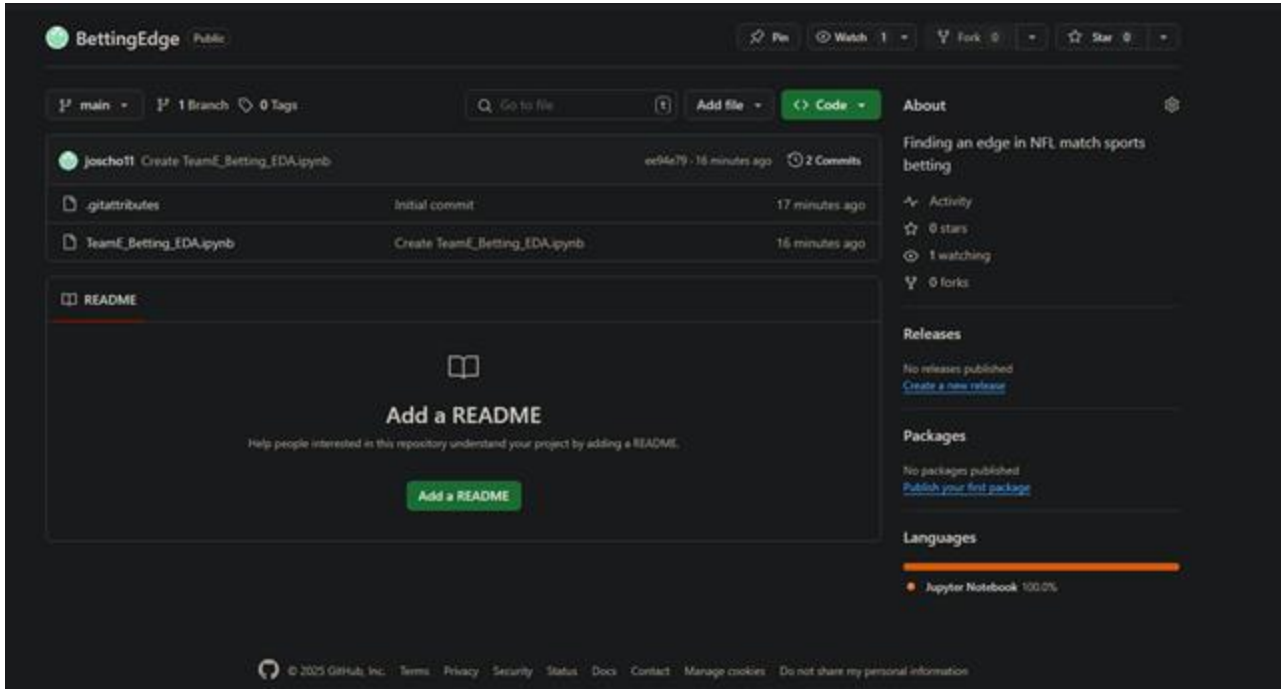
FINDING AN EDGE IN SPORTS BETTING

By Team E

Joseph Schoenbaum and Gabriel Kuykendall

GITHUB FOR OUR PROJECT



- https://github.com/joscho11/BettingEdge

OVERVIEW

Introduction to sports betting and our project

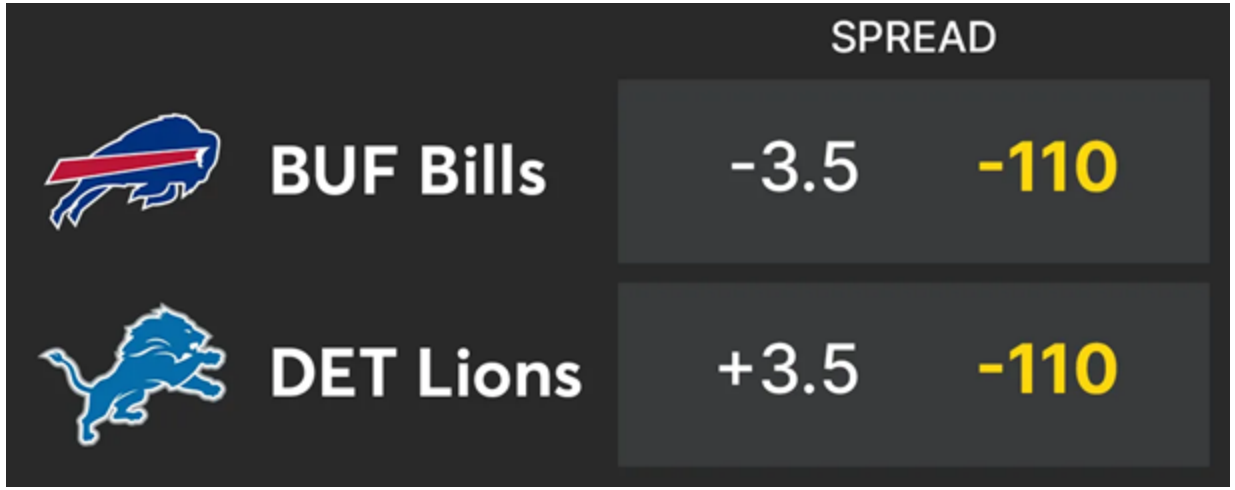
What can we takeaway from other sports betting projects?

Introduction to the dataset / Project Outline

Introduction to Sports Betting

Glossary:

- **Point Spread -** The expected difference in score between two teams
- **Plus/Minus -** Plus is associated with the underdog and Minus to the favorite
- **X.5 -** This allows sports books to avoid a tie between the bettor and the book because it is impossible to score half a point
- **Vigorish(vig) -** Represented as -110 in the graphic, this shows the payout for the betters. -110 means you can win $100 for every $110 wager, it is standard because the spread is expected to cover the odds, +110 would need a $100 wager to win $110



