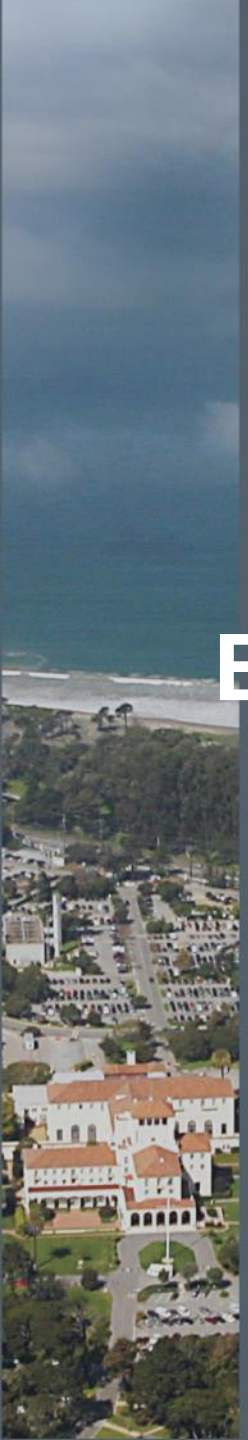


NAVAL
POSTGRADUATE
SCHOOL

Efficacy of NFL Game Predictors

OA3801 Final Project Presentation

ENS Ford Knippa
ENS Walker Lee
ENS Patrick Oliver
ENS Brennan Penafiel

OA3801 - Summer 2024

Overview

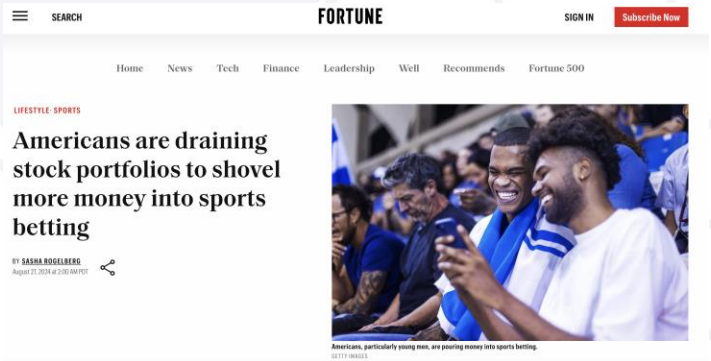
- Background
- Finding the Data
- Cleaning the Data
- Methodology
- Analysis
- Conclusions

The Question

How good is Vegas in predicting the outcomes of NFL Games?

- Predicting the outcome of NFL games before they occur is a major area of interest for financially motivated sports data analysts.
- Sportsbooks (i.e. casinos) employ oddsmakers to create these predictors considering many different factors.
- Terms to know:
  - The favorite is the team expected to win the game and is "giving" points.
  - The underdog is the team expected to lose and is "getting" points.
  - Oddsmakers set a point spread to represent the expected margin of victory by which the favorite is supposed to win the game.
  - When a spread is covered, it means that the team favored by the spread has won by more points than the spread itself, or the underdog has lost by fewer points than the spread or won outright.
  - The over and under refer to wagers on whether the combined total score of a game will be higher or lower than a number set by the sportsbook. Betting the "over" means you predict the score will exceed the set number, while betting the "under" means you expect it to fall short.

DRAFT KINGS



Finding the Data

- Using a data set from Kaggle, we were initially presented with three data frames:
  - Team characteristics (44 total teams)
  - Stadium characteristics (120 stadiums)
  - Game results (14,073 games from 1966 to 2023)
- What data should we use?

Cleaning the Data

- Remove games prior to 2002
  - League became 32 teams with updated divisions in 2002
- Account for team name changes and relocation
  - Redskins ⇨ Washington Football Team ⇨ Commanders
  - Oakland Raiders ⇨ Las Vegas Raiders
  - San Diego Chargers ⇨ Los Angeles Chargers
  - St. Louis Rams ⇨ Los Angeles Rams
- New stadium names
  - For example, Mercedes-Benz Superdome ⇨ Caesars Superdome
  - Only current NFL stadiums are included in our analysis
  - Weather data is included in stadium information data frame, need to account for name changes to include weather data in our analysis

Cleaning the Data

- Merge data frames (team characteristics, game results, stadium information)
- Calculate the result of each game
  - Given:
    - Spread
    - Over/Under line
    - Final score
  - Calculate:
    - Was the spread covered?
    - Did the over or under hit?
    - Did the favored team win?
- Game results used for analysis

Gambling Warning

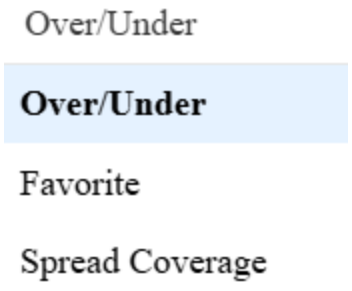
shutterstock_69389101.jpg (800×450) (albawaba.com)

Reminder: ***We do not condone the use of this information for Vegas sports-betting or other forms of sports-related gambling!***

First Dash Demonstration, 1

- Key Features:
  - Plotly Dash to visualize data
  - Two dropdown menus: Desired Prediction and Team
  - Toggleable between Regular Season vs. Playoffs
  - Red-Green Comparison corresponding to relevant outcome
    - Military Colors > Colorblindness

