From Data to Decision: Tactical Modeling for Optimizing Competitive Performance in College Soccer

DOI: 10.26573/2025.19.3.1
Volume 19, *Number* 3
September 2025, *pp.* 133-151

Yohann Mauger
Sima M. Fortsch
Northern Kentucky University
(maugery1@nku.edu)
(fortschs1@nku.edu)
Elena A. Khapalova
The State University of New York at Canton
(khapalovae@canton.edu)

We examine internal factors influencing success in NCAA men's college soccer, analyzing data from 62 teams in the 2023 national tournament. Three hypotheses were tested through regression and cumulative probability analyses to evaluate how tactical variables affect NCAA tournament, conference, and nonconference games. Findings show these variables explained 99% of tournament wins, 88% of conference wins, and 77% of nonconference wins. Possession was critical in conference play, while transitions and restarts had a greater impact in nonconference matches. Counterattacks delivered quick results, whereas sustained possession led to gradual gains. These findings offer strategic guidance for coaches and athletic decision-makers.

Keywords: Sport, Soccer, NCAA Competition, Strategic Decision, Data-Driven Decision Regression, Cumulative Probability Function

1. Introduction

Soccer is among the most popular and widely followed sports worldwide, and its popularity continues to grow, prompting researchers and analysts to examine soccer data (Boeker and Midoglu, 2023). In college soccer games, coaches, analysts, and players must rely on effective strategies to succeed. Talent and teamwork are crucial, but the tactical approach—focusing on possession, aggressive defense, quick transitions, or restarts—can greatly influence a game's outcome. While many scholars have studied soccer data to identify winning strategies, we have not found recent research exploring the potential connections or differences between NCAA, conference, and nonconference matches, nor studies measuring the effectiveness of a tactical approach over time. This presents a valuable opportunity to investigate how match type affects team strategies, player performance, game results, and whether these strategies remain effective. Understanding these differences can provide insights into how a winning strategy operates during an NCAA tournament or the familiarity of opponents in a league-like competition (conference). Coaches can make better decisions that influence athletic training programs if any differences between tournaments or diminishing returns of a tactical approach are identified. Therefore, we are motivated to address this gap.

The nonconference games offer a valuable strategic advantage for university soccer teams, providing a controlled yet competitive environment to assess strengths, identify weaknesses, and fine-tune tactics before conference play starts. From a national ranking perspective, nonconference matches impact RPI (Ratings Percentage Index) ratings, which truly excite coaches because nonconference success helps teams gain visibility beyond their local area and get national recognition. Now winning a conference tournament is a key outcome in competitive advantage as a conference title guarantees automatic qualification for the NCAA tournament, removing the uncertainty of selection committees and proving the team's ability to perform consistently against familiar opponents. A championship victory boosts the team’s legacy, enhances recruiting efforts, and gives players a legitimate reason to celebrate. Ultimately, winning the NCAA tournament is the highest achievement in college soccer. From an athletic standpoint, success in the NCAA brings the program into the national spotlight and raises its prestige.

Analysts and coaches focus heavily on identifying patterns in winning strategies, but there are few academic studies on this topic. Therefore, we collected extensive data from U.S. college soccer games, covering one season in 2023 and involving the 62 teams that qualified for the NCAA tournament. Our research seeks to find a dominant tactic that could lead to successful outcomes across different matches: NCAA tournament, conference, and nonconference games. Specifically, we develop and test several hypotheses to determine whether teams prefer defense over offense, or possession over transition and restart, to win. Additionally, we aim to understand the relationships between teams’ past performances and whether a successful tactical approach has diminishing returns.

Section 1 introduces the topic of our study, including the motivation, the literature gap, and our approach. Section 2 provides a relevant literature review. Sections 3 and 4 discuss the data and outline our research methodology, respectively. Section 5 presents our performance measures. The results are discussed in Section 6, while Sections 7 and 8 present our conclusions and the study’s limitations.

2. Literature Review

Analyzing soccer data has been a topic of interest for researchers, coaches, and sports scientists for decades. Numerous studies have aimed to understand the relative impact of game strategies on outcomes. One of the early studies was conducted by Orlick and Partington (1988), who examined 235 Canadian Olympic athletes who competed in the 1984 Olympic Games in Sarajevo and Los Angeles, identifying common elements of success and factors that hindered optimal performance. Hughes and Franks (2005) analyzed World Cup data and found that, for successful teams, longer passing sequences resulted in more goals per possession than shorter ones, while for unsuccessful teams, neither tactic had a clear advantage. Carling et al. (2005) studied numerous soccer matches, highlighting the use of innovative technologies to bridge the gap between research, theory, and practice. Lago-Peñas et al. (2010) analyzed men’s soccer competitions to identify which game-related statistics distinguish between winning, drawing, and losing teams.

Recent studies, such as Schoenfeld and Krieger (2019), have shown that training frequency is an important factor in disciplined resistance exercise. No significant differences were found in training frequencies across all categories: resistance, upper body, and lower body exercises. Bravo et al. (2021) studied the Southern Methodist University (SMU) men’s soccer team to evaluate the plays most linked to victories. The authors focused on individual players’ performances, creating team-level metrics and visualizations to improve coaching efficiency. Additionally, Wu et al. (2021) found that crossing the ball in soccer has a long history as an effective tactic for scoring goals, although the benefit of this move has recently been questioned. García-Aliaga et al. (2021) analyzed several on-field plays to see if player positions could be identified from their statistical performances.

Forcher et al. (2022) examined research papers on soccer and found that, compared to the variety and detailed analyses of offensive play, fewer studies focus on defensive play. However, the authors noted that in recent years, more studies on defensive play have been published. Later, Shen et al. (2022) analyzed soccer data and discovered that pace varies significantly—often highest in the offensive third—while remaining fairly consistent across leagues and increasing as team quality decreases. Cho et al. (2022) studied players' passing styles and improved accuracy, revealing that pass location, length, and direction are key factors. Similarly, Gedik and Egilmez (2023) used machine learning models to predict game outcomes. Rico-González et al. (2023) found that predictive statistical models based on decision-making and scientific evidence often lack strong predictive ability. The authors recommended using maximum likelihood (ML), but warned that ML models rely heavily on dataset size for accurate predictions. Lastly, Verma and Shrivastava (2024) offered a comprehensive review of sports analytics in soccer, highlighting key trends in the literature and noting a shift toward advanced statistical techniques, machine learning, and deep learning applications.