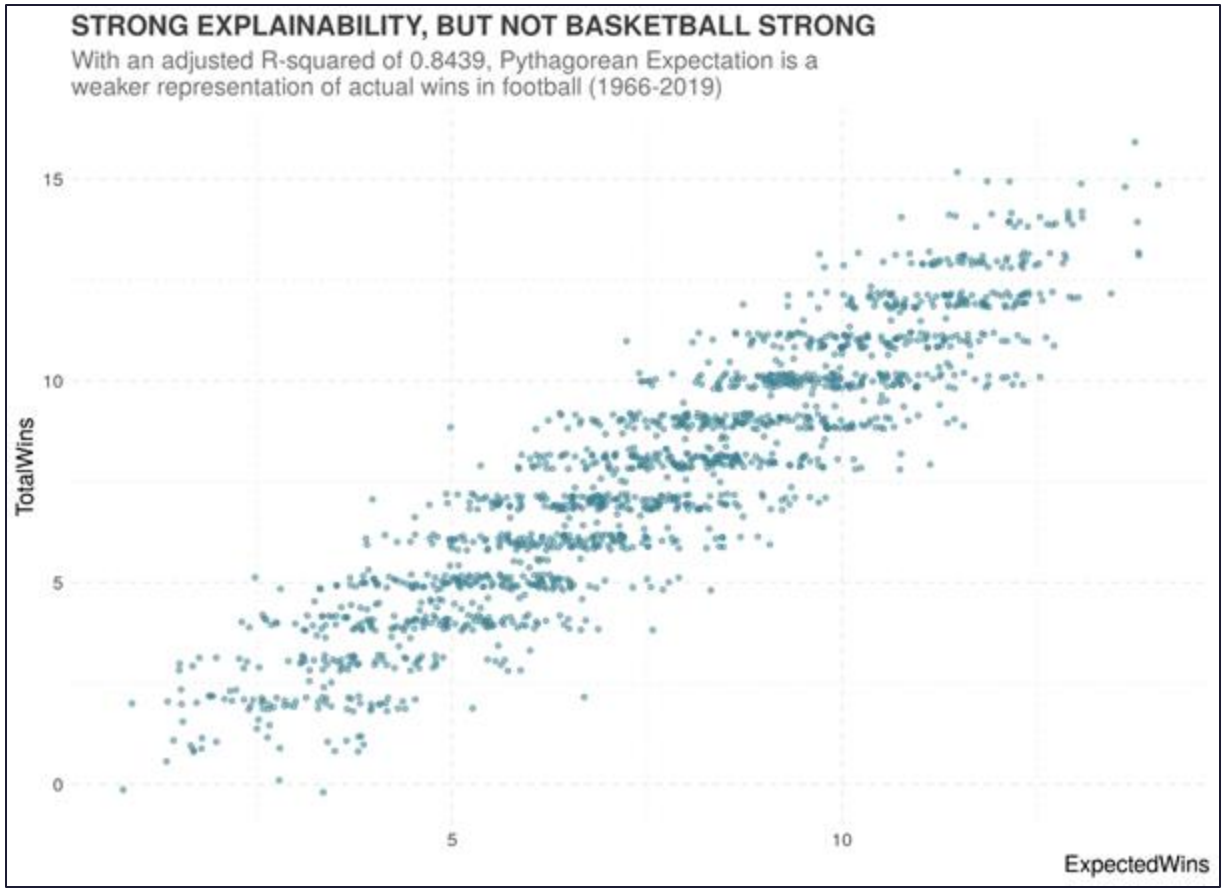
Project Introduction

- Using different match related features, create several models to predict the spread in NFL games
- Compare our new spreads to existing spreads
- Determine the effectiveness our spreads in improving the earnings of bettors

Other Project Takeaways

Applying Pythagorean Expectations - from Kaggle

- Pythagorean Expectations is a statistic created by Major League Baseball to predicted expected wins, usually used to see see if a team is underperforming or overperforming based on results
- This project attempted to use that same statistic and apply it to other major sports like basketball and football
- A key takeaway from this is that every sport had their own modifier applied to fit the statistic, the models we create will not be directly replicable for other sports



Other Project Takeaways

NFL Betting Model - from Kaggle

- Using Logistic Regression classifier model to determine if the favorite will “cover” or get over the spread
- Model Results show to be more favorable than random guessing
- Model uses less features than we are planning to use and only focuses on favored team instead of the spread