First Dash Demonstration, 2



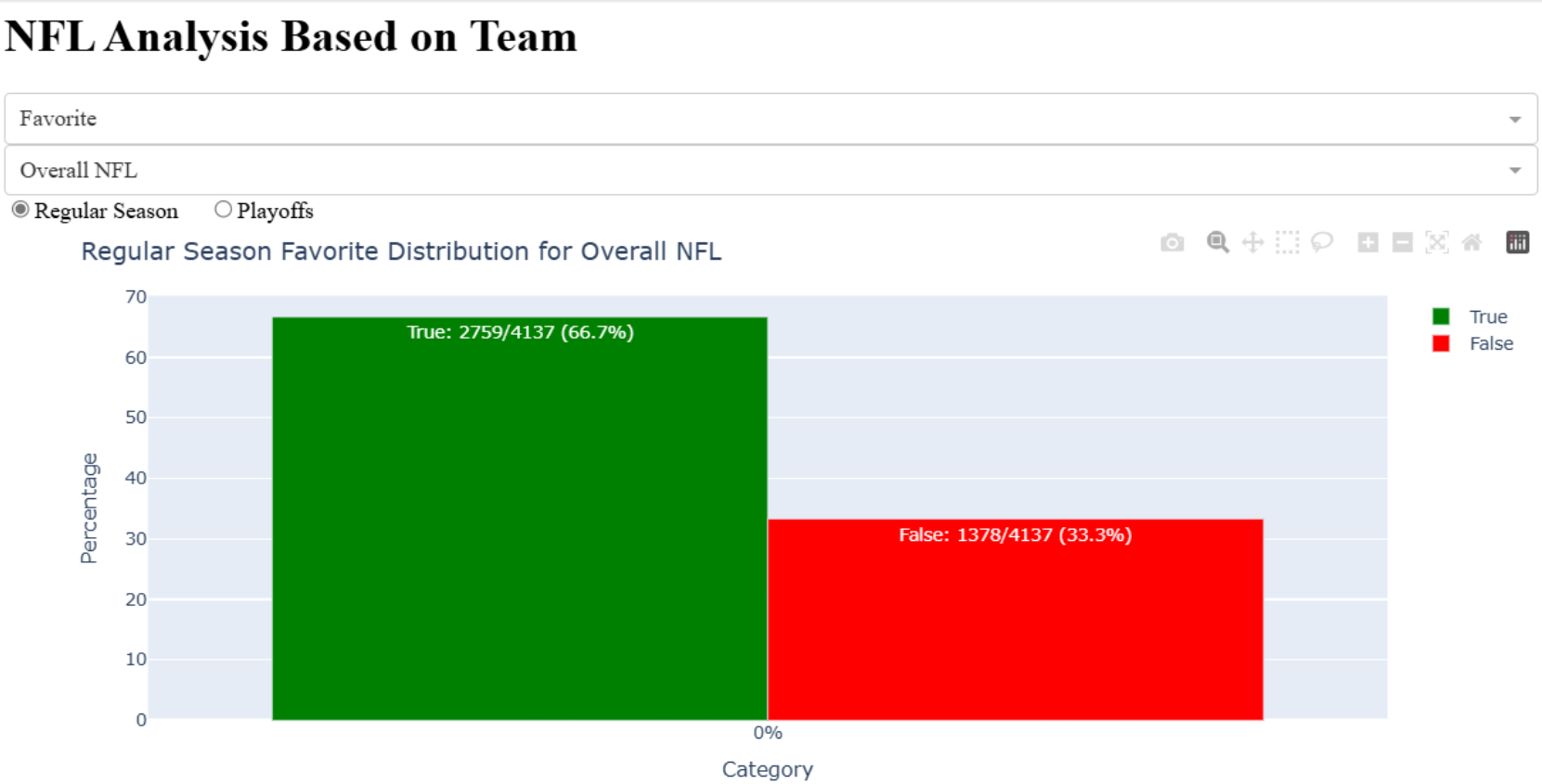
<---1st Dropdown:
-Lists the three most relevant areas for NFL game predictions

2nd Dropdown: --->
-Lists all 32 individual teams comprising the NFL, sorted alphabetically, as well as the NFL as a whole

First Dash Demonstration, 3

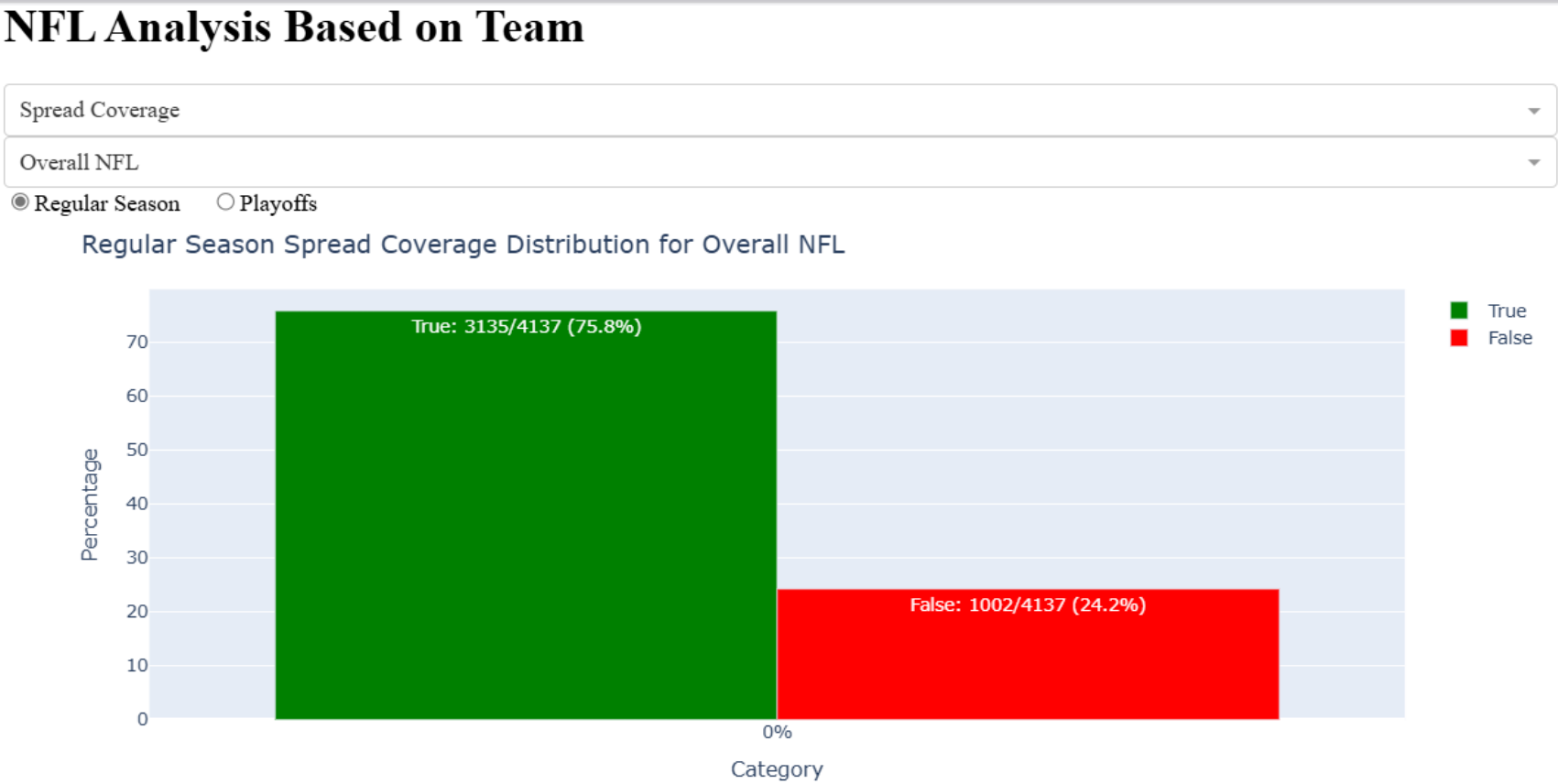
- Over/Under
  - Should be even and non-skewed
  - Over/Under is the threshold line that leads to an even outcome divergence
  - Vegas has accurately set a diverging threshold near the total score of 4,137 NFL games at a ratio of 1/2 for the past 21 years
    - Extremely successful!

