ball quickly and efficiently into scoring position. While there may be slight variations in personnel, alignment, and playcalling between teams, the two-minute script is generally predicated on field position, down, and clock management. These three factors should have an impact on where the ball is thrown and how aggressive or conservative the offense is. A new variable called `field_position_zone` is created that categorizes the existing `field_position` column into four critical zones: yellow, green, red, and goal line. The figure below shows the distributions of passing plays by `field_position_zone`, `down`, and `off_timeouts_remaining`.

Distribution of Passing Plays by Field Position

Distribution of Passing Plays by Down

Distribution of Passing Plays by Timeouts Remaining
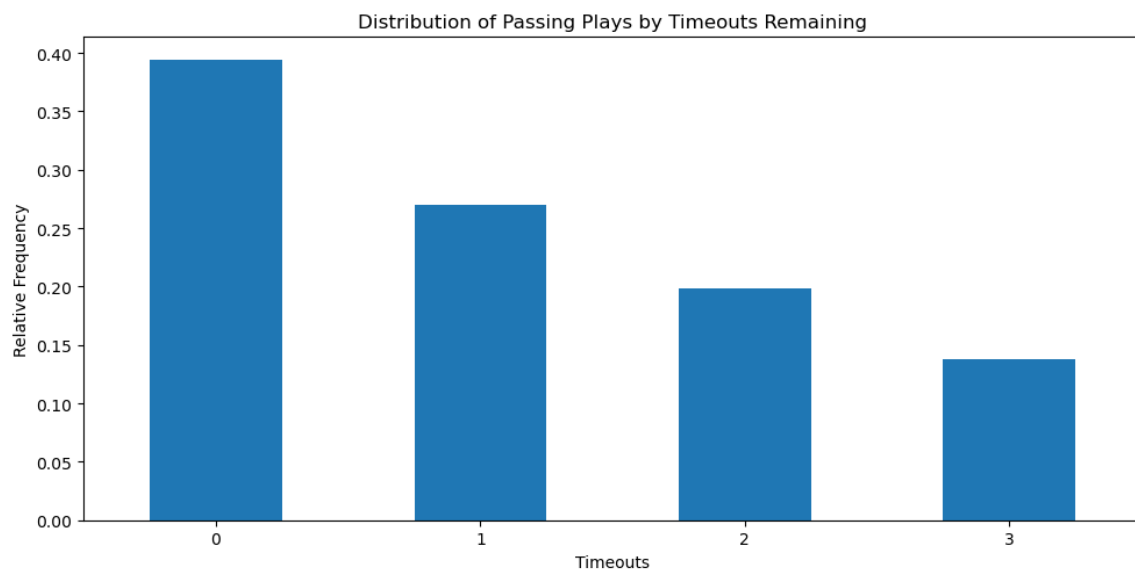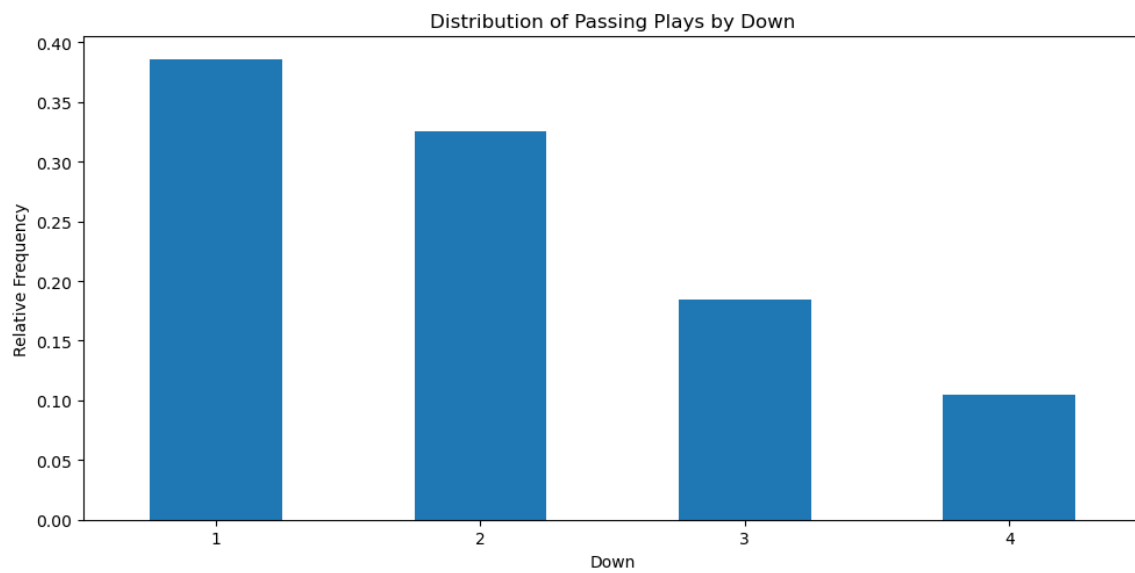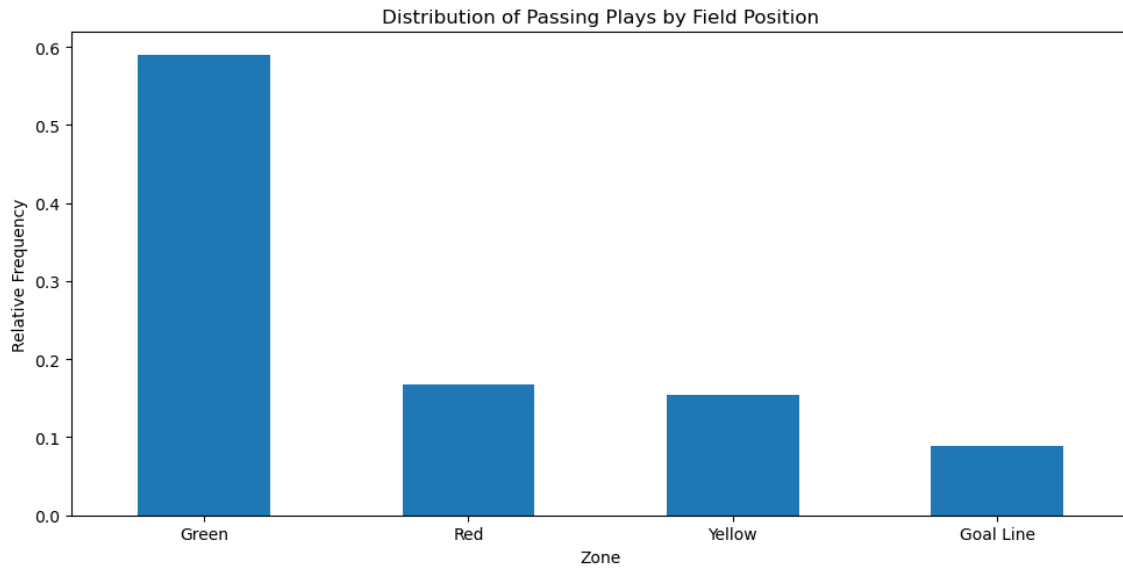
