DATA 6560 - Project Proposal October 31, 2025

Sydney Lockwood

The Impact of Defensive Pressure and Sacks on Win Probability

Project Overview

This proposal focuses on analyzing sacks and defensive pressure to determine whether generating consistent quarterback pressure on defense can impact win probability. The goal is to determine how defensive disruption and pressure influence team performance and overall success.

Problem Statement

In the NFL, defensive performance is often judged by points allowed or turnovers, but pressures and sacks may better reflect a defense's ability to control a game. Understanding this relationship is important for coaches and analysts seeking to optimize defensive strategies, allocate resources effectively, and identify undervalued defensive players who generate high-impact plays beyond traditional scoring statistics.

Intended Dataset

For this analysis, I will use datasets available on GitHub and data from Next Gen Stats. The NFLFastR GitHub dataset provides play-by-play statistics from every NFL game since 1999, including sacks, pressures, wins, losses, etc. This data source will help me calculate certain metrics, such as sack rate and pressure rate, to help answer my proposed question. Next Gen Stats from the NFL can also help in my research and data analysis, as they provide more granular data that I can marry with my other source.

Initial Plan/Method

1. Data Cleaning and Preparation: I am going to merge play-by-play data with team results
2. Analysis: Sum average pressure by defense and sack rates across teams, compared to their win percentage
3. Correlation and regressions: Use the data to complete correlation and linear regression models to test the relationship between sack rate and team win probability/EPA
4. Visualization: Create scatterplots and trendlines to illustrate how the increased pressure relates to certain outcomes

Next Steps

After submitting the proposal, I will:

1. Access and download the play-by-play data from GitHub and Next Gen Stats
2. Begin loading the data sets into an Excel sheet to prepare for data cleaning and merging
3. Ensure the data provided will help me answer my question (refine the proposed question or find more extensive metrics if needed)
4. Clean the data for easier readability and analysis.

Data Sources:

https://github.com/nflverse/nflfastR/blob/master/R/data_documentation.R
https://nextgenstats.nfl.com/stats/top-plays/fastest-sacks/2025/REG/all