Advanced Business Applications

Module 3: Sports Analytics

Group 4 – Andrew Dingman & Chuol Gatkek

The professional sports landscape is ever so evolving and adapting. New ways of achieving a competitive advantage are constantly emerging. Being able to not only adapt but to understand the root cause for what drives team success is pivotal, for not only coaches but the front office personnel as well. The National Football League is on the cutting edge of the evolution by embracing and utilizing data science and analysis in their decision making.

Wins and losses are ultimately decided by a team’s ability to score and prevent the other team from scoring. Through the use of linear regression, we built a model that can predict a team’s number of wins based off its offensive and defensive production. Not only can this assist coaches in play calling decisions, but also front office personnel on possible player rearrangements or recruitment.

As previously mentioned, our linear regression looks at a team’s total points scored, and points allowed to predict its total number of wins. We compiled data from the regular season from 2018 to 2023. Our model results show that the coefficient for total points scored was .029, meaning, with each point scored, a team’s total wins should increase approximately .029. Essentially, roughly every 35 points a team scores should result in 1 win. On the contrary, the coefficient for total points allowed is -.025. Indicating that for each point given up to an opposing team the number of wins decreases by -.025; Every 40 points given up costs the team approximately 1 win.

A screenshot of a computer

AI-generated content may be incorrect.

At face value, points scored, and points allowed up are neck and neck for value for a team. Total points scored have a slight edge over points allowed. Resulting in scoring having a larger impact on wins than defense. Which makes logical sense, to win a team has to score, it doesn’t matter how well or dominant the defense is if they can not get any points on the board.

The other analytical take aways from the linear regression include an adjusted R-square of .79, affirming that the model is a good fit for the data. Also, the standard error of the model was 1.38, so on average the model predicts total wins within 1.37 of the actual total wins for a team. Also, the P-Value for the predictors, total points scored, and total points allowed, were extremely low; meaning both are key to predicting team success.

A graph of a graph

AI-generated content may be incorrect.

The above graph shows our models win predictions for NFL franchises for the 2023 season compared to their actual number of wins, based off their total points scored and points allowed. This shows a team like Baltimore winning less games than our model predicted and Washington winning more than predicted. An explanation for this could be teams getting a bad break, maybe a call didn’t go their way in crunch time, or vice versa and a team got lucky. Another instance that could affect the model is a team wins one week by a large margin such as 70-14, and then the next week plays in a snowstorm and loses 0-3. Analytically, they should have 2 wins and no losses, however over the course of the season those outliers tend to work themselves out.

With the use of the coefficients, teams would be able to self-reflect on performance not only at the end of the season, but also midseason. For instance, If a team was 5-0 and had abysmal offensive production, and a great defense, The model can calculate what their wins could be at if their defense wasn’t so dominate, or if they had an average defense, and so on and so forth with similar instances. By utilizing the mode in those type of scenarios, front office and coaches can see if the current record is sustainable for the whole season or if its being held up by one side of the ball.

Below is a table showing the statistics for Baltimore, Houston, and Washington.

A close-up of a number

AI-generated content may be incorrect.

As we can see, the only prediction that the model was incorrect by over 1 was Houston, and with such a close total point scored and points allowed margin, it is safe to assume they had some games that could have gone either way.

The model can also be used to predict hypothetical wins for a team. A practical use would be to build additional models based off previous seasons that could forecast a team’s offensive and defensive statistics for the upcoming season. Those numbers could then our model to estimate wins. This not only shows the value and flexibility of our model, but the benefits of continued development and growth potential of continuing the utilization of analytics in sports.

We created a set of hypothetical teams and gave them randomly assigned values for total points scored and points allowed to demonstrate how possible preseason predictions would work. This scenario highlights the models potential for forecasting, which would in turn benefit the organization as a whole.

A black text on a white background

AI-generated content may be incorrect.

In summary, the model built has a magnitude of potential benefits for an organization. From preseason predictions to midseason assessment, and post season reflection. By using a team's offensive production and defensive performance, the model is able to explain nearly 80% of the variation in total wins over a season. This makes it a powerful tool for strategic planning, performance monitoring, and data-driven decision-making. However, the model has its own limitations, and outside factors can impact its accuracy. Some of those factors include weather, officiating, bad breaks, and even injuries. Therefore, the model is best used along with other tools to enhance decision-making.