Early-Down Playcalling Analysis

November 4th, 2024

Nathan Wright

Overview

In football, early-down playcalling is critical to setting up favorable third-down situations, increasing conversion rates, and ultimately enhancing a team's scoring potential. This project aims to analyze first and second-down playcalling combinations (run-run, pass-pass, pass-run, run-pass) and their impact on third-down distance, conversion rates, points per game, and winning percentages during neutral game situations (16-point spread or less (two possessions)).

Research Objectives

Factors and Data Sources for Model Development: This study seeks to build a model to assess the effectiveness of early-down playcalling strategies in football. The key factors and data sources essential for this model include:

- **NFL_Data_py Data:** Play-by-play data for all NFL games (2021-2023), capturing each play's type (run or pass), down, yards to go, environmental factors, involved players, outcomes, etc.
- **Pro Football Focus (PFF) Grades:** Providing contextual player performance data to assess the influence of offensive lines on play outcomes.

Methodological Approach

The research follows a structured approach to model development and statistical analysis:

1. **Data Preparation:**
    - Filter play-by-play data to include only first, second, and third-down plays within a neutral game situation.
    - Group first and second down plays to categorize playcalling combinations: run-run, pass-pass, pass-run, run-pass.
    - Record third-down distance, conversion outcome, and points per game following each combination.
2. **Exploratory Data Analysis (EDA):**
    - Descriptive statistics and visualizations to examine the frequency and distribution of each playcalling sequence.
    - Initial analysis of third-down average distance to go and conversion rates following different playcalling strategies.

Specifications

This analysis will be conducted using R, with data sourced from NFLFastR, PFF, and Pro Football Reference. The study will focus on neutral situations (16-point spread or less) and examine every first and second-down combination (run-run, pass-pass, pass-run, run-pass), and the corresponding third-down situation and outcome that follows the early down playcalling setup. Visualizations to be produced to compare the outcomes of different playcalling strategies.

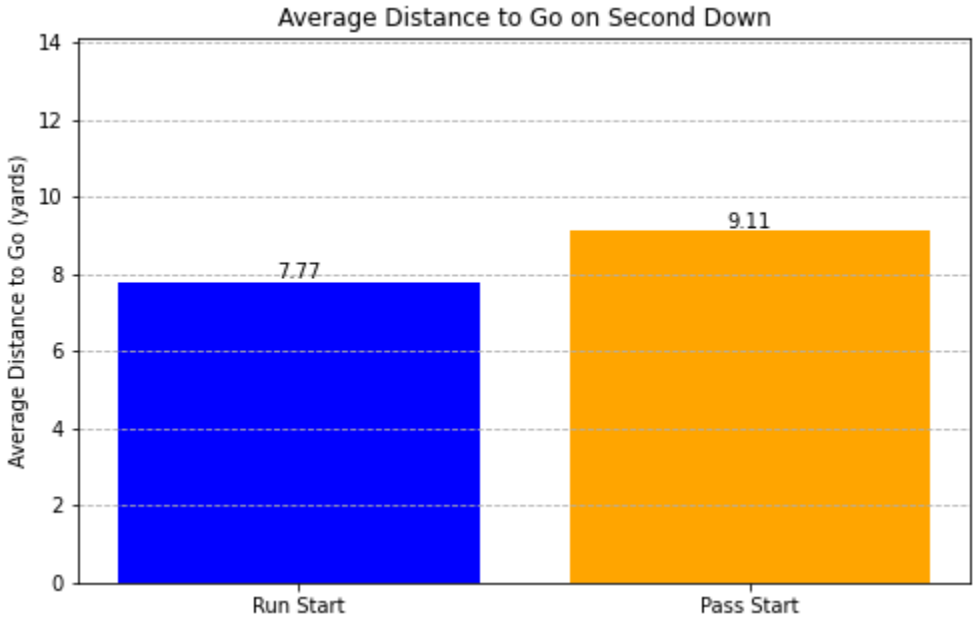
Applications and Sport Impact

This research aims to offer actionable insights for coaches and analysts in the NFL, helping to optimize early-down playcalling for better third-down outcomes. By identifying the most effective playcalling strategies in neutral game situations, teams can enhance their offensive decision-making, ultimately improving drive sustainability, scoring potential, and win rates. This project contributes to the growing importance of data-driven strategies in football, with a particular focus on early-down success and its impact on overall team performance through the lens of sustaining drives offensively.

