

# Brian\_Reppeto540\_Milestone1

January 7, 2024

## 0.1 Project Subject Area:

This project aims to analyze and derive insights from NFL data to include betting stats, incorporating information from three distinct sources: a flat file containing team data, an API providing game statistics, and a website with game schedule, betting stats, weather, and surface information.

### 0.1.1 Data Sources:

Flat File: Description: A CSV file containing historical team data, including details such as team names, abbreviations, conference, and division information. Link: [File:https://github.com/nflverse/nfldata/blob/master/data/teams.csv](https://github.com/nflverse/nfldata/blob/master/data/teams.csv)

API (Sportradar): Description: The Sportradar API provides comprehensive NFL game statistics including team name, game date, player-related data, and stadium information. Link: [http://api.sportradar.us/nfl/official/trial/v7/en/games/7d06369a-382a-448a-b295-6da9eab53245/statistics.json?api\\_key=k8uqtrmhauf3rmkm6f8t9f54](http://api.sportradar.us/nfl/official/trial/v7/en/games/7d06369a-382a-448a-b295-6da9eab53245/statistics.json?api_key=k8uqtrmhauf3rmkm6f8t9f54)

Website (Rotowire): Description: The Rotowire website offers valuable data on NFL game schedules, kickoff times, betting stats, weather conditions, scores, and surface information. Link: <https://www.rotowire.com/betting/nfl/archive.php>

### 0.1.2 Relationships:

The flat file (teams.csv) Contains information about NFL teams such as team names and abbreviations.

The API data contains game dates and team names along with detailed stats of each game.

The website (Rotowire) provides game schedule information with team abbreviations and game dates.

All 3 of these data sources are related by NFL team names. The CSV file has a 1 to many relationship with both the Website NFL team name and has a one to many relationship with the API data by NFL team name as well. Both the Website data and the API data will have a many to many relationship with one another as the NFL team name and the game date will be needed to link each file to on another.

### 0.1.3 Project Approach:

To accomplish this NFL data analysis project, the initial step involves cleaning and preprocessing the data from the three data sources. This includes addressing missing values, standardizing formats, and ensuring consistency in team identifiers. A data dictionary will be established to

document the meaning and structure of each variable. Each data source will then be loaded into three separate tables and linked together by the relationships identified above to create a joined table that can be queried for valuable insights. The establishment of relationships between datasets is critical. Historical team data will be linked with real-time data through team names, and game dates to providing a comprehensive overview of team performance over time.

In the interpretation phase, insights derived from the betting lines will be integrated with the overall analysis. This involves assessing whether specific team or player trends correlate with betting odds and outcomes. Exploring anomalies and unexpected patterns in the data may reveal valuable insights into the relationship between on-field performance and betting predictions.

Once all the data is merged into the database, five visualizations will be created to help explain the data in a visual form. The data that will be displayed is not yet known as the data has not yet been evaluated for insights.

Ethical considerations, including privacy and adherence to league regulations, will remain a priority throughout the project. Protecting sensitive player data, particularly in injury-related scenarios, will be carefully managed.

Challenges may arise during web scraping due to potential changes in the Rotowire website structure. Regular updates and monitoring will be necessary to adapt to any modifications. Handling real-time API data will require robust error handling and continuous validation.

By the project's conclusion, the aim is to have an understanding between NFL performance trends and betting lines. This additional analysis will contribute to a comprehensive understanding of the factors influencing both on-field and betting outcomes in the NFL.