

# CSE 6242 Final Poster - Team 176 - Modeling Spreads for NFL Betting Odds

**Joshua McLennan Mayanja**

**Aditya Thakur**

**Kevin Buck**

**Ryan Buck**

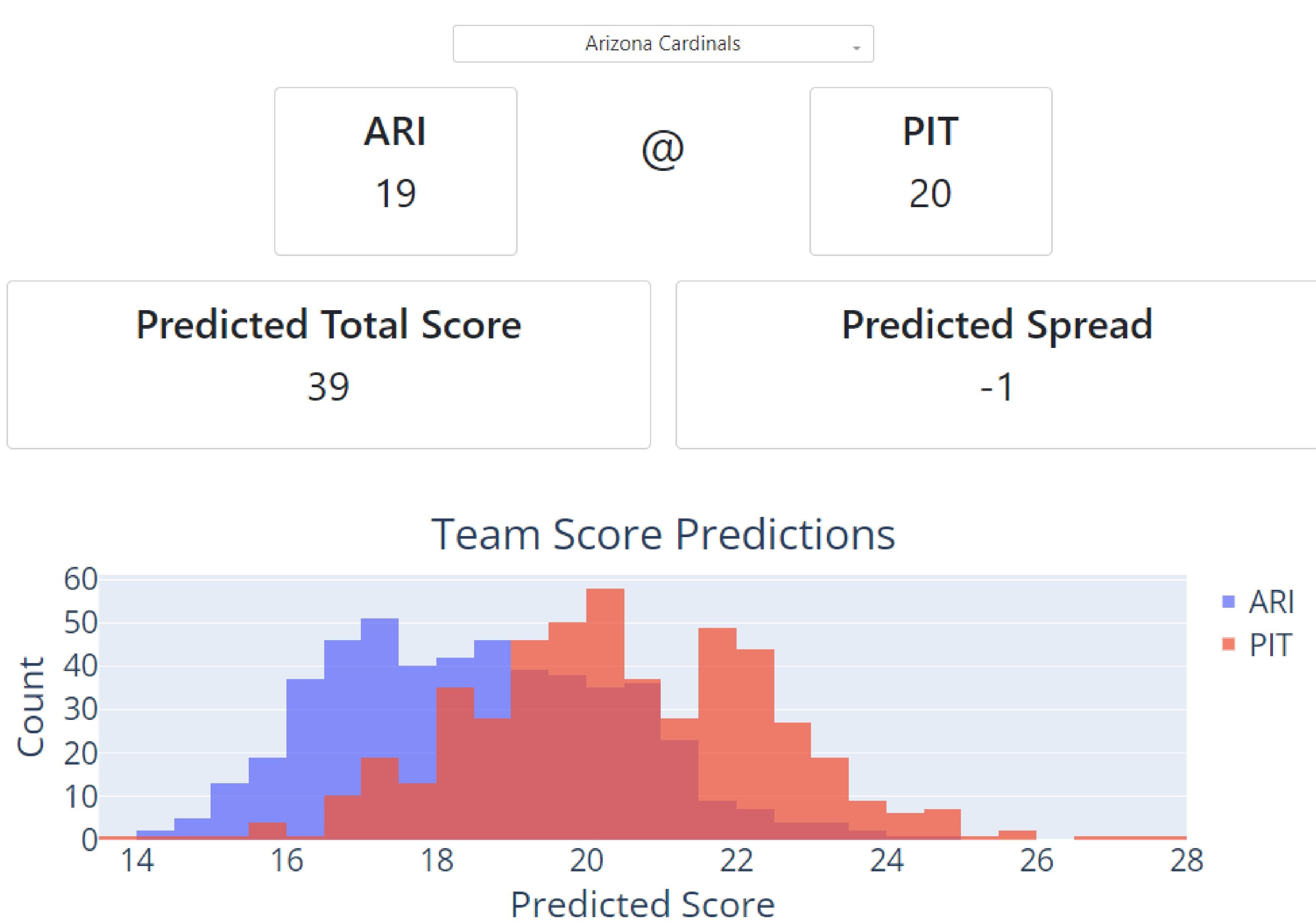
**Sameer Rau**

## Problem

With the legalization of sports betting in the U.S., there's a lack of tools to verify the accuracy of betting lines set by sportsbooks. This gap poses risks of potential line manipulation, disadvantaging millions of bettors.

Addressing this issue is crucial to ensure fairness and transparency in a burgeoning industry, protecting the interests and trust of bettors across the country.