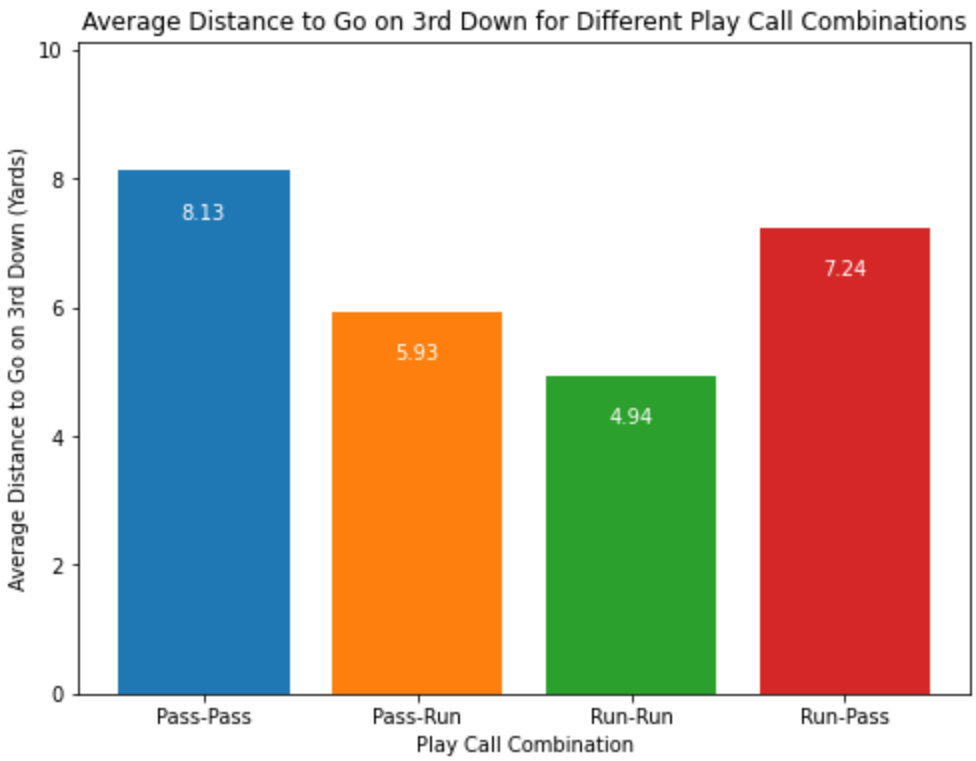
Step-by-Step Process

1. **Data Acquisition:**
- Retrieved comprehensive play-by-play data from the nfl_data_py library, encompassing the 2021, 2022, and 2023 NFL seasons.
2. **Filtering Criteria:**
- The dataset was narrowed to include only plays from "neutral game situations" (defined as games with a spread of 16 points or less), focusing exclusively on first, second, and third downs.
3. **Series Identification:**
- Filtered for series that included consecutive 1st, 2nd, and 3rd-down plays, ensuring each sequence was complete to support full situational analysis.
4. **Play Call Segmentation:**
- Sorted the data based on the first-down play type, dividing the series into those beginning with a pass and those beginning with a run.
5. **Second-Down Combination Classification:**
- Further segmented each group into four categories based on second-down play selection following a first-down play call: Pass-Pass, Pass-Run, Run-Run, and Run-Pass.
- Excluded series with penalty-affected plays to maintain the integrity of distance-to-go data and prevent skewed results.
6. **Third Down Analysis:**
- For each play call combination, calculated the average distance to go on third downs and assessed conversion rates, providing insights into how early down choices influence third-down scenarios.
7. **Team-Specific Play-Calling Tendencies:**
- Calculated team-level tendencies by determining the frequency of each play call combination for each team.
- Converted these frequencies into percentages, allowing for a comparative analysis of team play-calling strategies across different combinations.
8. **Visualization of Findings:**
- Developed visualizations to illustrate average third-down distances and third-down conversion rates across play call combinations.
- Produced a stacked bar chart for team play-call tendencies, showing each combination as a percentage and clearly comparing play-calling patterns among teams.