While many of the above studies focus on identifying tactics that lead to winning soccer games, our research further explores the differences among NCAA tournament, conference, and nonconference matches, as well as the long-term effects of a tactical approach—an area that has not yet been studied.

3. Data and Variables

Data for this research was obtained from Wyscout (now owned by Hudl), the world's most extensive library of soccer videos and data. The data is available to researchers at the following website: https://www.hudl.com/products/wyscout. This platform collects videos from millions of games worldwide and calculates the performance of teams and individual players based on a specific dataset through its proprietary data analytics. We selected the 62 men's soccer teams that qualified for the NCAA tournament during the fall of 2023. The matches analyzed included conference, nonconference, and NCAA games. A total of 938 games were included in this study, averaging 15 games per team over the season, with 94 variables considered.

After careful consideration—guided by one of our coauthors who is connected with several coaches and regularly helps them develop winning strategies—we narrowed our variables to those that directly affect our goal. From the original 94 variables, we selected 15. Three variables were chosen for each of the five categories we studied. Y1 (Offense) includes the number of goals (GOALS), expected goals (xG), and the percentage of shots on target (PSOT). Y2 (Defense) contains the total number of interceptions (INT), total goals conceded (TGAG), and total defensive duels won (TDFEDW). Y3 (Possession) incorporates the average number of passes per possession (APASSP), total passes in the final third (TPASFT), and total positional attacks with shots (TPOSAWS). Y4 (Transition) includes the number of counterattacks with shots (COUNTWS), total smart passes (TSPAS), and average pass lengths (AVEPASSL). Finally, Y5 (Restart) features the total number of free kicks with shots (TFKWS), corners with shots (CORNWS), and total offensive duels won (TOFDW). Table I shows the list of our variables and the categories we studied.

Table Ia: *Variables and their Definitions*

Variable	Definition
APASSP	Average passes per possession
AVEPASSL	Average pass length
CORNWS	Corners with shots
COUNTWS	Counterattacks with shots
GOALS	Number of Goals
INT	Total interceptions
PSOT	Percentage shot on target
TDEFDW	Total defensive duels won
TFKWS	Total free kicks with shots
TGAG	Total conceded goals
TOFDW	Total offensive duels won
TPASFT	Total passes final third
TPOSAWS	Total Positional attacks / with shots
TSPAS	Total smart passes
XG	Expected goals
Dependent variable: WIN. Independent variables: Y1 – Y5. Y1 (Offense)= GOALS xG PSOT Y2 (Defense)= INT TGAG TDEFDW Y3 (Possession)= APASSP TPASFT TPOSAW Y4 (Transition)= COUNTWS TSPAS AVEPASSL Y5 (Restart)= TFKWS CORNWS TOFDW	

To evaluate the degree of correlation among our chosen variables, we analyzed the correlation matrix. As shown in Table Ib, only one pair—Goals and Expected Goals (XG)—shows a moderate correlation (0.63). The other variables either have weak (less than 0.50) or statistically insignificant correlations (less than 0.30). Therefore, multicollinearity is not an issue in our analysis.

Table Ib *Correlation Matrix*

Non-Conference	*GOALS*	*xG*	*PSOT*	*TGAG*	*TDEFDW*	*INT*	*COUNTWS*	*TSPAS*	*AVEPASSL*	*CORNWS*	*TFKWS*	*TOFDW*	*APASSP*	*TPASFT*	*TPOSAWS*
GOALS	1														
xG	0.63	1													
PSOT	0.41	0.12	1												
TGAG	-0.23	-0.25	-0.10	1											
TDEFDW	-0.06	-0.13	0.03	0.13	1										
INT	-0.04	-0.13	0.09	0.06	0.23	1									
COUNTWS	0.30	0.27	0.07	-0.07	-0.03	0.13	1								
TSPAS	0.17	0.24	0.08	-0.10	-0.04	-0.02	0.12	1							
AVEPASSL	-0.05	-0.09	-0.01	-0.04	0.00	0.03	0.00	-0.19	1						
CORNWS	0.24	0.44	-0.08	-0.16	-0.15	-0.17	0.05	0.15	-0.03	1					
TFKWS	0.03	0.22	-0.08	0.04	0.02	-0.01	0.03	0.05	0.10	0.08	1				
TOFDW	0.28	0.39	-0.01	-0.16	-0.06	-0.12	0.15	0.21	-0.20	0.21	0.08	1			
APASSP	-0.12	-0.20	0.02	0.21	0.08	0.08	-0.03	-0.06	-0.05	-0.14	-0.05	-0.29	1		
TPASFT	-0.17	-0.30	0.03	0.19	0.10	0.27	0.02	-0.17	0.17	-0.22	-0.04	-0.33	0.52	1	
TPOSAWS	-0.18	-0.25	-0.03	0.33	0.15	0.25	-0.02	-0.06	0.01	-0.19	-0.03	-0.25	0.24	0.44	1

4. Research Methodology

This paper aims to identify an effective tactic that increases the chances of winning in NCAA tournament, conference, and nonconference matches. Of course, winning a soccer game often depends on a mix of factors, many of which can be completely outside a team's control. These external factors include bad weather, referees, unpredictable opponents' performances, injuries to key players at the wrong time, unruly fans, poor field conditions, travel fatigue, media distractions, and more. Similarly, our data does not account for important internal factors that can determine a match's outcome, such as team chemistry, physical conditioning, players' focus, mental toughness, discipline, practice, and motivation. Given this mix of unmeasurable internal and external factors, we fully expect some level of uncertainty and error in our analysis. Soccer, after all, is as much about art and luck as it is about science and stats. However, we collected data on several measurable internal factors listed in Table I. These variables give us a strong basis for exploring what drives a soccer victory—assuming all other factors remain constant (or at least equally unpredictable). To reach our goal, we propose the following hypotheses.

Hypothesis 1: Teams rely more on defense than offense to win NCAA matches.
Hypothesis 2: In conference and nonconference games, teams use both offense and defense equally to secure a win.
Hypothesis 3: Teams use restarts less frequently than possession and transition strategies to win.