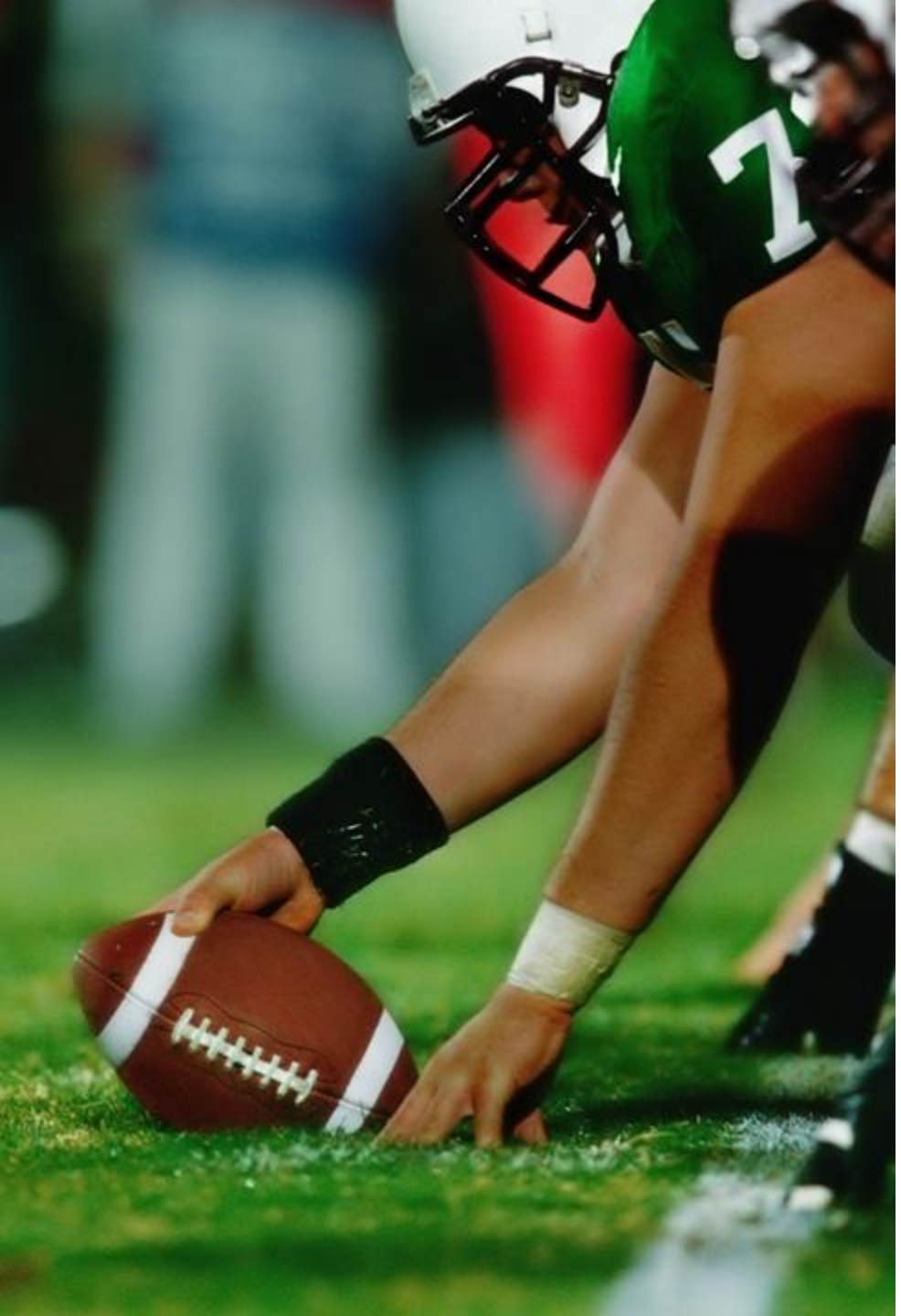
In [26]:

```
# initial features possible for model
X = df[['schedule_season', 'schedule_week', 'over_under_line', 'spread_favorite', 'weather_temper
ature', 'weather_wind_mph',
        'home_favorite', 'hm_avg_pts_diff','aw_avg_pts_diff', 'elo1', 'elo2', 'elo_prob1']]

y = df['result']
```

INTRODUCTION TO THE DATASET

- For our project we will be using data from a python library called nfl_data_py
- nfl_data_py is a Python library for interacting with NFL data sourced from nflfastR, nfldata, dynastyprocess, and Draft Scout
- Includes import functions for play-by-play data, weekly data, seasonal data, rosters, win totals, scoring lines, officials, draft picks, draft pick values, schedules, team descriptive info, combine results and id mappings across various sites
- Our project will use game data from 2000 to 2025
  - Ex: each row will include information about an event – 02/09/25 - Kansas City Chiefs vs Philadelphia Eagles (22 – 40)
- Original dataset
  - 6732 rows and 46 columns
- Link to library - https://pypi.org/project/nfl-data-py/
- Link to github of library - https://github.com/nflverse/nfl_data_py/tree/main