First Dash Demonstration, 4



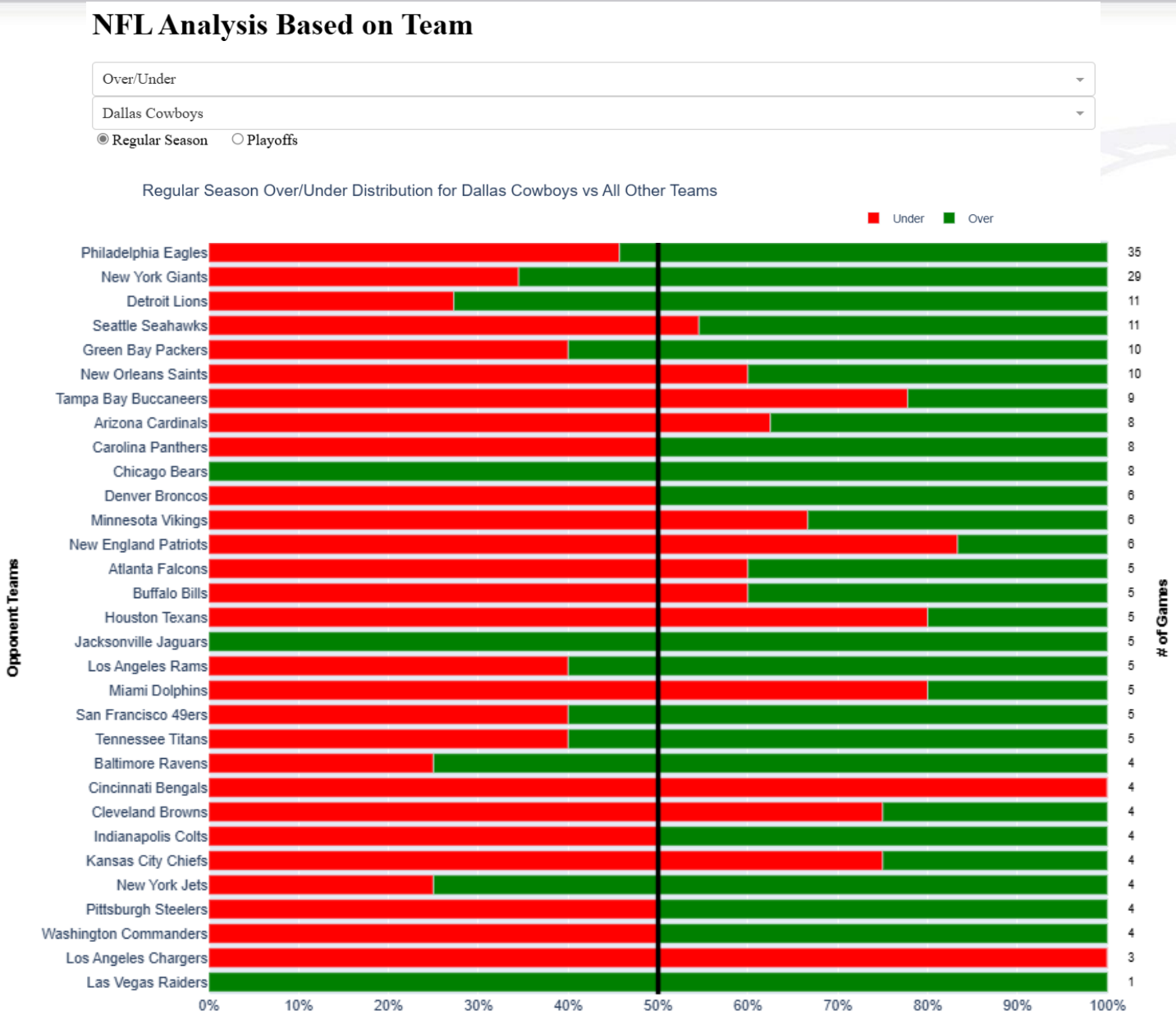
- Favorite
  - Should be skewed
  - Favorite is the team that is determined to have greater odds of winning
  - Vegas has accurately predicted the winning team in 4,137 NFL games at a ratio of 2/3 for the past 21 years
    - Significantly successful!

First Dash Demonstration, 5



- Spread Coverage:
  - Should be skewed
  - Spread is the predicted point differential that the favored team will win by
  - Vegas has accurately predicted the spread being covered in 4,137 NFL games at a ratio of 3/4 for the past 21 years
    - Extremely successful!