We use ordinary least squares (OLS) for some analyses, while other analyses involve a categorical dependent variable (WIN and not-WIN). Logistic regression is the preferred method for such models, where the dependent variable is binary (1 = win, 0 = not win). Binary models show how input variables affect the outcome by comparing class 1 (win) to class 0 (not win). The outcomes predict probabilities between 0 and 1, which are then transformed into a scale from −∞ to +∞. The equation for the logit (logistic) model is as follows.

$$Logit(P) = \ln\left(\frac{P}{1-P}\right) = \beta_0 + \beta_1 X \qquad \text{Eq. 1}$$

Here, X represents the desired independent variable, and P indicates the probability of our categorical dependent variable. The coefficients are estimated using maximum likelihood estimation (ML). Similar to linear regression, the outcome yields β values that maximize the likelihood of the observed data. We use STATA software to obtain our results. For more detailed information about logistic regression, we refer readers to the works of Das, A. (2024) and Best and Wolf (2015).

5. Performance Measure

We use two performance measures: R^2 and the cumulative probability function. To compare the models, we evaluate the regression outcomes based on R^2. However, when the dependent variable is categorical, we report the Pseudo R^2 (similar to R^2 in OLS). While not the only method for assessing results, it offers valuable insights. For example, it indicates how much of the variance in the dependent variable is explained by the model. It also reflects the percentage of changes in the dependent variable that are related to our selected independent variables. Therefore, the magnitude and changes in R^2 suggest the significance of our independent variables, as they influence the model's explanatory power. Accordingly, we rely on both R^2 and *Pseudo* R^2 to compare different models.

A cumulative probability function shows how the impact of a variable on the dependent variable (WIN) decreases, as the function ranges from 0 to 1 (0 = no impact, 1 = full impact). Because of the logarithmic nature of our model and the binary structure of the dependent variable, we calculate the probability functions using the following equation.

$$P = 1/\left(1 + e^{-(\beta_0+\beta_1 X)}\right) \qquad \text{Eq. 2}$$

Cumulative probability functions help identify which variables provide the most benefit early on. If a variable shows quick improvements in winning chances, it becomes a clear priority. Recognizing where the function levels off allows us to make smarter choices and adopt a more strategic approach, avoiding wasted time and effort once the gains diminish. For example, if increasing a variable from 70% to 90% only increases the chance of winning by 2%, then the time spent on training could be better allocated to developing a different skill. By visualizing diminishing returns, we can shift the question from "Does this skill work?" to "When does it stop being useful?" This graph is called the zone of marginal gains because it shows that the payoff is not guaranteed or becomes less predictable beyond a certain point.

6. Results and Discussions

The first set of results was straightforward: a combination of offense and defense is a significant predictor of winning games, indicating that all our selected variables related to offense and defense are important for forecasting game outcomes (fter all, to win a soccer game, a team needs to score more goals than the opponent, hence Goals and TGAG will be the 2 key variables). Monitoring the t-statistics (or p-values), it is not surprising that in NCAA matches, the most predictive variables with high t-statistics were scored goals (GOALS), total goals against (TGAG), and average passes per session (APASSP). For conference games, total interceptions (INT), GOALS, and TGAG had high t-statistics, while for nonconference matches, GOALS and TGAG emerged as significant predictors. Although some variables did not show strong t-statistics (indicating insignificance), the regression variables were jointly and strongly significant across all models, as evidenced by a strong F-statistic probability (P (F) = 0.000). This confirms that all variables are jointly significant at the 99% confidence level. Therefore, we conclude that all selected variables are important for predicting a win, even though some did not individually demonstrate high t-statistics.

Table II presents the results of the logistic regression, where the dependent variable is categorical (Win vs. Not Win), along with the Pseudo R^2. The results in Table II highlight several key findings. *First,* Hypothesis 1 demonstrates that variations in winning are more strongly explained by defense (Y2) than offense (Y1) in NCAA tournament matches. Table II shows that Y2 has a 17% higher Pseudo R^2 0.384) than Y1 (0.327). Similarly, Hypothesis 2 suggests that variations in winning related to offense and defense are equally matched when it comes to securing wins in conference and nonconference matches. For conference matches, the Pseudo R^2 for defense (Y2) is statistically similar to that of offense (Y1), with defense being 2% higher. For nonconference matches, the Pseudo R^2 for offense is 3% higher than for defense.

Table II. *Logistic Regression Output for Dependent Variable: WIN*

Included Independent Variables	***NCAA***		***Conference***		***Nonconference***	
	Pseudo R^2	R^2 direction	Pseudo R^2	R^2 direction	Pseudo R^2	R^2 direction
All (Y1 – Y5)	**0.992**	-	**0.882**	-	**0.774**	-
Y1	0.327	-67%	0.267	-70%	0.300	-61%
Y2	0.384	-61%	0.273	-69%	0.291	-62%
Y3	0.069	-93%	0.035	-96%	0.049	-94%
Y4	0.051	-95%	0.031	-96%	0.025	-97%
Y5	0.032	-97%	0.015	-98%	0.033	-96%
All variables are jointly significant for all equations. P (F-stat) =0.000. Y1: offense variables; Y2: defense variables; Y3: Possession variables; Y4: Transition variables; Y5: Restart variables.						

Second, although all categories used the same independent variables, our models explain 99% of wins in NCAA tournament games, 88% in conference games, and 77% in nonconference games. This suggests that the chosen variables are more effective in predicting wins for NCAA tournament matches. This can be linked to the nature of NCAA tournament games, which are elimination matches (Cup games) where only one team can advance. When the score is tied after 90 minutes, teams benefit from an extra 30 minutes (two 15-minute periods) to outscore their opponent. As a result, teams have more time to secure a win in NCAA tournament games compared to conference or nonconference games, potentially increasing the predictive power of the variables.

Third, all chosen variables are essential; using only the three variables linked to each specific category (offense, defense, possession, transition, or restart) greatly enhances the model's explanatory power. Omitting variables causes the R^2 to decrease by 67–97% for NCAA tournament games, 69–98% for conference games, and 61–96% for nonconference games, reducing the model's effectiveness in predicting wins.

