

Analysis of NFL Teams

By: Jake Sorrentino

Using the provided NFL team statistics data since 2000, (Kaggle, 2024) we analyzed key factors that correlate with wins, including points scored points allowed, points differential, and other performance metrics like margin of victory, strength of schedule, and offensive/defensive rankings and more important variables. Using the data acquired I look to answer the following question: What is the most statistically significant factor influencing NFL team wins and losses? Using key findings such as summary statistics, correlation and predictive analysis I can help to increase a teams ability to win now and for the future.

**Figure 1:**

*NFL Standings and Data Excel Sheet*

A screenshot of a computer

AI-generated content may be incorrect.

Note. Created from Microsoft Corporation. (2018). *Microsoft Excel*. Retrieved from https://office.microsoft.com/excel

**Figure 2:**

A screenshot of a computer

AI-generated content may be incorrect. *SQL: Input Table NFL Standings and Data*

pgAdmin Development Team. (2025). *pgAdmin: Open-source administration and management tool for PostgreSQL* (Version 7.5) [Software]. <https://www.pgadmin.org>

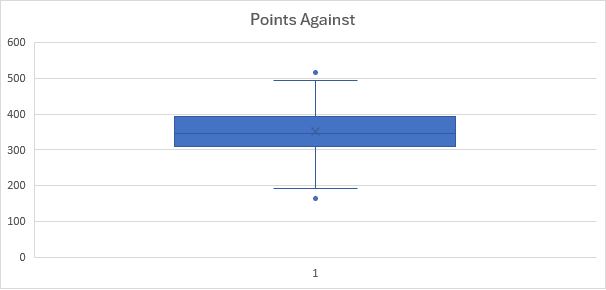
**Figure 3:**

A screenshot of a computer

AI-generated content may be incorrect. *SQL: Code for Summary Statistics/ Points For Points Against*

pgAdmin Development Team. (2025). *pgAdmin: Open-source administration and management tool for PostgreSQL* (Version 7.5) [Software]. <https://www.pgadmin.org>

**Figure 4:**

 *Points Against Box Plot*

Note. Created from Microsoft Corporation. (2018). *Microsoft Excel*. Retrieved from https://office.microsoft.com/excel

**Figure 5:**

A screenshot of a computer

AI-generated content may be incorrect. *SQL: Points Against Box Plot Code*

pgAdmin Development Team. (2025). *pgAdmin: Open-source administration and management tool for PostgreSQL* (Version 7.5) [Software]. <https://www.pgadmin.org>

**Figure 6:**

A screenshot of a computer

AI-generated content may be incorrect. *SQL Correlation Wins*

pgAdmin Development Team. (2025). *pgAdmin: Open-source administration and management tool for PostgreSQL* (Version 7.5) [Software]. <https://www.pgadmin.org>

**Figure 7:**

*A screenshot of a computer

AI-generated content may be incorrect. SQL Correlation Losses*

pgAdmin Development Team. (2025). *pgAdmin: Open-source administration and management tool for PostgreSQL* (Version 7.5) [Software]. <https://www.pgadmin.org>

**Figure 8:**

*A screenshot of a computer

AI-generated content may be incorrect. SQL: Line of Best Fit*

pgAdmin Development Team. (2025). *pgAdmin: Open-source administration and management tool for PostgreSQL* (Version 7.5) [Software]. <https://www.pgadmin.org>

**Figure 9:**

*A screenshot of a computer

AI-generated content may be incorrect. SQL Slope Check*

Note. Created from SAS Enterprise Miner Program. 2022. SAS Institute Inc 2013. SAS/ACCESS® 9.4 Interface to ADABAS: Reference. Cary, NC: SAS Institute Inc.

**Figure 10:**

*SAS Enterprise Miner MLB Attendance Transform Node Part 1*

pgAdmin Development Team. (2025). *pgAdmin: Open-source administration and management tool for PostgreSQL* (Version 7.5) [Software]. <https://www.pgadmin.org>

**Figure 10** **:**

*Line of Best Fit Pd to Loss*

Note. Created from Microsoft Corporation. (2018). *Microsoft Excel*. Retrieved from https://office.microsoft.com/excel

**Figure 11:**

*A screenshot of a computer

AI-generated content may be incorrect.SQL Code: Loss Prediction*

pgAdmin Development Team. (2025). *pgAdmin: Open-source administration and management tool for PostgreSQL* (Version 7.5) [Software]. <https://www.pgadmin.org>

**Figure 12:**

*Line of Best Fit: Pd to Wins*

Note. Created from Microsoft Corporation. (2018). *Microsoft Excel*. Retrieved from https://office.microsoft.com/excel

**Figure 13:**

*A screenshot of a computer

AI-generated content may be incorrect.SQL Code: Wins Prediction*

pgAdmin Development Team. (2025). *pgAdmin: Open-source administration and management tool for PostgreSQL* (Version 7.5) [Software]. <https://www.pgadmin.org>

This SQL code performs an extensive statistical analysis of football statistics stored in the public."standings" table, focusing on metrics like points scored, points allowed, and game outcomes (wins and losses). The first query retrieves all records from the table, providing the raw dataset for further analysis.

Subsequent queries calculate descriptive statistics for points\_for (points scored by teams) and points\_against (points scored against teams), including their count, minimum, maximum, average, and standard deviation. These summary statistics describe the distribution of offensive and defensive performance across all teams.

A separate query calculates percentile statistics for points\_against, determining the minimum value, first quartile (Q1), median, third quartile (Q3), and maximum value. These percentiles provide a more detailed understanding of the defensive capabilities of teams and highlight how these metrics are distributed. This also helps to establish relationships and determine any outliers

The next set of queries evaluates the correlation between various performance metrics and game outcomes. For wins, correlations are calculated with metrics like points scored, points allowed, points differential, margin of victory, strength of schedule, simple rating, offensive ranking, and defensive ranking. These correlations indicate the strength and direction of relationships, helping to identify which factors are most strongly associated with winning games. Similarly, correlations are calculated between losses and these metrics to analyze their association with poor performance. We can see in both wins and loss the highest correlation variable seems to be point\_differential with 0.916 and -0.916. The negative correlation indicates an inverse relationship.

A regression analysis is then performed using the REGR\_SLOPE and REGR\_INTERCEPT functions, which determine the slope and intercept of the linear relationship between wins and points differential. This relationship is expressed in the form y=mx+by = mx + by=mx+b, where m is the slope and b is the intercept. This means with an output of .028 for the slope and 7.98 for the y incercept. This makes the line of best fit y= (.028)x + 7.98 This model can predict the number of wins and losses a team might achieve based on its points differential.

Before running the regression, I needed to double check the slope. By knowing corr=r and r+ystdev/xstdev= slope. I queried the code above and got .028 once again. The code also predicts losses based on points differential using a custom regression formula. It calculates the predicted loss value for a given input (e.g., points differential of 200) by combining the mean, standard deviation, and correlation of losses and points differential.

By running the linear regression framework that is used to predict wins based on points differentia, it calculated a predicted number of wins for a specified input value by leveraging averages, standard deviations, and the correlation coefficient between wins and points differential. The line of best fit to predict the y value would be, y=.028(200) + 7.98. Using this input, the predicted y value for wins when the point differential is 200 is approximately 14 while losses is about 2. From the above data I have determined the point\_differential is one the most signifact variables in the NFL for determining wins and losses.

o