First Dash Demonstration, 6



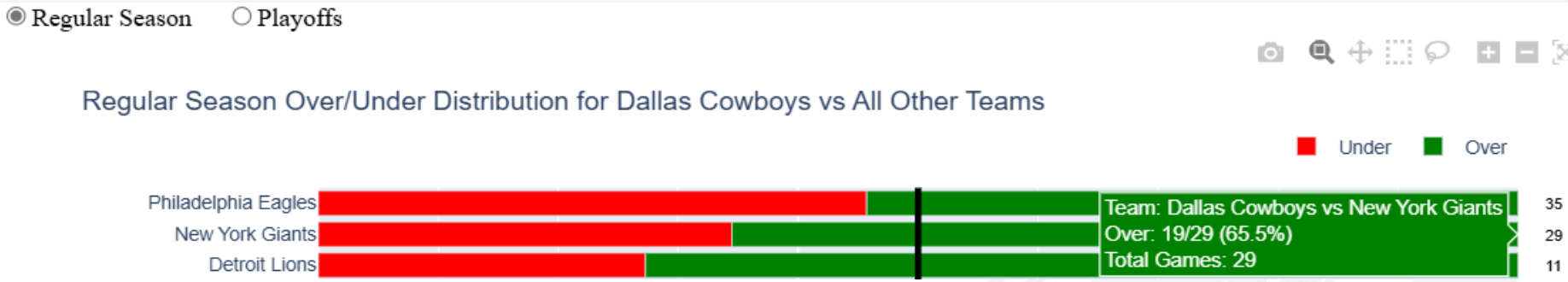
First Dash Demonstration, 7

- Interactive interface allows for exploration of large amount of data
- Opponent teams are ordered by number of matchups from most to least

First Dash Demonstration, 8

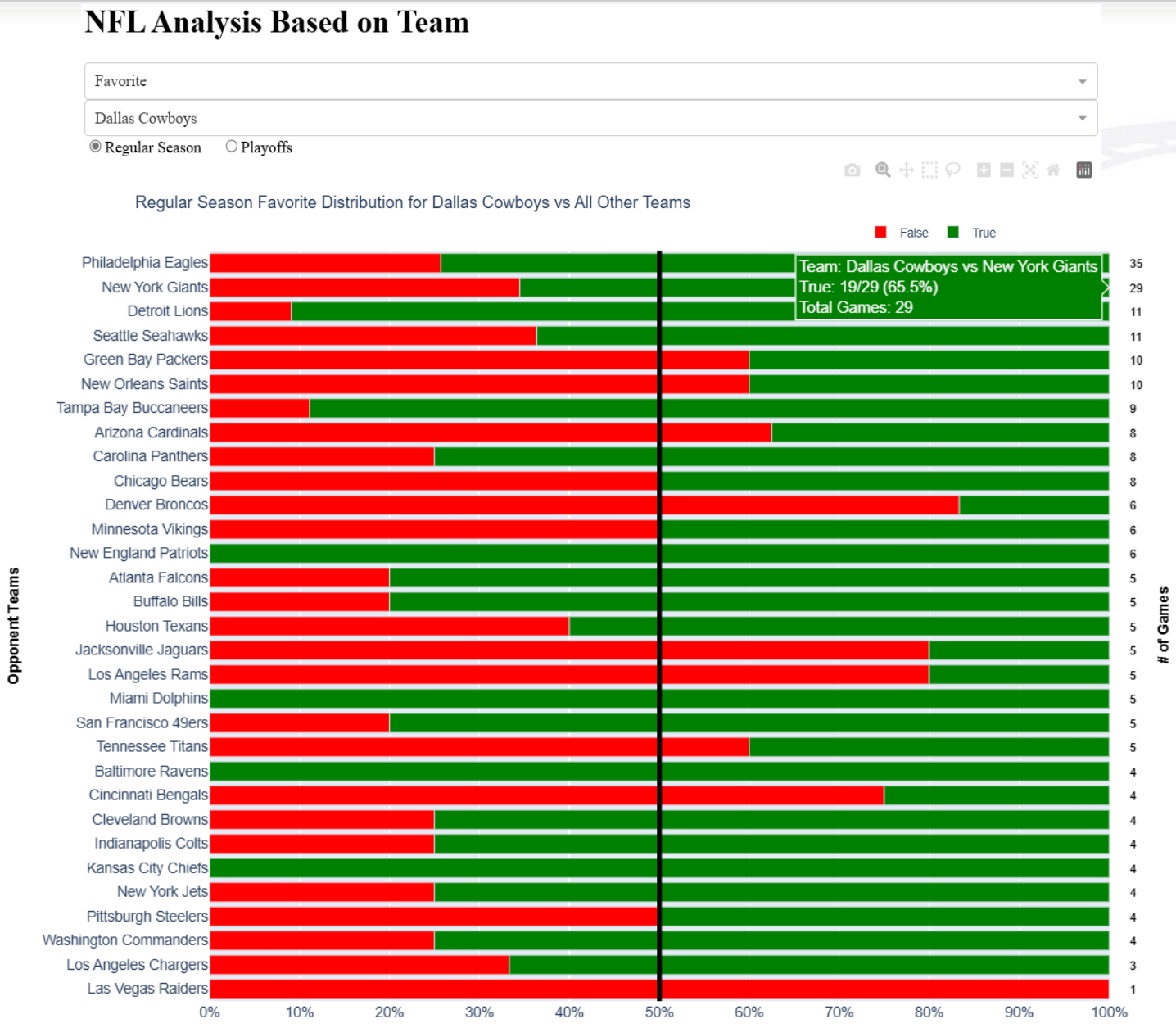
Sat, Aug 24 · Final	Pre Season · Week 3	Sun, Sep 8	Week 1
Chargers 26		Cowboys	1:25 PM
Cowboys 19		Browns	
Sun, Sep 15	Week 2	Sun, Sep 22	Week 3
Saints	10:00 AM	Ravens	1:25 PM
Cowboys		Cowboys	
Thu, Sep 26	Thursday Night Football	Sun, Oct 6	Sunday Night Football
Cowboys	5:15 PM	Cowboys	5:20 PM
Giants		Steelers	
Sun, Oct 13	Week 6	Sun, Oct 27	Sunday Night Football
Lions	1:25 PM	Cowboys	5:20 PM
Cowboys		49ers	
Sun, Nov 3	Week 9	Sun, Nov 10	Week 10
Cowboys	10:00 AM	Eagles	1:25 PM
Falcons		Cowboys	

