CS - NFL Fan Sentiment Analysis: Predicting Game Outcomes Through Social Media

DS 2000 - Spring 2025
Due: TBD
Submission format: GitHub repository (submitted by link to Canvas)
Individual Assignment

General Description: Submit to Canvas a link to your GitHub repository containing all code, data, analysis, and findings from your investigation into the relationship between NFL fan sentiment on Twitter and actual game outcomes.

Preparatory Assignments: None

Why am I doing this? This case study allows you to apply data science techniques to a real-world question in sports analytics. You'll experience the entire data science workflow from data preprocessing to insight generation, while working with natural language data—a key skill in today's data science landscape. This project provides hands-on experience with sentiment analysis, statistical modeling, and drawing actionable conclusions from social media data.

What am I going to do? Using the provided dataset of NFL-related tweets, you will conduct a sentiment analysis to determine if fan attitudes before games correlate with actual outcomes. You'll clean and prepare the data, analyze sentiment patterns, apply statistical methods to test for relationships, visualize your findings, and communicate your conclusions effectively.

Tips for success:

- Start by exploring the dataset thoroughly to understand its structure
- Carefully consider how to define "pre-game" tweets (how many days before a game?)
- Pay attention to potential confounding variables in your analysis
- Create clear visualizations that effectively communicate your findings
- Document your code thoroughly with comments
- Consider alternative hypotheses when interpreting your results

How will I know I have Succeeded? You will meet expectations on this case study when you follow the criteria in the rubric below.

Spec Category	Spec Details
Repository Structure	• Your GitHub repository should contain: ○ README.md file with project overview ○ Data folder with the dataset ○ Notebook or script files showing your analysis ○ Visualization outputs ○ Executive summary document

Data Preprocessing	• Goal: Clean and prepare the Twitter data for analysis • Parse timestamps correctly • Filter for relevant pre-game tweets • Handle any missing values appropriately • Match tweets to specific games and outcomes
Sentiment Analysis	• Goal: Analyze and quantify sentiment in tweets by team • Calculate average sentiment scores for each team before games • Explore sentiment distribution across teams • Compare sentiment patterns for winning vs. losing teams
Statistical Analysis	• Goal: Test the relationship between pre-game sentiment and game outcomes • Apply appropriate statistical model (logistic regression recommended) • Test statistical significance of relationship • Consider possible confounding factors • Interpret p-values and confidence intervals correctly
Visualization	• Goal: Create informative visualizations that clearly communicate findings • Include at least 3 different types of visualizations • Ensure visualizations have proper titles, labels, and legends • Use appropriate color schemes to distinguish different categories
Findings & Conclusions	• Goal: Clearly communicate results and their implications • Provide a concise answer to the research question • Discuss statistical significance of findings • Address limitations of the analysis • Suggest potential improvements or future work • Limit to one page maximum