Fourth, the results in Table II suggest that Y1 (offense) and Y2 (defense) show similar patterns across all match types, indicating they can be exchanged. In contrast, Y3 (possession), Y4 (transition), and Y5 (restart) differ between match types, suggesting they are not interchangeable. This may imply that these categories reflect unique or context-specific aspects of gameplay that are difficult to replicate—meaning Y3, Y4, and Y5 are variables tied to tactical decisions made by the coaching staff, depending on the opponent. The low R^2 score for Y5 (0.032) is expected, as restarts make up a small part of the overall game. On average, over a 90-minute match, teams in our sample committed 13.6 fouls and earned 4.8 corners per game. As is well known, a soccer team can win by playing offense or defense. Restart tactics matter when games are close, prompting teams—such as Arsenal in England—to hire coaches with expertise in restarts to capitalize on critical moments.

While Table II helps test Hypotheses 1 and 2, it does not permit testing Hypothesis 3 because of the high variability in teams' preferences for using possession, transition, or restart as winning strategies. Therefore, we analyzed each team separately to identify which method—possession, transition, or restart—is more commonly used to win. NCAA tournament matches could not be included in the team-level analysis due to insufficient data (several games per team were needed for accurate predictions, and some teams were eliminated after only one match). However, we separated the teams based on conference and non-conference matches. Tables IIIa and IIIb show the regression results for individual teams in conference and nonconference games, respectively. While Table II included results for the 62 teams, some teams were excluded from Tables IIIa and IIIb either due to too few observations (not enough games) or because their records consisted solely of wins or losses, making it impossible to analyze factors contributing to both outcomes. Table IIIa features 47 of the original 62 teams for conference play, while Table IIIb (nonconference games) includes 45 teams. Despite these minor exclusions, the data remaining are sufficient to draw meaningful insights into conference and nonconference performance.

Testing Hypothesis 3 produced mixed results. The data in Tables IIIa and IIIb show that, in conference matches, restart was the least common factor influencing a win compared to possession and transition (supporting Hypothesis 3). In contrast, possession was used less often in nonconference matches than transition and restart (rejecting Hypothesis 3). Table IIIa reveals that in conference matches, teams scored goals through possession in 26 of 47 cases (55%), followed by transition plays in 16 cases (34%), and restarts in only 5 cases (10%). Therefore, in conference matches, possession results in more goals. Conversely, in nonconference matches, teams scored through possession in just 12 of 45 cases (26%), while transition contributed 17 goals (37%) and restarts contributed 16 goals (35%).

We believe this is because teams that qualify for the NCAA tournament are among the best in their conferences; therefore, they tend to dominate opponents by maintaining possession more often. In nonconference games, some teams follow the same tactics as in conference matches, while others adjust their strategies. Table IIIb shows that teams from lower-ranked conferences, when facing higher-ranked opponents, tend to score less through possession (Y3) and more through transition (Y4) due to reduced ball control. As a result, these teams shift from possession to transition. Conversely, teams from higher-ranked conferences facing less challenging opponents tend to increase their possession, as they are more likely to control the ball. Table IIIb indicates that teams do not stick to a single strategy but instead choose the best method to win based on their opponents.

The impact of restart (Y5) on scoring goals seemed linked to the type of opponent faced. Once again, it appeared that when teams played against higher-ranked opponents, the effect of the restart strategy diminished, while against lower-ranked teams, the impact grew. There were noticeable differences among teams in their use of the restart strategy. Some teams, like the Bryant Bulldogs, SC Gamecocks, and Pittsburgh Panthers, frequently emphasized restarts, whereas others, such as the Louisville Cardinals, Georgetown Hoyas, and SMU Mustangs, did not. Top-performing teams tended to prioritize possession, while lower-performing teams relied more on transition and restarts to secure victories. Additionally, the restart strategy was nearly absent in conference games but was as prominent as transition in nonconference matches. However, we lack significant statistical evidence to confirm how restart is used or its exact impact on winning.

Table IIIa *Output for Conference Teams. Dependent Variable: Number Of Goals. (26): Possession was more Significant for GOALS in 26/47. (16): Transition was more Significant in 16/47. (5): Restart was more Significant in 5/47 times*

Conference Teams	***R^2 Per Category***			
	Offense/Defense	**Possession[26]**	**Transition[16]**	**Restart[5]**
Boston Terriers	0.838	0.383	0.226	0.226
Bryant Bulldogs	0.924	0.885	0.797	0.810
California Baptist Lancers	0.964	0.749	0.504	0.414
Charlotte 49ers	0.741	0.327	0.882	0.197
Clemson Tigers	0.883	0.768	0.760	0.459
Dayton Flyers	0.529	0.220	0.448	0.395
Denver Pioneers	0.862	0.703	0.472	0.595
Duke Blue Devils	0.498	0.391	0.831	0.188
FAU Owls	0.998	0.875	0.966	0.697
FIU Panthers	0.448	0.615	0.299	0.050
Georgetown Hoyas	0.828	0.378	0.642	0.102
Green Bay Phoenix	0.824	0.326	0.312	0.551
High Point Panthers	0.883	0.246	0.469	0.416
Hofstra Pride	0.883	0.295	0.551	0.307
Indiana Hoosiers	0.716	0.586	0.362	0.435
JMU Dukes	0.816	0.593	0.453	0.136
Kentucky Wildcats	0.240	0.325	0.468	0.576
LIU Sharks	0.479	0.426	0.380	0.705
LMU Lions	0.860	0.503	0.285	0.287
Louisville Cardinals	0.551	0.312	0.172	0.065
Marshall Thundering Herd	0.608	0.262	0.511	0.096
Memphis Tigers	0.823	0.483	0.665	0.428
Missouri State Bears	0.798	0.685	0.250	0.260
New Hampshire Wildcats	0.882	0.604	0.447	0.110
North Carolina Heels	0.682	0.378	0.242	0.185
Notre Dame Fighting Irish	0.834	0.850	0.682	0.124
Oregon State Beavers	0.667	0.429	0.178	0.293
Pittsburgh Panthers	0.965	0.209	0.549	0.601
Portland Pilots	0.999	0.405	0.864	0.662

