

Grace Coccagna, Jenna Nowland, George Thorp, Ryan Thompson

21 December 2024

Data 300 Final Project Report

College Football is one of the most profitable businesses in America, bringing in 1.3 billion in revenue and over 4 billion in TV deals in just the 2023 season alone (247 Sports). College football season is beyond exciting, with upsets, massive crowds, and high-profile games spread across September to January. However, the impact of these games and the success of a team goes far beyond just the field.

In 1984, Boston College took on the Miami Hurricanes in the Orange Bowl. Boston College quarterback, Doug Flutie, threw a Hail Mary Touchdown pass as the clock struck 0 to win the game, giving Boston College an upset victory over Miami. This play made national news, and in the two years following the victory, applications to Boston College increased by about 30% (Eggers). This phenomenon became known as the ‘Flutie Effect’ - when athletic success contributes to an increase in enrollment and number of applications from students.

This project aims to examine the success of SEC and BIG 10 football programs, how this translates to changes in enrollment, and their U.S. College News Ranking. These two conferences were selected because they generate the highest revenue in college football and are the most successful in terms of on-field performance. Our dataset was constructed using the years 2005-2019. We excluded data effected by COVID-19; in addition, we excluded data that could be skewed by the impact of NIL and new conference mergers. The data was manually scraped and assembled from various public datasets. The dataset consisted of on field stats – such as, win percentage, conference wins, strength of schedule, and AP National Rank.

Additionally, the data includes academic statistics for those years, such as enrollment, applications, and U.S. College News Ranking.

Each group member had defined roles established before starting the project. However, there was an overlap, as we worked very closely with each other, and we were able to share multiple responsibilities, making it a true group effort. Ryan constructed the dataset, wrote the report, and helped with analysis and interpretation of the model. George also constructed the dataset, oversaw the GitHub, built the power point slides, and helped with analysis and interpretation of the model. Grace was the main coder for building the model, put the poster together, and helped with analysis and interpretation of the model. Jenna helped with analysis and interpretation of the model, helped build the power point, and the poster.

The experiments aimed at exploring the relationship between football performance and Fall enrollment. The correlation analysis gave insight to calculate the Pearson correlation coefficient. This coefficient helped measure the strength and the direction (positive or negative) of the relationship between football performance statistics, including wins and fall enrollment. In addition, a logistic regression model was developed with football statistics, such as wins, with the independent variable and fall enrollment as the dependent variable. The regression was compiled in order to measure how well on-field performance can predict a positive change in enrollment. The analysis of this regression helped uncover the strength of the impact on-field success has to enrollment through our p-value and our coefficient of determination. Our approach focused on simplicity, aiming to isolate football statistics to identify a positive predictor for enrollment. We performed multiple logistic regressions on our individual football statistics, enrollment, and U.S. College Rank. We explored if statistics – such as wins, win percentage, conference wins, strength of schedule, AP rank, and others – could predict an increase in enrollment. Additionally, we completed a binary regression on whether the school

that specific season qualified to attend a bowl game; a school was classified as 1 if they did and 0 if they did not (to classify for a bowl game, a school must achieve 6 wins or more in the season). It was also broken down by the conference; this assisted in determining if either the Big10 or the SEC had a higher increase in enrollment on average, depending on the success of their teams in each respective conference.

Discussion of Results:

The results from our experiments showed several mixed results. Our correlation analysis revealed a correlation of .1977 between total team wins and Fall enrollment at that school. However, we obtained a p-value of .0005, noting that we obtained a statistically significant result. This suggests that on-field success has some effect on enrollment, but it is not the strongest predictor. Our regression model uncovered similar findings. We obtained a relatively low R^2 value, denoting that total wins only explains a small bit of the variability in Fall enrollment numbers. The correlation was small, showing that each win shows just a small increase in Fall enrollment. It is important to note that it is very possible that although football success does contribute to enrollment increases, there are other potential factors that also lead to an increase in enrollment. For example, academic programs, new facility upgrades, or a well-known academic reputation.

From our K-Means Clustering algorithm, we used it to identify patterns in Fall Enrollment and Wins. Based on the Elbow Method, the optimal number of clusters is 3, and each point in the dataset was assigned a cluster label (0, 1, or 2). From Cluster 0, it represented schools with lower enrollment and a mix of lower-to-mid win totals. Schools in this cluster may not see significant increases in wins, despite modest enrollment. Cluster 1 consists of schools

with moderate enrollment and lower win totals. These schools might struggle to achieve football success consistently and may lack the enrollment size needed to boost competitiveness.

Additionally, Cluster 2 demonstrated that schools in this cluster show higher win totals and are generally associated with higher enrollment. This group likely represents successful football programs with larger enrollments, where the relationship between success and enrollment increase is more evident. In the context to our project, these findings suggest that schools with higher enrollment tend to perform better in football (as seen in Cluster 2), which aligns with the theory that greater resources, funding, and recruitment opportunities at larger institutions support competitive football programs. On the other hand, schools with lower enrollment (Clusters 0 and 1) face challenges in achieving similar levels of success. These results reinforce the hypothesis that football success is more prominent at larger universities and could influence enrollment trends positively. We would like to note that due to the way our data is constructed, a panel model would have been best suited. However, we did not have the time or resources to successfully accomplish this.

There were various lessons regarding specific Machine Learning topics that our group was able to learn due to the measures taken to complete this project. For us, the group emphasized normalizing the data. When looking at enrollment, a lot of the subjects, or schools observed, have various average enrollments, so it was important to ensure the data was completely normalized. In addition, one of the biggest things that we learned throughout the project is the importance of having a complete understanding of your data and making note of all external factors that could impact your findings. Although we feel very confident in our results and findings, it is important to acknowledge that there are other factors that help contribute to increased enrollment that is not solely the success of a football team.

Some challenges that we faced were due to time constraints; we were not able to tell if there were other factors that contributed to the increased enrollment in these schools in addition to football success. There are many different agendas that large universities have, so it is impossible to be able to truly cover all the factors that can lead to increased enrollment. One main goal that we were unable to accomplish is to see if the quality of students increased at these schools as well. For example, if a school had a successful season, would they then see an incoming class that has a higher average high school GPA, or higher average test scores than the previous class.

With recent changes in the NCAA realm, student athletes are now allowed to profit from their name, image, and likeness (NIL). For example, Colorado University's quarterback, Shedeur Sanders, has ranked up around 6.2 million dollars and over 2 million Instagram followers. This shows how the future of college athletics is shifting. This can raise questions about if these successful programs will experience even greater enrollment benefits due to the increased visibility of star athletes, and whether these financial incentives might create competitive imbalances. In the future, examining these trends can help to provide insight into the evolving role of athletics in higher education and its potential impacts on enrollment and institutional growth.

Another area for future study would be the impact of large upsets on applications. In this study we focused solely on how the success of large college football teams created increases in student enrollment. It would be interesting to investigate how enrollment would change for a smaller school if they were able to beat an SEC or a BIG10 school. In addition, if you look at other sports, specifically basketball, which participates in a tournament called March Madness. Every year, a smaller, lesser-known team makes an incredible run in the tournament and gets

national coverage. It would be intriguing to see how that type of run impacted their academic standing.

In conclusion, our project highlights the detailed relationship between college football success and enrollment trends, particularly within the SEC and BIG10 conferences. While our findings indicate a statistically significant correlation between on-field success and enrollment, the overall impact is modest, suggesting that football performance is just one of many factors influencing student decisions. Through the Flutie effect, schools may lay a foundation to examine how shifts in athletic success may further impact enrollment and institutional growth.