Many plays occur in the green zone, the largest chunk of the field between your own 26 yardline and the opponent's 21 yardline. This zone is generally where the offense runs its "bread and butter" plays and passing concepts to efficiently move the ball down the field. Just under 20% of plays occur in the yellow and red zones while about 10% of plays occur near the goal line. Given the late-game time constraint, it makes sense that the offense has zero or one timeouts for the majority of these plays. Nearly 70% of these plays occur on first or second down – a possible reflection of the offense's aggressive play-calling in these pressure situations to move the chains as quickly as possible. Additionally, examining the average air yards per pass (labeled as `pass_depth`) in each zone could be an indication of just how aggressive offenses are based on field position.
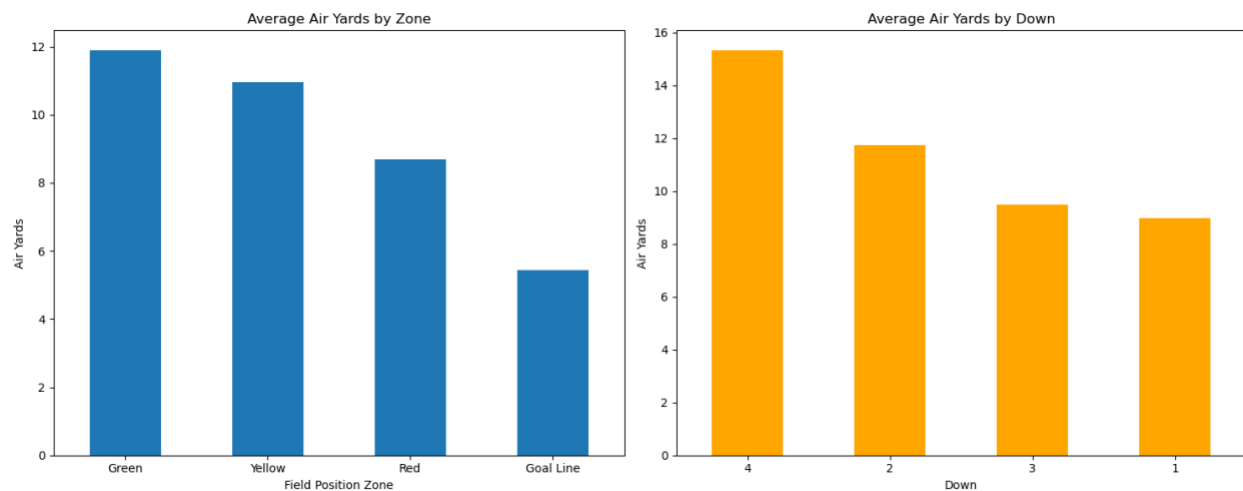