First Dash Demonstration, 9

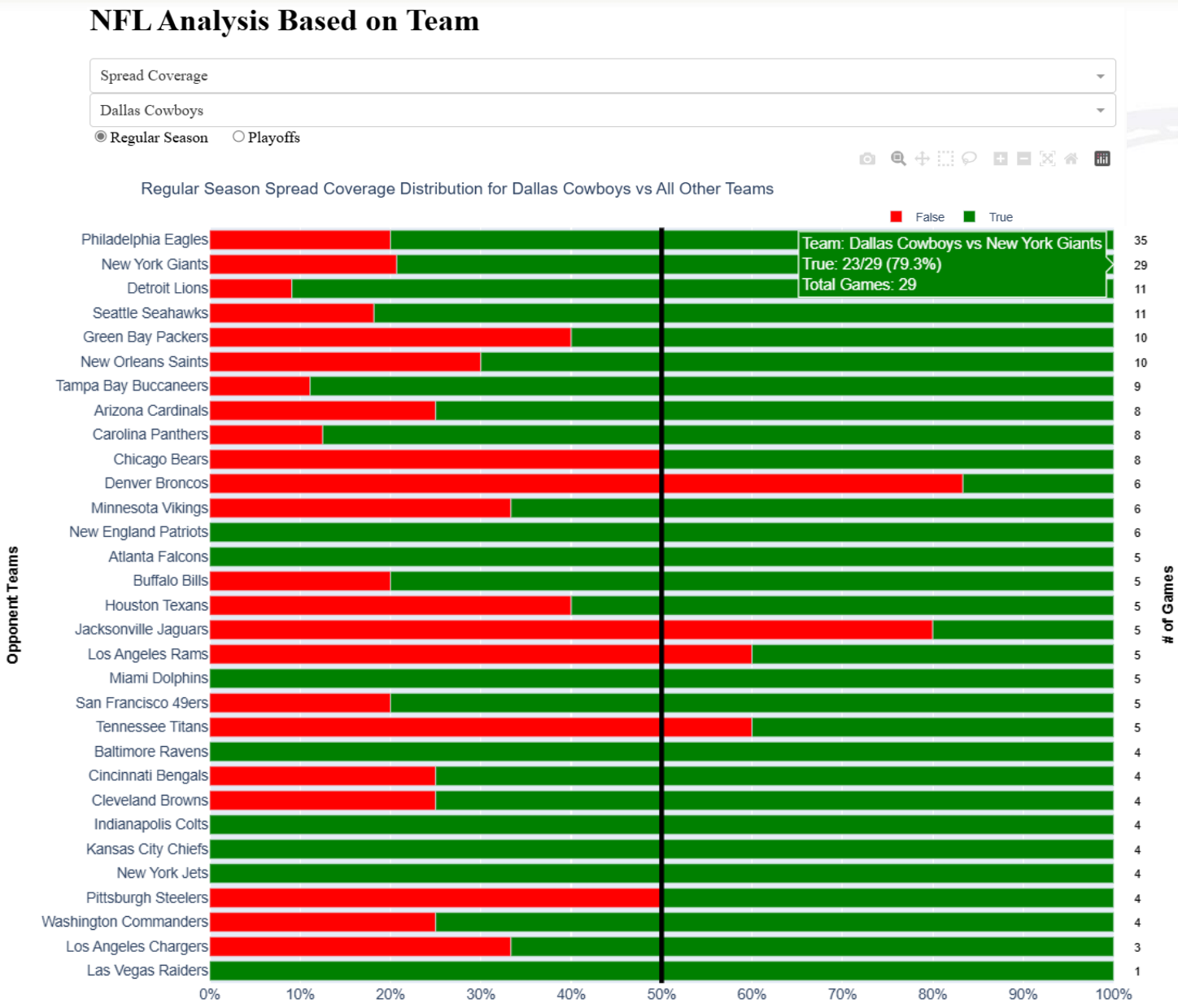


- Why is this dashboard useful?
  - When the Cowboys play the Giants, the total score tends to be over the Over/Under line in 65.5% of the 29 games played from 2002-2023.
    - The projected total score tends to be underestimated more often in this matchup compared to the overall NFL prediction rate of 49.1%
    - Impress your friends for no financial gain with this information!