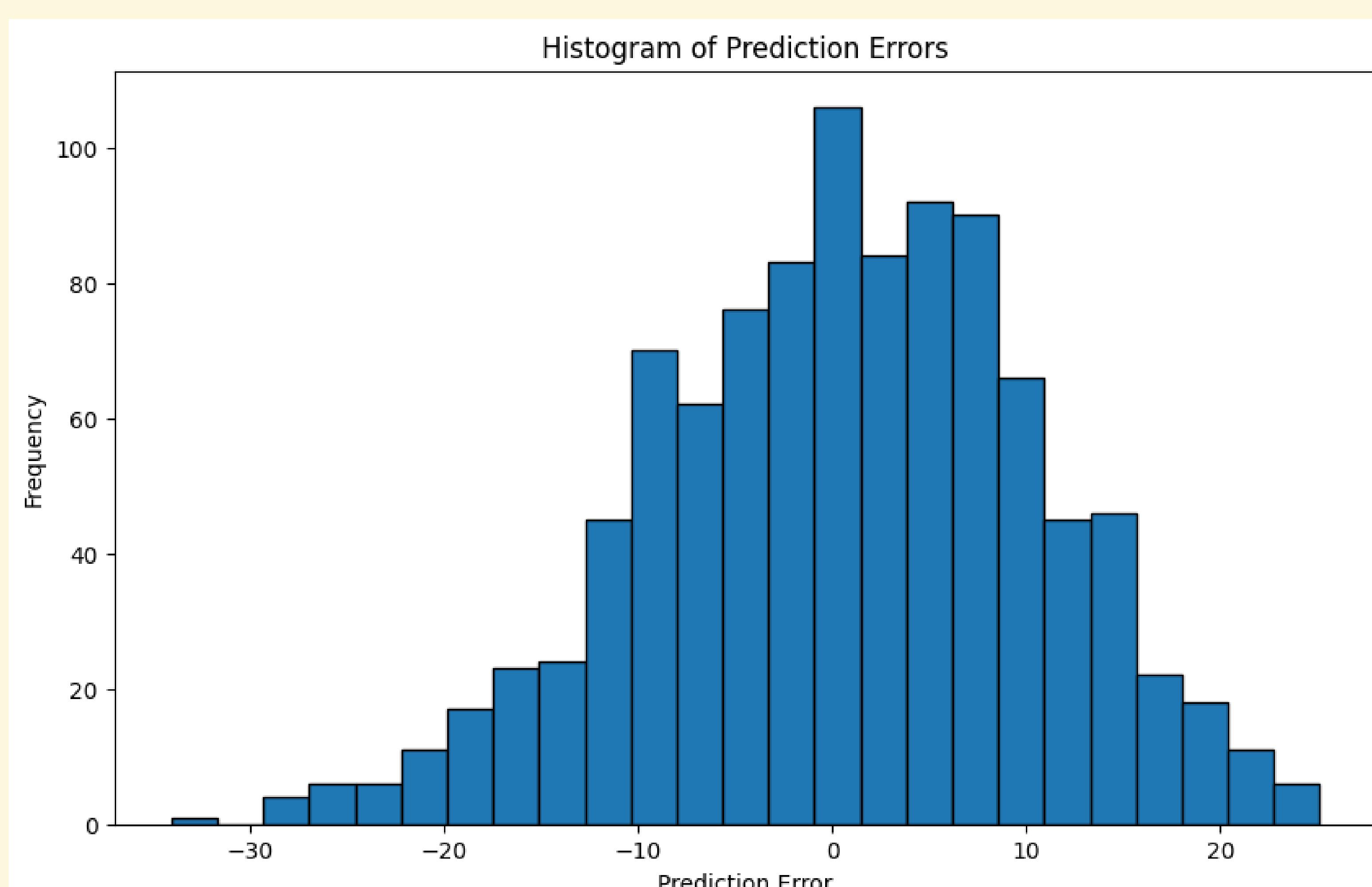
## Beat The Bookie



## Data

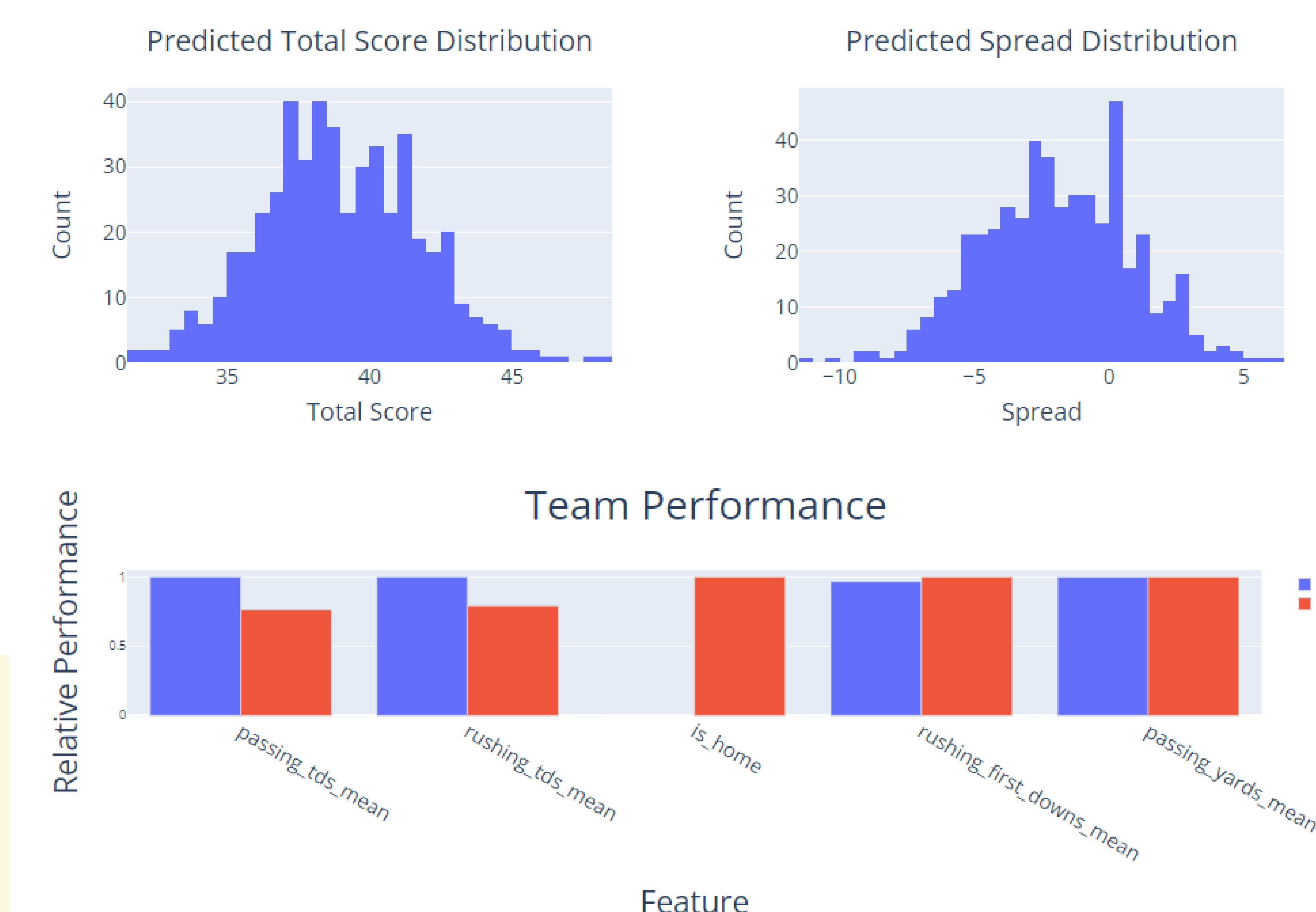
The data for this project was sourced from a CSV file on Kaggle, which includes NFL game and betting results since 1966. Additional data was extracted through 'nfl\_data\_py'.

The refined dataset focuses on games from 2002 onwards, featuring all 32 NFL teams. It encompasses about 5MB on disk, with approximately 5600 data points and 94 features per data point. The data includes spread and points totals, game scores, stadium details, and extensive team performance metrics.



## Approach

Our project features an NFL point spread model using a random forest model to evaluate sportsbooks' betting lines' accuracy, presented through a plotly dashboard for clear visualizations. It analyzes extensive historical data from 32 NFL teams, incorporating factors like game location and team performance for precise predictions. This innovative approach, enhanced by detailed player data and matchup analyses, offers a comprehensive and accurate tool for assessing betting line integrity, setting a new standard in sports betting analysis.



## Experiment and Results

Our random forest model's performance was assessed using Mean Absolute Error (MAE), Mean Squared Error (MSE), and R-squared statistics, yielding an MAE of 7.979, MSE of 99.076, and an R-squared of 0.074. The MAE suggests that, on average, our model's predictions deviate from the actual scores by approximately 8 points, a margin that is impactful given the typical point spreads in NFL games. The low R-squared value indicates that the model captures a small portion of the variance in game outcomes, reflective of the unpredictable nature of NFL game results. Validation was conducted using a robust cross-validation approach, ensuring the model's generalizability across various data subsets and enhancing confidence in its predictive power.

Our user-friendly dashboard provides bettors with an easy-to-understand visual representation of predictions, distinguishing our approach from more complex models. It offers a balance of simplicity and comprehensive data analysis, tailored for the average user.