Average air yards are roughly four yards higher in this specific two-minute scenario than the rest of the data, indicating that offenses tend to pass the ball more downfield. It's no surprise that average air yards are about three yards higher in the green and yellow zones compared to the red zone and goal line. Being further away from the end zone, offenses have more room to operate and take advantage of over-aggressive defenses, thus giving them an incentive to take downfield shots. Closer to the end zone, defensive backs are more likely to play closer to the line of scrimmage, and the risk associated with a turnover is higher the closer a team is to scoring, naturally leading to shorter pass attempts.

On a similar note, offenses attempt shorter passes on first and third down, opting to make safer passes to keep the drive alive and/or set up for a more manageable fourth-down scenario. Average air yards are the highest for fourth-down plays, about 15.3, which could be a byproduct of fourth-and-long desperation passes/hail marys, but are still a signal that offenses tend to attempt longer, riskier passes.

From a defensive perspective, secondaries might adjust their coverages to account for these air-yard discrepancies depending on the field position and down. In situations where average air yards are higher, the secondary could play a softer zone or give a few

extra yards of cushion in man coverage, and vice versa in situations where average air yards are lower.

In addition to examining how far a pass is thrown, insights into the direction in which the pass is thrown could be advantageous for defenses. Football intuition would say that offenses look to throw the ball near the sideline more often during the two-minute drill to make it easier for receivers to get out of bounds and stop the clock. The data, however, shows that passes are thrown over the middle about 4% more often in two-minute situations when the offense needs a touchdown. While this may seem counterintuitive, there are other contextual factors shown in the figure below that could influence the pass direction.