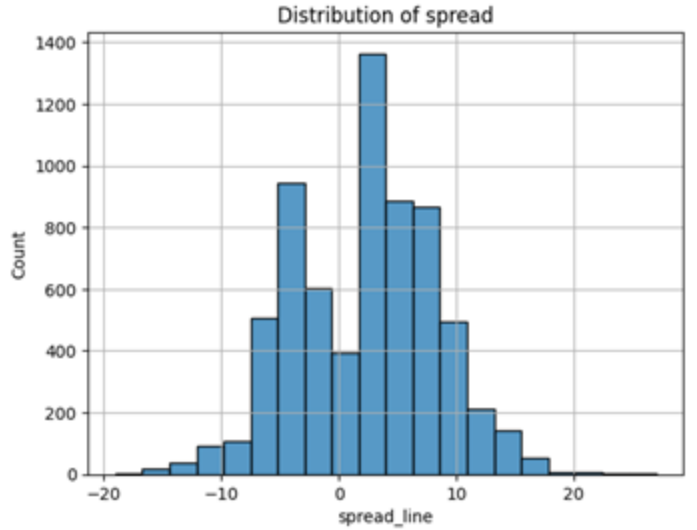


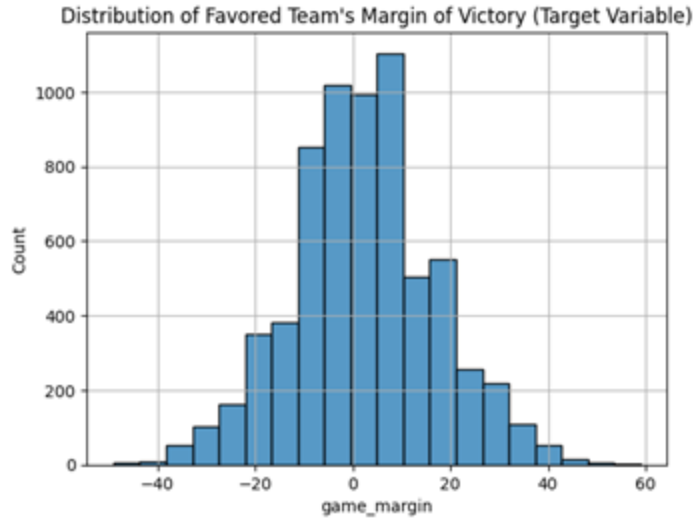
MANUALLY ADDED FEATURES

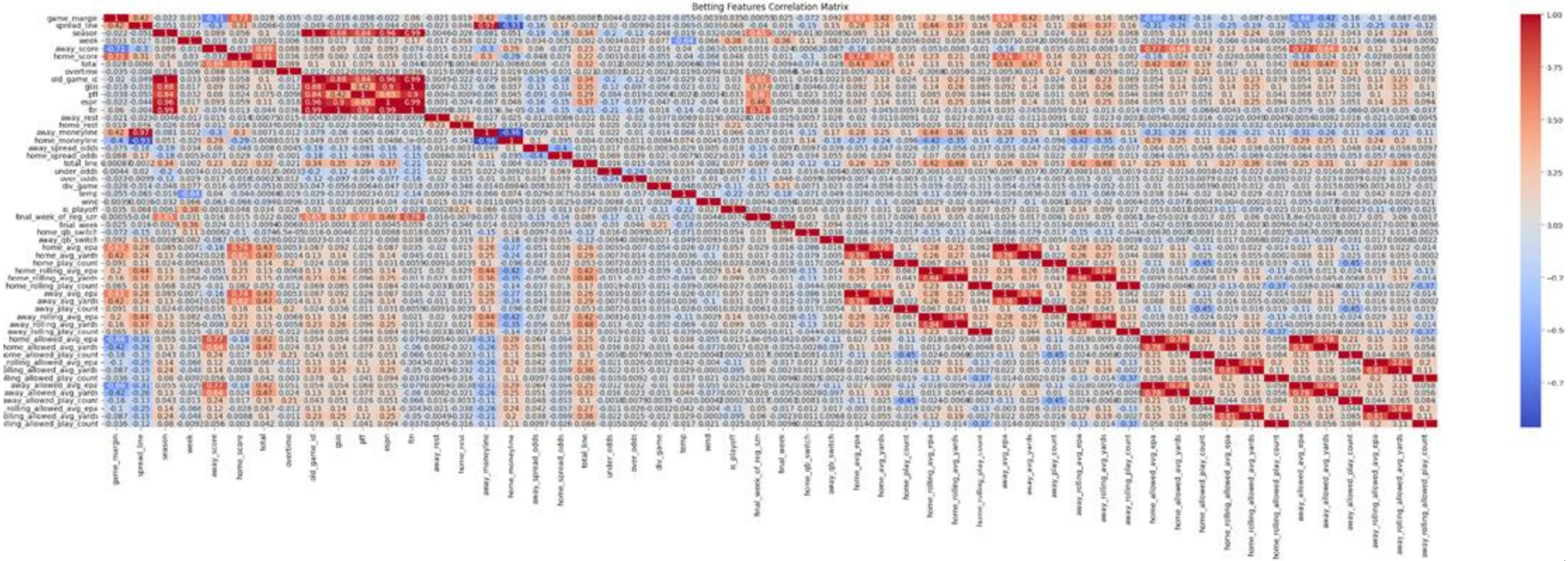
- is_playoffs – teams tend to play harder in playoff games so scores may be lower and closer
- is_final_week – teams tend to rest players in the last week of the regular season
- home_qb_switch, away_qb_switch – the quarterback is the most important position in team sports, so if they changed the qb from one week to another it will most likely have a major impact on the result (differentiate between home and away team)
- home_rolling_avg_epa, away_rolling_avg_epa – rolling average of the home and away teams' expected points added which is a metric that measures the impact of each play on a team's chances of scoring
- home_rolling_allowed_avg_epa, away_rolling_allowed_avg_epa – rolling average of the home and away teams' allowed expected points added (how well the defense can prevent the offense from having productive plays)
- home_rolling_avg_yards, away_rolling_avg_yards – rolling average of the home and away teams' average yards per play
    - If a team has a lot of yards, they tend to score more points
- home_rolling_allowed_avg_yards, away_rolling_allowed_avg_yards – rolling average of the home and away teams' allowed average yards per play
    - Shows how well a team is at stopping their opponent from moving the ball down the field
- home_rolling_play_count, away_rolling_play_count – rolling average count of the plays the home and away teams run
    - The more plays you run, the more time you have the ball, the more likely it is you will win the game