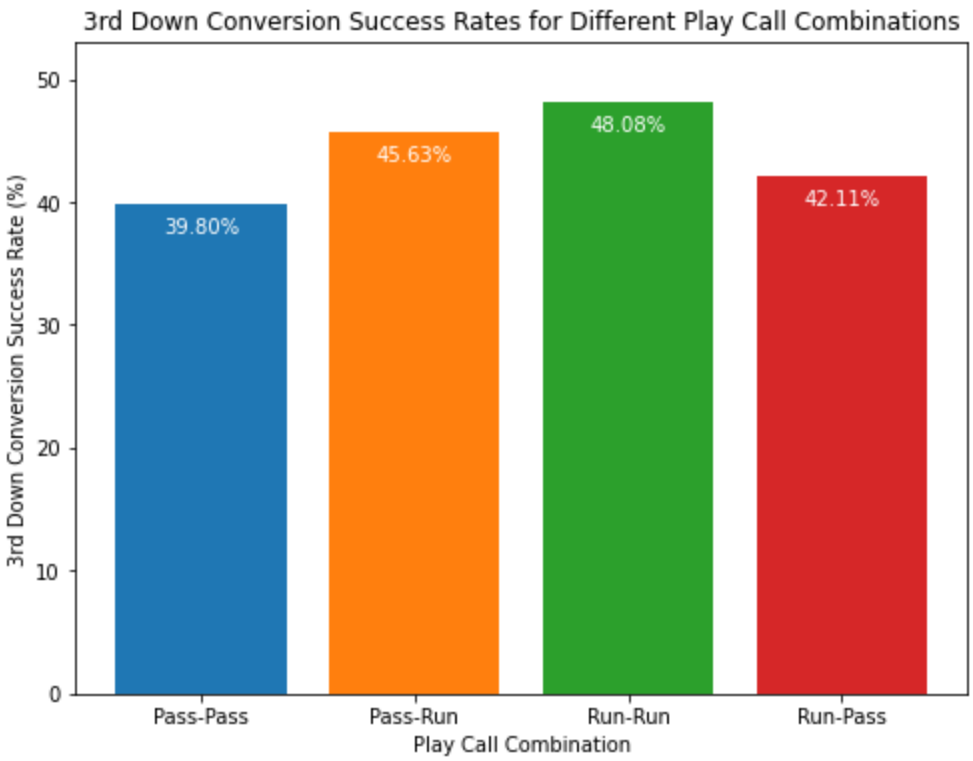
Analysis Results



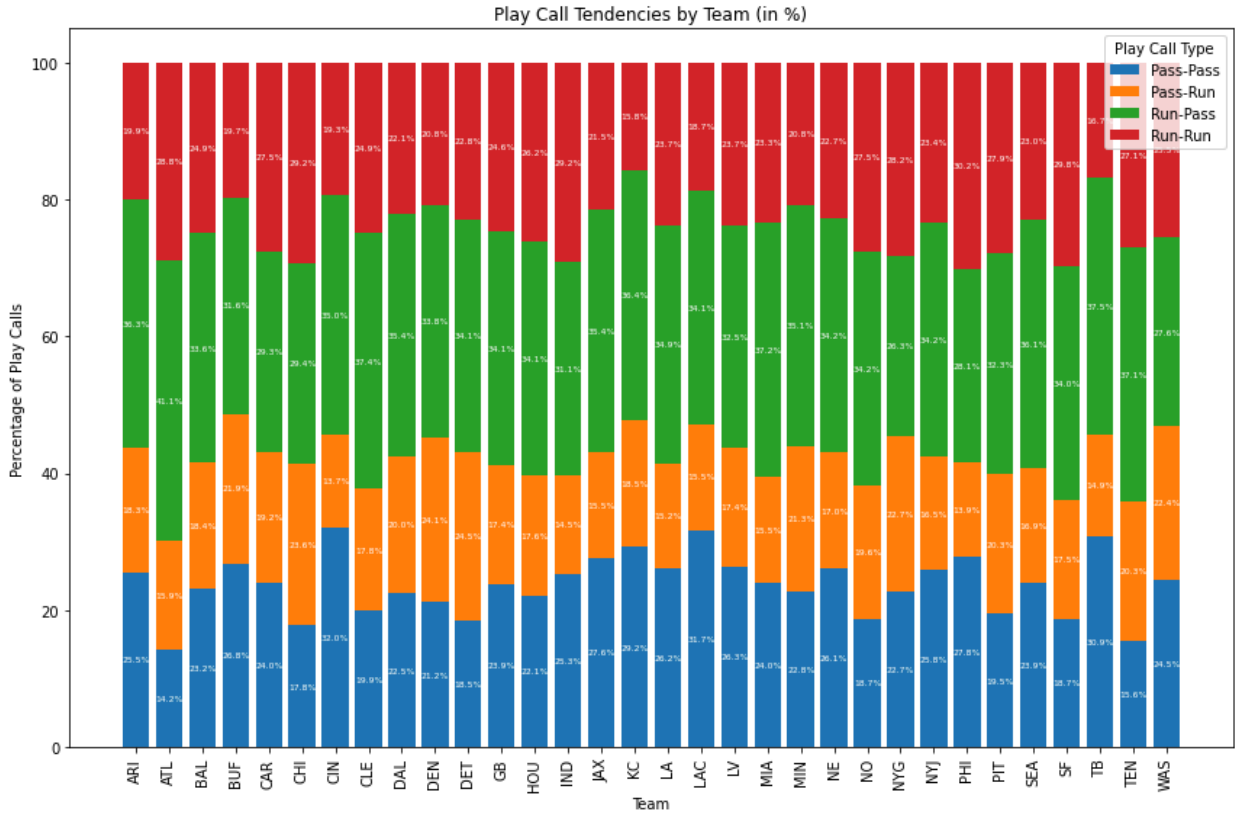
Detailed above is the head-to-head comparison of the average distance to go after starting a series with a pass vs a run play. As shown, starting a series with a run play ended up providing a significant advantage over a pass start, leading to teams facing 1.34 fewer yards to go on second down versus starting with a pass. This backs up the timeless adage of coaches looking to “establish the run”, evidencing the advantages of having a strong run game setting up a team for success on a given set of downs.