First Dash Demonstration, 10



First Dash Demonstration, 11



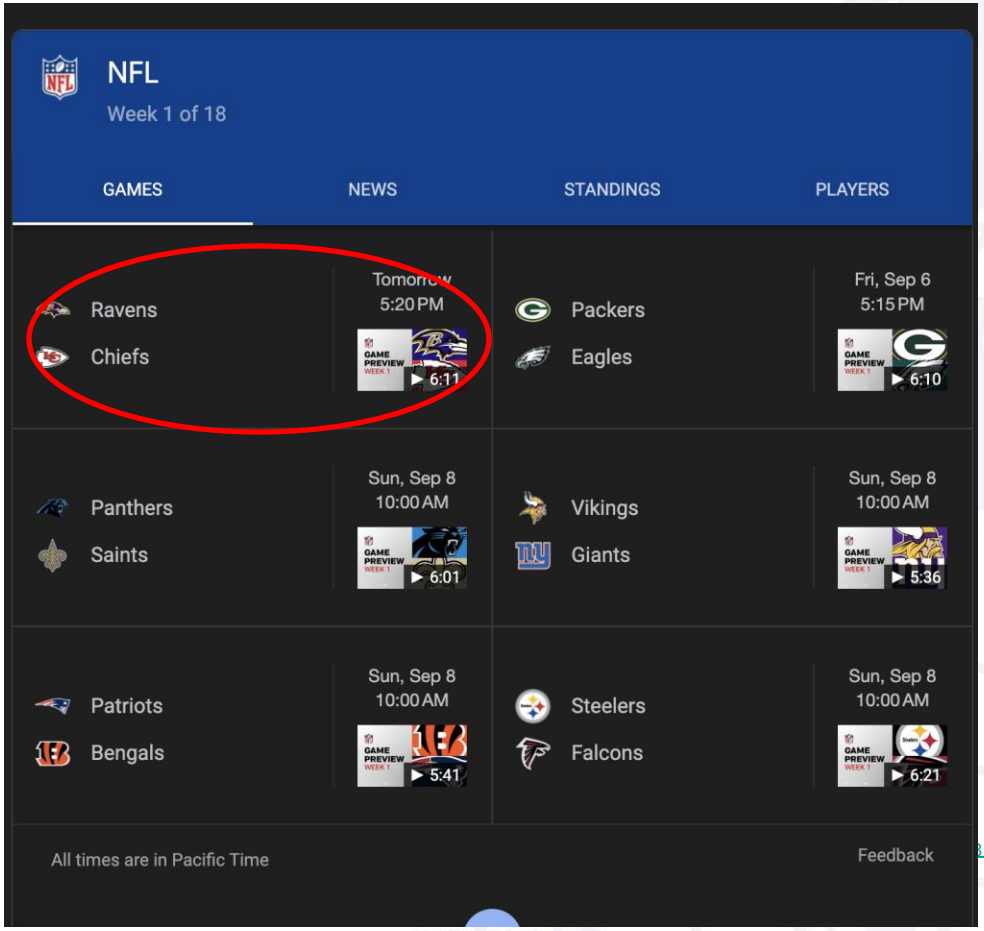
First Dash Demonstration, 12

- Main Conclusion from First Dashboard:
  - Vegas is extremely successful at accurately predicting NFL game outcomes when analyzing the Over/Under, Favorite, and Spread Coverage for the past 21 years
  - By knowing this, you can equip yourself with accurate knowledge (to impress your friends, not to make money)

Circa-Sportsbook-768x576.jpg (768×576) (vegasalways.com)

Second Dash Demonstration, 1

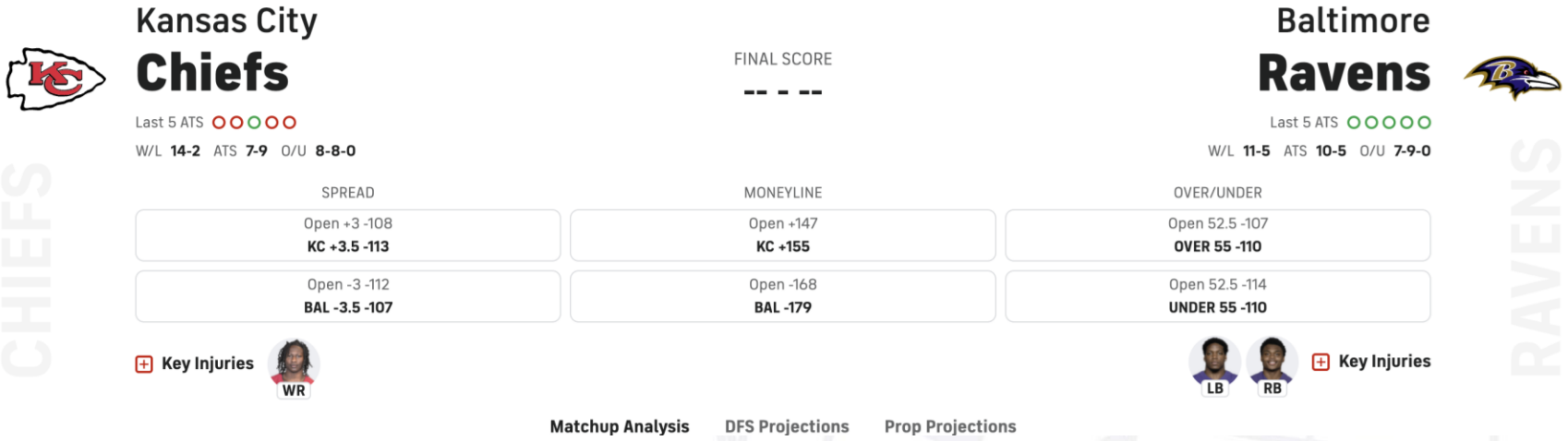
Let's look at some external factors that may affect tomorrow's season opener!



8x576.jpg (768×576) (vegasalways.com)

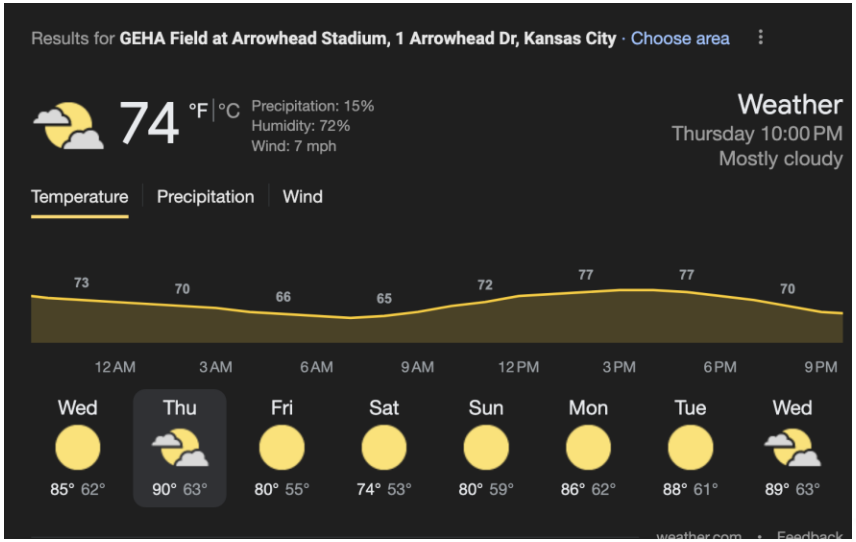
Second Dash Demonstration, 2

- From the Sportsline graphic below (which we used solely as a prediction tool and for no betting related purposes!) we can see that Baltimore is favored for tomorrow's game at –3.5, with an over/under line of 55