Rider Broncs	0.613	0.625	0.215	0.346
San Diego Toreros	0.985	0.907	0.523	0.104
Seattle Redhawks	0.905	0.641	0.217	0.479
SIUE Cougars	0.597	0.540	0.513	0.665
SMU Mustangs	0.854	0.319	0.120	0.070
SC Gamecocks	0.999	0.893	0.518	0.868
Stanford Cardinal	0.721	0.617	0.275	0.152
Syracuse Orange	0.634	0.241	0.120	0.188
UC Irvine Anteaters	0.893	0.238	0.209	0.039
UCF Knights	0.287	0.445	0.237	0.203
UCLA Bruins	0.739	0.322	0.578	0.285
Vermont Catamounts	0.713	0.671	0.826	0.439
Virginia Cavaliers	0.974	0.443	0.603	0.291
Wake Forest Demon Deacon	0.587	0.590	0.206	0.152
West Virginia Mountaineers	0.607	0.181	0.672	0.580
Western Michigan Broncos	0.675	0.266	0.370	0.238
Xavier Musketeers	0.653	0.585	0.580	0.401
Yale Bulldogs	0.550	0.363	0.535	0.241

Table IIIb *Output for Nonconference Teams. Dependent variable: number of goals. (12): Possession was more significant for GOALS in 12/45. (17): Transition was more significant in 17/45. (16): Restart was more significant in 16/45 times*

Nonconference Teams	***R^2 Per Category***			
	Offense/Defense	**Possession[12]**	**Transition[17]**	**Restart[16]**
Boston Terriers	0.936	0.633	0.270	0.181
Bryant Bulldogs	0.336	0.410	0.122	0.242
California Baptist Lancers	0.680	0.081	0.950	0.840
Charlotte 49ers	NCR	0.811	0.515	0.779
Clemson Tigers	0.495	0.220	0.743	0.367
Dayton Flyers	0.898	0.718	0.510	0.398
Denver Pioneers	0.622	0.072	0.283	0.193
Duke Blue Devils	0.963	0.411	0.821	0.829
FAU Owls	Excluded for insufficient data			
FIU Panthers	0.885	0.609	0.443	0.499
Georgetown Hoyas	0.625	0.105	0.783	0.802
Green Bay Phoenix	0.999	0.444	0.920	0.725
High Point Panthers	0.848	0.516	0.291	0.366
Hofstra Pride	0.608	0.252	0.547	0.626

Indiana Hoosiers	0.836	0.666	0.547	0.619
JMU Dukes	0.846	0.313	0.806	0.295
Kentucky Wildcats	0.372	0.303	0.769	0.245
LIU Sharks	0.762	0.511	0.276	0.286
LMU Lions	0.933	0.216	0.319	0.598
Louisville Cardinals	0.928	0.756	0.781	0.249
Marshall Thundering Herd	0.991	0.401	0.720	0.883
Memphis Tigers	NCR	0.855	0.916	0.829
Missouri State Bears	0.631	0.509	0.637	0.551
New Hampshire Wildcats	0.778	0.173	0.765	0.279
North Carolina Heels	0.997	0.995	0.599	0.866
Notre Dame Fighting Irish	0.842	0.454	0.442	0.827
Oregon State Beavers	NCR	0.489	0.528	0.624
Pittsburgh Panthers	0.971	0.646	0.650	0.342
Portland Pilots	0.757	0.183	0.358	0.438
Rider Broncs	NCR	0.652	0.335	0.653
San Diego Toreros	0.351	0.122	0.279	0.459
Seattle Redhawks	0.236	0.266	0.184	0.278
SIUE Cougars	0.961	0.946	0.353	0.461
SMU Mustangs	0.868	0.325	0.790	0.860
SC Gamecocks	Excluded for insufficient data			
Stanford Cardinal	0.861	0.252	0.474	0.473
Syracuse Orange	0.959	0.280	0.332	0.727
UC Irvine Anteaters	0.823	0.718	0.459	0.625
UCF Knights	NCR	0.476	0.683	0.800
UCLA Bruins	0.955	0.944	0.880	0.678
Vermont Catamounts	0.564	0.461	0.483	0.729
Virginia Cavaliers	0.724	0.327	0.191	0.676
Wake Forest Demon Deacon	0.822	0.473	0.775	0.248
West Virginia Mountaineers	0.682	0.244	0.698	0.204
Western Michigan Broncos	0.994	0.253	0.769	0.191
Xavier Musketeers	0.991	0.740	0.954	0.834
Yale Bulldogs	0.923	0.478	0.479	0.147

Table IV shows the odds of winning, calculated as PR (Win) / PR (Not Win). It illustrates how the odds of winning change when the left-side variables increase by one unit. When the values are close to one, the effect is minimal; when they are greater than one, the odds increase; and when they are less than one, the odds decrease. For example, a one-unit increase in the number of goals results in the odds of winning rising by 8.19 times (not 8 percent) in NCAA tournaments, 7.20 times in conference games, and 7.75 times in nonconference matches. These findings are expected, given that scoring goals significantly influences winning, especially in NCAA games. Conversely, when total goals scored against a team (TGAG) increase by one unit, the odds of winning drop markedly—by a factor of 9 (0.09) in NCAA games, 11 (0.11) in conference games, and 9 (0.09) in nonconference games. This reflects a strong negative impact across all match types.

All variables associated with the offense equation (Y1: number of goals (GOALS), expected goals (xG), and percent shot on target (PSOT)) have odds ratios greater than one across all match types, signifying that offense significantly affects the odds of winning. Among these, the number of goals and expected goals exert the strongest influence. In contrast, PSOT has a more moderate impact, with odds ratios of approximately 1.04, 1.02, and 1.02 for NCAA, conference, and nonconference matches, respectively. Furthermore, all offensive variables have a slightly greater impact in NCAA matches compared to the other two categories.

On the defensive side, the selected variables (Y2: total interceptions (INT), total goals conceded (TGAG), and total defensive duels won (TDEFDW)) show varying impacts on the odds of winning. INT has minimal effect when its value increases by one, with odds ratios of 1.08 in NCAA matches, 1.04 in conference games, and 1.02 in nonconference games. However, both TGAG and TDEFDW significantly decrease the odds of winning when their values increase by one. Specifically, when the number of goals conceded rises by one, the odds of winning decrease by 11 times (0.11) in conference games and by 9 times (0.09) in both NCAA and nonconference matches. Conversely, an increase of one unit in total defensive duels won results in a slight negative effect on the odds of winning across all tournaments—0.98 in NCAA, and 0.99 in both conference and nonconference matches.