The graphic above shows the distance to go based on different playcalling combinations from first and second down plays. The resulting average distances to go provide further evidence of the "establish the run" agenda, showing that when remaining game time permits, even when down by two possessions, teams should still be running the ball at some point every series as it leads to shorter third down distances.



Listed above is an analysis run of the effectiveness of each play call combination in terms of third down conversion rates following the two play combination setup on first and second down. When comparing this to the previously listed graphic of average distances to go, a strong correlation can be seen between the two, with two down setups involving a run play being more effective than purely passing.



The above graphic shows the playcalling tendencies of teams over the past three seasons on first and second-downs, with teams favoring opening series with run plays. This approach of teams backs up what this study evidences as the superior approach as NFL teams, and more importantly, their play-callers, have keyed into the expected elevated success by establishing the run in order to be more successful on each series of downs.

Research Summary

This research project delves into the critical role of early-down playcalling in the NFL, specifically examining how first and second-down combinations influence third-down outcomes. By analyzing playcalling sequences—run-run, pass-pass, pass-run, and run-pass—this study establishes a framework to evaluate their impact on key performance metrics in the form of average third-down distance to go and conversion rates within neutral game situations (games with a point spread of 16 or fewer).

Utilizing comprehensive play-by-play data from NFL_data_py, this analysis employs both linear and logistic regression models to derive actionable insights. Key findings indicate that initiating a series with a run play results in a significantly shorter average distance to go on subsequent downs, reinforcing the longstanding strategic emphasis on establishing the run. The research also highlights strong correlations between specific playcalling tendencies and third-down conversion

rates, underscoring the importance of early-down decision-making in enhancing overall offensive effectiveness.

Future Applications

The implications of this research extend beyond theoretical insights, providing practical applications for NFL coaches and analysts. By identifying the most effective playcalling strategies in early downs, teams can refine their offensive game plans to maximize scoring potential and improve win rates by more effectively sustaining drives offensively. This data-driven approach encourages teams to prioritize run plays in their initial down sequences, which has been shown to lead to shorter third-down distances and higher conversion success as a result of the shorter distance to go.

Moreover, the findings can inform training and preparation methods, guiding coaches in developing playbooks that leverage successful early-down strategies. As the NFL continues to embrace analytics, this research contributes to the evolving landscape of football strategy, emphasizing the need for teams to adopt evidence-based methodologies in their decision-making processes.

Future research can build on these findings by exploring the impact of situational variables—such as game context, opponent tendencies, and player performance—on playcalling effectiveness, further enriching the understanding of strategic dynamics in professional football.