Second Dash Demonstration, 3

- It looks like it will be fairly warm for tomorrow's game



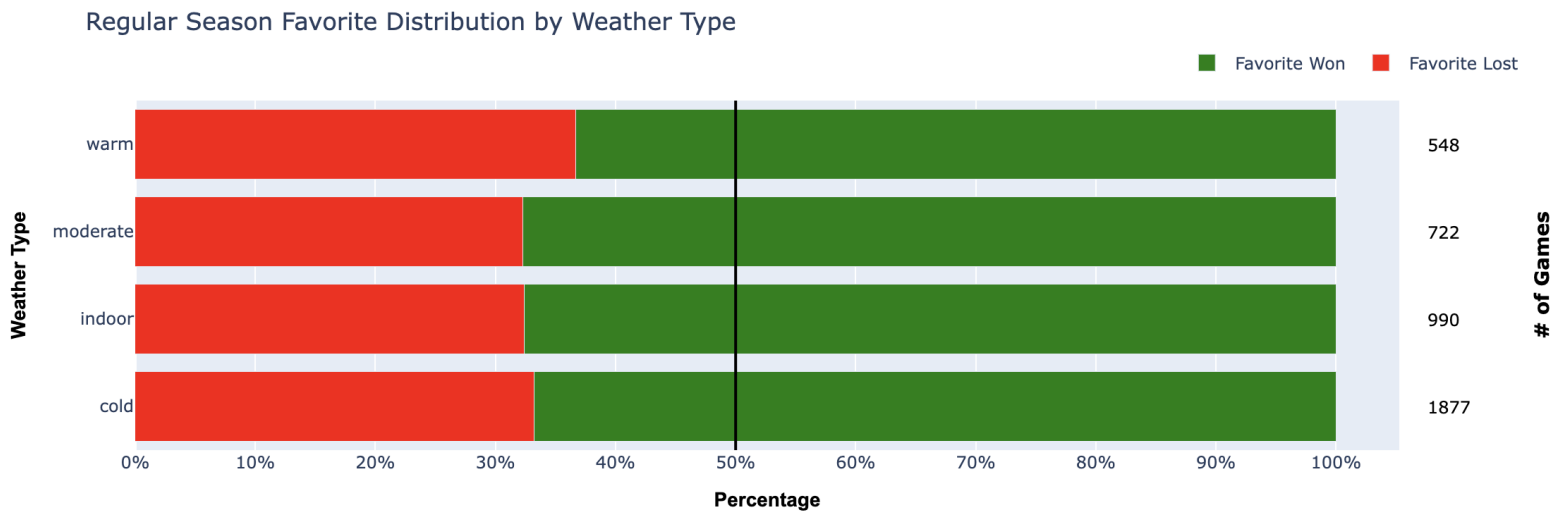
Second Dash Demonstration, 4

NFL Analysis Based on Location Factors

Favorite

Weather Type

◉ Regular Season ○ Playoffs



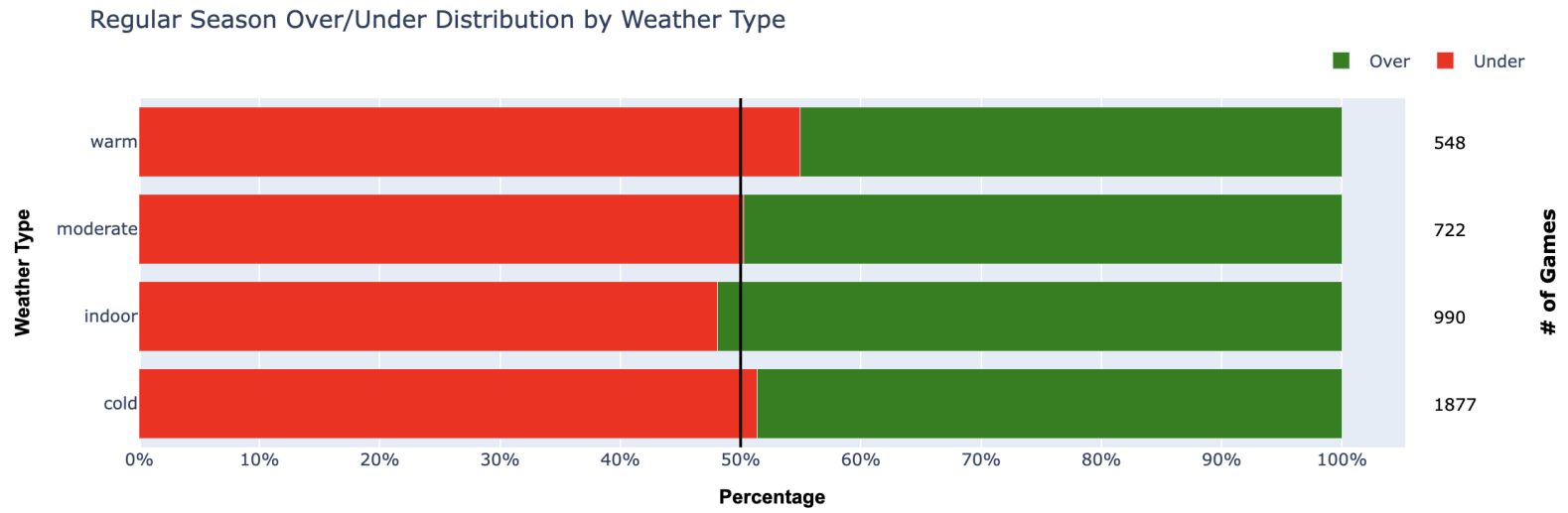
Second Dash Demonstration, 5

NFL Analysis Based on Location Factors

Over/Under

Weather Type

◉ Regular Season ○ Playoffs



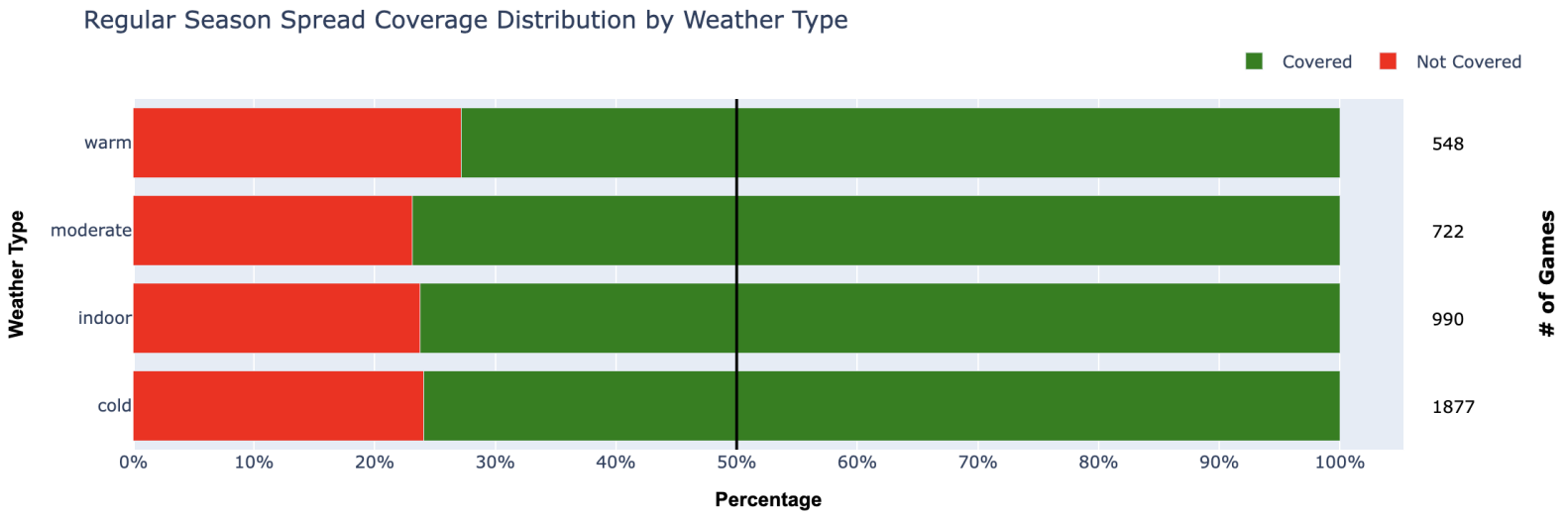
Second Dash Demonstration, 6

NFL Analysis Based on Location Factors