Defense is the slightly more influential factor! The NCAA's most impactful variables are GOALS (offense) and TGAG (defense). An increase of one unit in GOALS can raise a team's chances of winning by eight times; however, increasing TGAG by one unit against a team can decrease its chances of winning by nine times. Overall, the findings are expected—teams need to score more goals than their opponents to win, making GOALS and TGAG the most significant variables that influence the odds of winning in opposite directions. However, Table IV also highlights other important variables, such as the probability of making a shot that results in a score (expected goals: xG), the use of quick transitions (COUNTWS), and a restart strategy involving corner shots (CORNWS) and free kicks (TFKWS). These variables show a similar impact across all matches, while scoring goals has a greater effect on winning NCAA games and, by extension, on expected goals.

Another key finding from Table IV is that the external defense variable (TGAG) has a more significant effect on the chances of winning than the internal variables (INT and TDEFDW). Total interceptions (INT) are mainly influenced by players' anticipation, positioning, and defensive awareness—reflecting the team's effort, tactics, skill, and execution. Similarly, total defensive duels won (TDEFDW) depend on physical and tactical skills shaped by players' aggression, technique, and teamwork—factors largely guided by coaching strategies and training. Therefore, these two variables are considered internal (within the team's control). In contrast, the total goals against (TGAG) variable— which has a strong and negative impact on the chances of winning when its value increases by one unit—can be viewed as external, affected by factors such as the opponent's skill, refereeing decisions, penalties, and more.

Table IV: *The odds of Winning for an Increase of one unit for each Variable*

Variables	NCAA		Conference		Nonconference	
	β	Odds of Winning	β	Odds of Winning	β	Odds of Winning
Y1: Offense						
GOALS	2.103	8.19	1.974	7.20	2.048	7.75
XG	0.503	1.65	0.220	1.25	0.5894	1.80
PSOT	0.036	1.04	0.018	1.02	0.018	1.02
Y2: Defense						
INT	0.075	1.08	0.039	1.04	0.016	1.02
TGAG	-2.407	0.09	-2.193	0.11	-2.357	0.09
TDEFDW	-0.021	0.98	-0.012	0.99	-0.014	0.99
	β	Odds of Scoring	β	Odds of Scoring	β	Odds of Scoring
Y3: Possession						
APASSP	0.096	1.10	0.106	1.11	0.047	1.05
TPASFT	0.006	1.01	0.005	1.01	0.009	1.01
TPOSAWS	0.118	1.12	0.132	1.14	0.040	1.04
Y4: Transition						
COUNTWS	0.348	1.42	0.400	1.49	0.485	1.62
TSPAS	0.071	1.07	0.031	1.03	0.079	1.08
AVEPASSL	0.111	1.12	0.028	1.03	0.036	1.04
Y5: Restart						
TFKWS	0.125	1.13	0.153	1.16	0.095	1.10
CORNWS	0.129	1.14	0.229	1.26	0.264	1.30
TOFDW	0.002	1.00	0.026	1.03	0.391	1.48

β: beta coefficient.
Odd: numbers > 1 indicate the variable increases the odd of winning, < 1 shows it decreases the odd, and ≈ 1, signifies the variable does not change the odd of winning significantly, but has moderate impact.

The results of the cumulative functions are summarized in Tables IVa through IVc. These functions highlight the fastest (or slowest) variables in terms of their impact on winning a game. The vertical axes represent the impact of a variable, while the horizontal axes show time. Variables with the quickest impacts are displayed by graphs that reach one (100% impact) rapidly, whereas slower impacts are represented by more gradual, linear graphs that reach one slowly.

Possession variables. Starting with Table Va, which lists our selected possession variables, we found that teams can quickly influence winning through APASSP (average passes per session) in NCAA games, followed by nonconference games. However, in conference matches, APASSP has a slower effect. We believe this is because opponents in conference games are more likely to use counterattack strategies, which reduce the impact of APASSP. Conversely, TPOSWS (total positional attacks with shots) has the fastest impact on winning in conference matches, although its effect in NCAA games is somewhat slower. Among all three possession variables, the slowest response is from TPASFT (total passes in the final third) across all match types, likely because it takes longer to break down the opponent's defense. Still, although TPASFT's impact on winning is slower, it also suggests that the effort invested in this variable continues to pay off over time—longer than the effects of APASSP and TPOSWS—indicating that its influence is sustained rather than fleeting.

Transition variables. Table Vb summarizes the cumulative functions for our selected transition variables. First, we observed that among all variables (possession, transition, and restart), COUNTWS (counterattacks with shots) has the quickest impact on winning a game. However, this effect also diminishes the fastest, after 4 occurrences in NCAA matches and 5 occurrences in conference and nonconference matches. These results are consistent across all match types. The second quickest impact among the transition variables comes from AVEPASSL (average pass length), which has the second-fastest effect on winning in NCAA games but is somewhat slower in conference and nonconference matches. Finally, the slowest impact is from TSPAS (total smart passes). While its effect is not immediate, it is continuous and slowly diminishing, with the quickest impact observed in nonconference games.

Restart variables. The cumulative functions for our selected restart variables are shown in Table Vc. The fastest impact on winning comes from CORNWS (corners with shots) in nonconference games, followed by conference and NCAA matches. While TFKWS (total free kicks with shots) has a slower impact on winning in NCAA matches, it produces a quicker effect in conference games, followed by nonconference matches. The slowest return in terms of impact on winning is attributed to TOFDW (total offensive duels won), as its graph is nearly linear. However, this also suggests that although its impact is among the slowest, it continues to yield benefits over a longer duration compared to TFKWS and CORNWS.

7. Conclusion

Based on our research findings, it is clear that both offensive and defensive strategies are vital in predicting NCAA tournament game outcomes. The logistic regression analysis showed that defense is slightly more important than offense in determining wins in NCAA tournament games, supporting Hypothesis 1. In contrast, Hypothesis 2 was also confirmed, indicating that teams tend to balance offense and defense relatively equally in both conference and nonconference games. Therefore, Hypothesis 1 is supported, though only marginally. Additionally, the strength of our models—particularly the NCAA model, which accounts for 99% of game outcomes—underscores the overall importance of all selected variables. Simplifying these models by removing any category (offense, defense, possession, transition, or restart) caused significant drops in explanatory power, highlighting the need for a comprehensive approach to performance analysis.