Distribution of Pass Direction by Zone and Timeouts Remaining
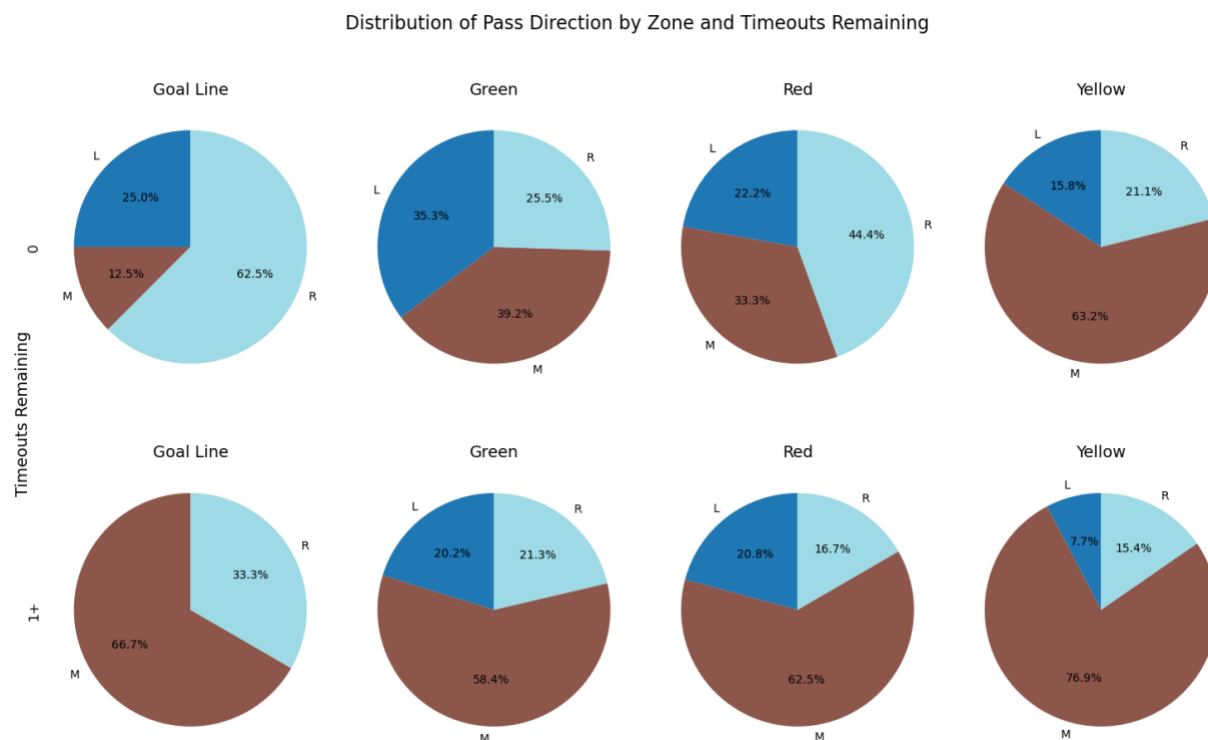
Clearly, teams are less afraid to throw the ball over the middle of the field when they have at least one timeout to burn. Interestingly, when in the yellow zone, offenses throw over the middle of the field over 60% of the time with 0 timeouts and over 75% of the time with 1 or more timeouts. This could be because defenses might be anticipating shorter and safer routes near the boundary, like flats and quick outs, exposing the middle of the field for big yards-after-catch (YAC) opportunities. When offenses have one or more timeouts, defenses might even allow more throws over the middle to force the offense to use their timeouts. Therefore, defenses must scheme for the tradeoff between time and yards, i.e. either allowing either the receiver to step out of bounds or gain more yards over the middle.

After exploring passing depth and direction in two-minute drills, a more granular approach looking at the primary passing concepts can help the defense be better prepared for certain situations. Without considering for any of the previous contextual factors, the following five passing concepts account for over half of the plays: Verticals (37%), Slot Under (5%), Ohio (4%), China (4%), and Drive (4%). Verticals, to no surprise, has the highest average air yards at 16 yards. Medium-yard passing concepts like Slot Under, Ohio, and Drive have average air yards of 6.8, 8.9, and 7.8 yards, respectively. China has the shortest average air yards out of these five with 4.5 yards.