Spread Coverage

Weather Type

◉ Regular Season ○ Playoffs



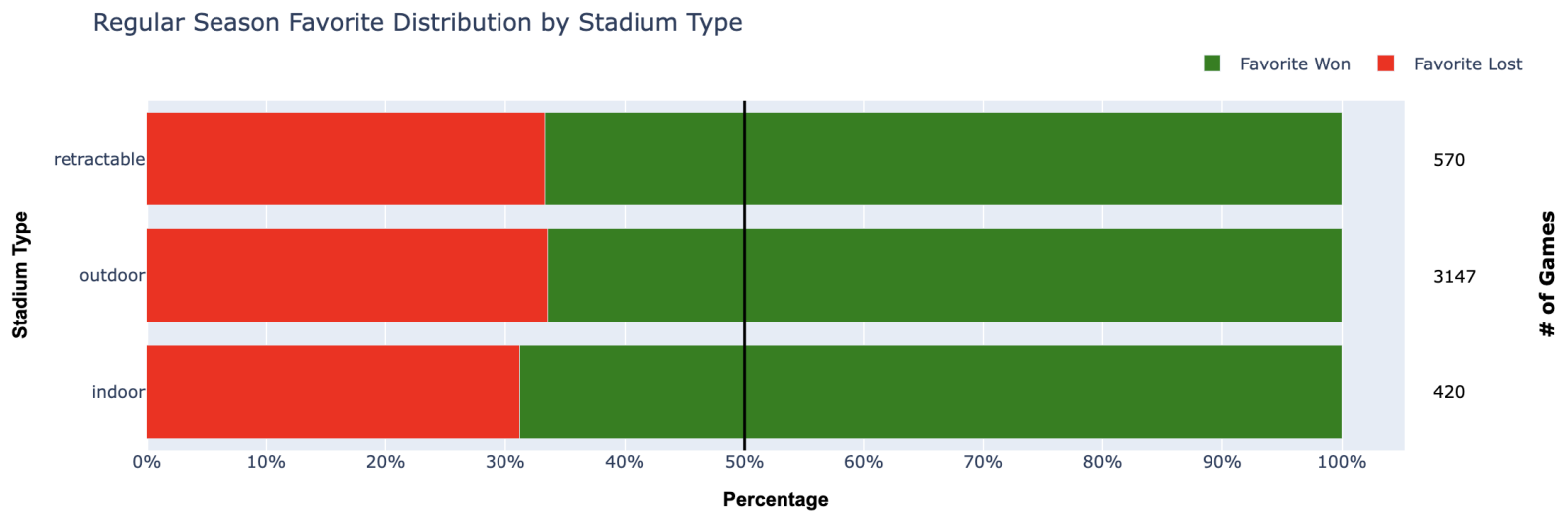
Second Dash Demonstration, 7

NFL Analysis Based on Location Factors

Favorite

Stadium Type

◉ Regular Season ○ Playoffs



Second Dash Demonstration, 8

- What can we take away from this?
- Vegas is still extremely good at predicting spread, over/under, and picking favorites
- External factors do not substantially shift the outcomes one way or another

Future Work

- Expanding to look at other sports
- Incorporate more real-time data

Questions?