Table Va. Possession: Cumulative Probability Graphs (ROX).

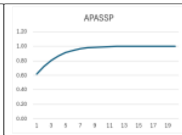
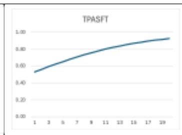
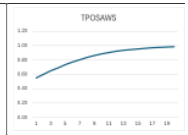
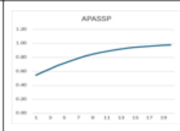
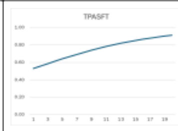
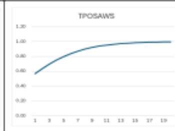
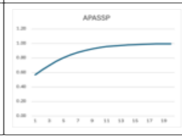
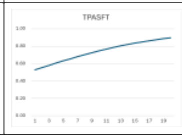
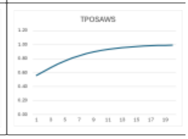
NCAA			
Conference			
Nonconference			

Table Vb. Transition: Cumulative Probability Graphs (ROX).

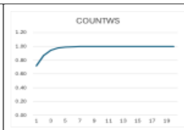
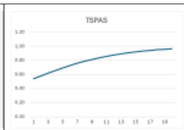
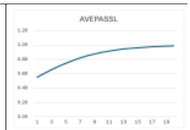
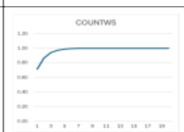
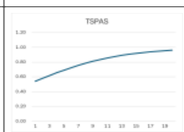
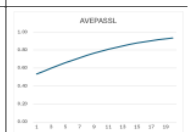
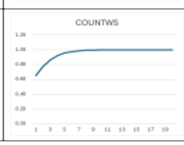
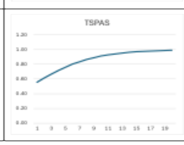
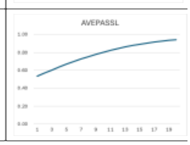
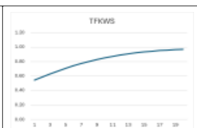
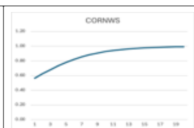
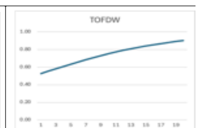
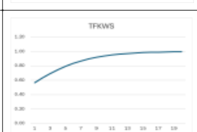
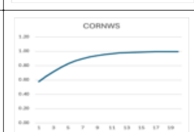
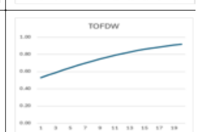
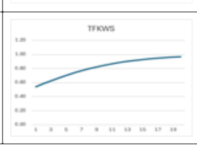
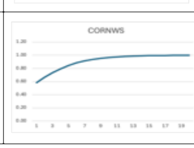
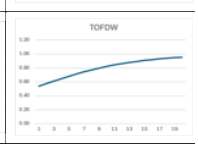
NCAA			
Conference			
Nonconference			

Table Vc. Restart: Cumulative Probability Graphs (ROX).

NCAA			
Conference			
Nonconference			

While Hypothesis 3 yielded mixed results, it offered valuable insights into team strategies. In conference games, possession and transition methods were more often linked to winning than the restart strategy, supporting Hypothesis 3. However, in nonconference matches, possession led to wins less frequently than transition and restart strategies, leading to the rejection of Hypothesis 3. These contrasting patterns emphasize the variability in team preferences, especially regarding the restart strategy, which was used inconsistently and showed little statistical evidence of impact.

The analysis of cumulative functions showed that average passes per session (APASSP) significantly influenced NCAA tournament and nonconference wins, while counterattacks with shots (COUNTWS) and corners with shots (CORNWS) were the quickest routes to scoring success in transition and restart strategies, respectively.

Overall, the results highlight the dynamic and context-dependent nature of winning strategies in college sports. Teams that combined both offensive and defensive tactics while remaining adaptable in their use of possession, transition, and restart strategies were better able to secure consistent wins. These findings have practical implications for coaching, suggesting a data-driven approach to optimizing specific variables that can produce quick and steady improvements in performance. Our findings, therefore, could help shape coaching strategies to win games. However, we stress that a winning approach also depends on the team's player profiles and their ability to maintain possession over long periods. Coaches need to decide whether to adjust their strategy based on the opponent or stick to their own plan, depending on their players' strengths.

General Contribution Beyond the NCAA. While this study centers on NCAA men's college soccer, its analytical framework and findings have broad application beyond sports—especially when viewed through the lens of strategic management theory. For instance, the Resource-Based View (RBV) states that a sustained competitive advantage comes from an organization's ability to leverage resources that are valuable, rare, inimitable, and non-substitutable. Our identification of tactical approaches that produce either quick or long-lasting benefits reflects this principle, highlighting specific capabilities—such as efficient operations or superior customer engagement—that can act as strategic differentiators. Just as soccer teams customize their tactics to match the strengths of their players, companies can align their strategic decisions with employee capabilities to enhance performance. Our findings also support the concept of Dynamic Capabilities, which refers to an organization's capacity to integrate, develop, and reconfigure internal and external skills in response to changing conditions. The time-dependent effects we observed in tactical performance can be directly applied to business scenarios—such as choosing adaptive strategies (e.g., agile innovation or responsive marketing) that offer the greatest and most enduring impact. In soccer, coaches adjust tactics based on both their team's strengths and those of their opponents. Similarly, in business, strategic choices should consider internal capabilities and awareness of competitors' moves. Knowing a rival's strategy—and a company's ability to respond—is often crucial to success. Success depends on choosing the right strategy, depending on the competitive landscape. Finally, our analysis of diminishing returns aligns with principles from Game Theory and Porter's Five Forces. Overusing strategies that are easy to copy or widely adopted can decrease effectiveness and profitability. Recognizing when a tactic's impact stalls allows organizations to optimize resource use, avoid strategic fatigue, and strengthen long-term positioning. By adopting a dynamic, data-driven approach, companies can better manage both short-term decisions and long-term strategies, ultimately building a sustainable competitive advantage.