- home_rolling_allowed_play_count, away_rolling_allowed_play_count – the rolling average count of plays the defense can limit the offense to
    - The less amount of the plays the better

EDA

- Dataframe info
- Feature counts
- Heatmap
- Distribution of target variable (game margin)
- Distribution of spread line







HEAT MAP OF FEATURES FROM ORIGINAL DATASET

CLEANED DATASET

- Key features from the original dataset included: roof, surface, temp, wind, home_qb_name, away_qb_name, referee, stadium_id, away_rest, home_rest, total_line
- After cleaning (only take data after week 3), our dataset has 3831 rows and 34 features

NEXT STEPS – PROJECT OUTLINE

finalizing the dataset / features for modeling

build various predictive models that output a result for a given match

Compare model output with truth value

We will then evaluate those models in comparison to real spreads given to us in the dataset

Compare how well our models do in comparison to the other models we have implemented for this project to conclude what model may be best suited for this task.

Some of the models we are thinking about using for this project include linear regression, random forest regressor, gradient boosted trees, or a multi-layer perceptron

To evaluate our models, something such as MAE may be best because betting results can be very noisy. To account for this, we don't want to penalize outliers too heavily.

References

- https://www.foxsports.com/stories/nfl/point-spread-over-under
- https://letmebet.com/blog/what-is-a-point-spread/
- https://www.kaggle.com/code/jaseziv83/applying-pythagorean-expectation-to-major-sports/input
- https://www.kaggle.com/code/twalters20/nfl-betting-model#Feature-and-Model-Testing

THANK YOU

- Questions?