8. Limitations and Impending Research

We believe our research provides a foundation for further analysis of soccer game strategies. One limitation is the scope and variability of the 2023 soccer game data, which could impact the generalizability of the results. Although the models showed strong explanatory power for NCAA tournament, conference, and nonconference games, the high variability in team strategies—especially in possession, transition, and restart tactics—creates challenges in drawing consistent conclusions across different contexts. Additionally, our inability to test Hypothesis 3 with strong statistical significance limits our understanding of the restart strategy. Future research should explore how restarts influence game outcomes to gain a better understanding of this tactic. We also recommend analyzing a larger, multi-season dataset to improve the accuracy and reliability of the findings. Similarly, studying professional games, women's college games, and teams that did not qualify for the NCAA tournament could reveal whether similar patterns are present. Finally, investigating teams using a case-study approach during an entire season can also show how the level of the opponent impacts the strategy selected and the outcome. Let this article be the playmaker that will allow other researchers to score by adding their talented research to move forward the science on optimizing outcomes in soccer.

9. References

1. Bai, L., Gedik, R., & Egilmez, G. (2023). What does it take to win or lose a soccer game? A machine learning approach to understand the impact of game and team statistics. Journal of the *Operational Research Society*, 74(7), 1690-1711.
2. Best, H., & Wolf, C. (2015). Logistic regression. *The SAGE handbook of regression analysis and causal inference*. Los Angeles: Sage, 153-171.
3. Boeker, M., & Midoglu, C. (2023). Soccer athlete data visualization and analysis with an interactive dashboard. *International Conference on Multimedia Modeling*. *Springer International Publishing.* pp. 565-576.
4. Bravo, A., Karba, T., McWhirter, S., & Nayden, B. (2021). Analysis of Individual Player Performances and Their Effect on Winning in College Soccer. *SMU Data Science Review*. 5(1), 8.
5. Carling, C., Williams, A. M., & Reilly, T. (2007). Handbook of soccer match analysis: A systematic approach to improving performance. *Routledge*.
6. Cho, H., Ryu, H., & Song, M. (2022). Pass2vec: Analyzing soccer players' passing style using deep learning. *International Journal of Sports Science & Coaching*. 17(2), 355-365.
7. Das, A. (2024). Logistic regression. In Encyclopedia of quality of life and well-being research. *Springer International Publishing*. pp. 3985-3986.
8. Forcher, L., Altmann, S., Forcher, L., Jekauc, D., & Kempe, M. (2022). The use of player tracking data to analyze defensive play in professional soccer-A: A scoping review. *International Journal of Sports Science and Coaching*. 17(6), 1567-1592.

9. García-Aliaga, A., Marquina, M., Coteron, J., Rodríguez-González, A., & Luengo-Sanchez, S. (2021). In-game behaviour analysis of football players using machine learning techniques based on player statistics. *International Journal of Sports Science and Coaching*. 16(1), 148-157.
10. Hughes, M., & Franks, I. (2005). Analysis of passing sequences, shots, and goals in soccer. *Journal of Sports Sciences*. 23(5), 509-514.
11. Lago-Ballesteros, J., Lago-Peñas, C., & Rey, E. (2012). The effect of playing tactics and situational variables on achieving score-box possessions in a professional soccer team. *Journal of Sports Sciences*. 30(14), 1455-1461.
12. Lago-Peñas, C., Lago-Ballesteros, J., Dellal, A., & Gómez, M. (2010). Game-related statistics that discriminated winning, drawing, and losing teams from the Spanish soccer league. *Journal of sports science and medicine*. 9(2), 288.
13. Orlick, T., &Partington, J. (1988). Mental links to excellence. *The sport psychologist*. 2(2), 105-130.
14. Rico-González, M., Pino-Ortega, J., Méndez, A., Clemente, F.,& Baca, A. (2023). Machine learning application in soccer: a systematic review. *Biology of sport*. 40(1), 249-263.
15. Schoenfeld, B. J., Grgic, J., & Krieger, J. (2019). How many times per week should a muscle be trained to maximize muscle hypertrophy? A systematic review and meta-analysis of studies examining the effects of resistance training frequency. *Journal of Sports Sciences*. 37(11), 1286-1295.
16. Shen, E., Santo, S., & Akande, O. (2022). Analyzing pace-of-play in soccer using spatio-temporal event data. *Journal of Sports Analytics*. 8(2), 127-139.
17. Verma, R. K., & Shrivastava, M. K. (2024). Analysis of research trends in sports analytics for Soccer from 2012 to 2023. Sports Analytics: *Data-Driven Sports and Decision Intelligence*. 145-162.
18. Wu, L. Y., Danielson, A. J., Hu, X. J., & Swartz, T. B. (2021). A contextual analysis of crossing the ball in soccer. *Journal of Quantitative Analysis in Sports*. 17(1), 57-66.

About Our Authors

Yohann Mauger is an Associate Professor in the Department of Management at the Haile College of Business, Northern Kentucky University. His research explores decision making at both individual and organizational levels across industries such as sports, labor markets, and smart cities. His work has been published in peer-reviewed journals, including *Baltic Journal of Management* and *Technology Analysis & Strategic Management*, and presented at international conferences like the Academy of Management and Academy of International Business. Dr. Mauger enjoys global collaboration with faculty and students and serves as a consultant for professional sports organizations worldwide.

Sima M. Fortsch is an Assistant Professor of Global Supply Chain Management at the Hail College of Business, Northern Kentucky University. She holds a Ph.D. in Supply Chain Operations and Strategic Management from SUNY Buffalo, an Executive MBA from the University of Rochester, and a B.S. in Chemical Engineering from Northeastern University. With over 17 years of industry experience at major S&P

500 companies, she has held roles as both an engineer and a manager. Her research uses econometric and statistical methods to develop industry-specific best practices. Her expertise includes supply chain management, logistics, product development, risk management, Lean, and TQM.

Elena A. Khapalova is an Assistant Professor in the Business Department of the School of Business and Liberal Arts at the State University of New York at Canton. Her research interests lie in operations and supply chain management with a focus on simulation. Her work has been published in peer-reviewed journals like The Annals of Applied Statistics, Journal of Forecasting, and Operations Research for Health Care. She has presented at multiple Decision Sciences Institute Annual Meetings.