# NFL Betting Data

Group 2:
Gage Anderson
David Ferguson
Brandon Ibarra
Quintele Jackson

# Introduction

What is a spread?

- A spread is the number of points the favored team is expected to win by.
  - Example: Tom bets on NFL games, and he is presented with the following scenario.
    - Team A is favored over Team B. The spread is -6.5. That would mean Team A is expected to beat Team B by at least 6.5 points. If Tom bets on Team A to “cover” the spread and Team A beats Team B by 6.5 points or more, the spread is “covered”.
    - If you look at the spread from Team B’s point of view, their spread would be +6.5. If Tom bets on Team B to “cover” the spread, then Team B would have to lose by less than 6.5 points or win the game.

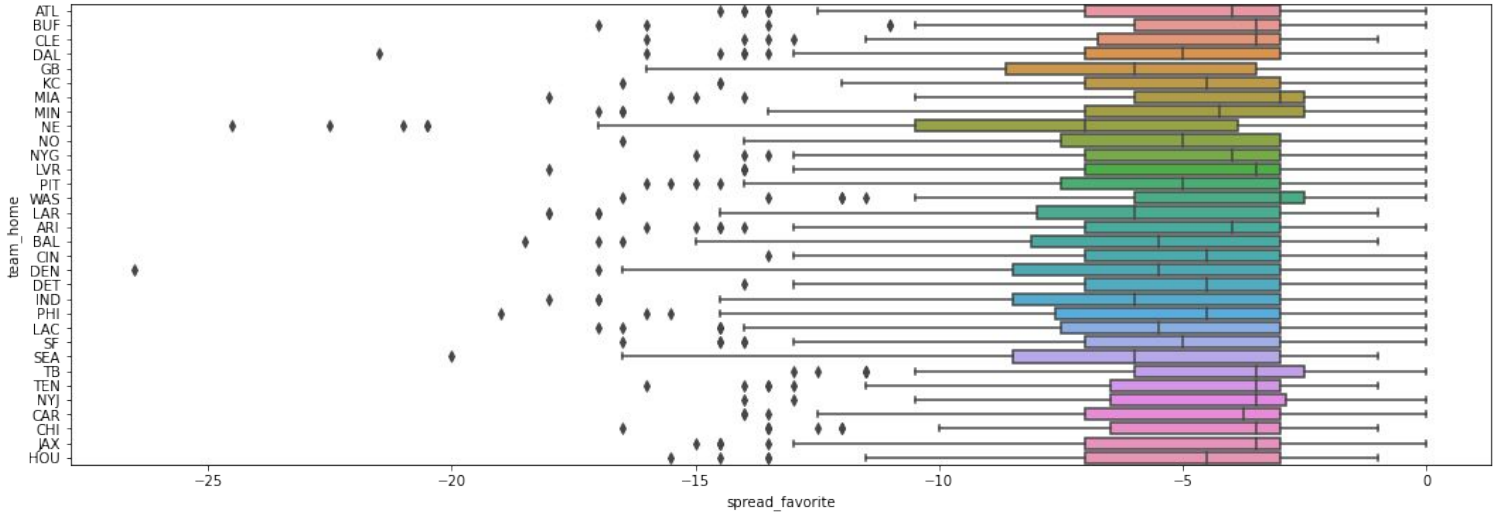
# Data Breakdown

Our analysis used three datasets constrained to a time frame of 2000 to 2019

- A list of all NFL Stadiums used, excluding neutral sites
- A list of all 32 NFL teams
- A list of all NFL games, excluding neutral site games and playoffs, including the scores, the favored team, the spread, and the over/under

# How often do favored teams win?

Looking at the data lots of team that where favored won their games as the graph below shows.This graph shows by how much points the favored team would win compared to the team that is not favored.As you may see denver broncos and new england patriots have the 2 highest spreads out of all the teams

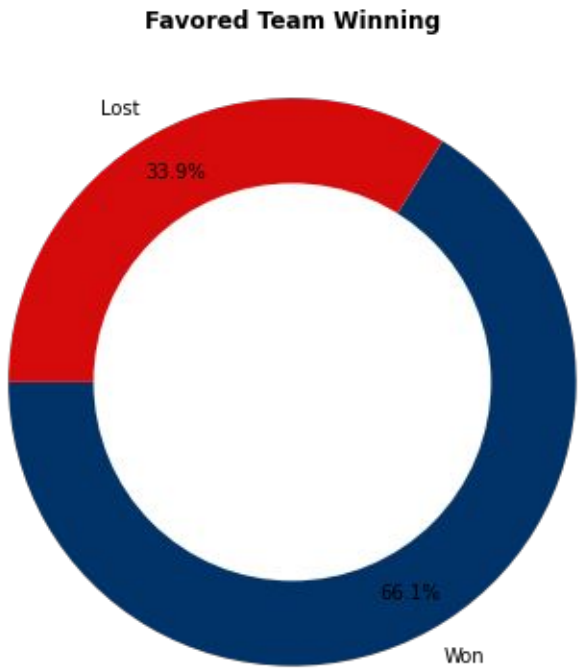


# Favored Continued

Attached are 2 screenshots. The first is a True/False, where True means that the favored team won, while False means the underdog team won. The second shows the count of all favored teams who won, with the Patriots having the most. This makes sense due to the Brady/Belichick dynasty. “PICK” means no team was favored to win.

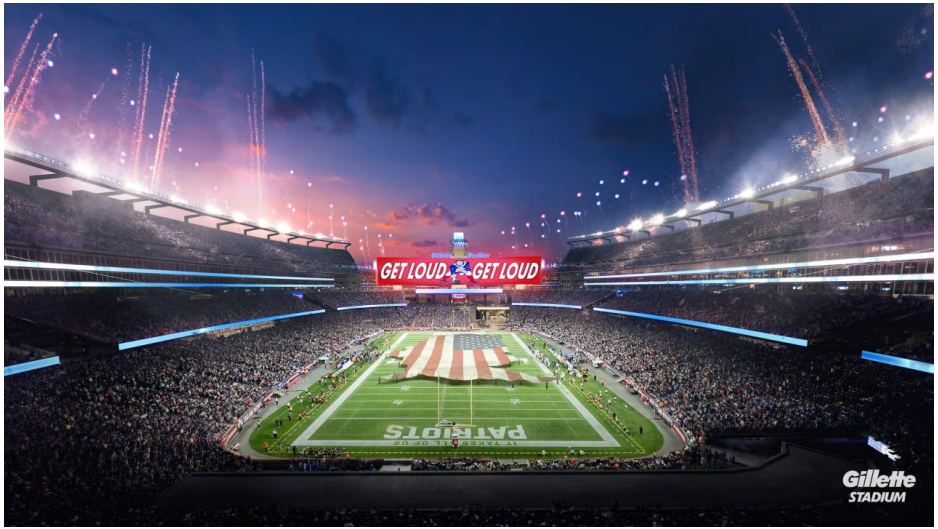
```
True     3347
False    1714
Name: favorite_won, dtype: int64
```

```
NE      197
IND     158
PIT     157
GB      156
BAL     147
PHI     143
DEN     134
NO      132
SEA     131
DAL     116
MIN     110
LAC     109
KC      109
ATL     108
NYG     106
TEN     101
CAR      99
LAR      98
CHI      93
SF       93
NYJ      86
CIN      83
TB       81
HOU      79
BUF      74
MIA      73
ARI      70
DET      61
JAX      61
WAS      60
LVR      60
CLE      42
PICK     20
Name: team_favorite_id, dtype: int64
```

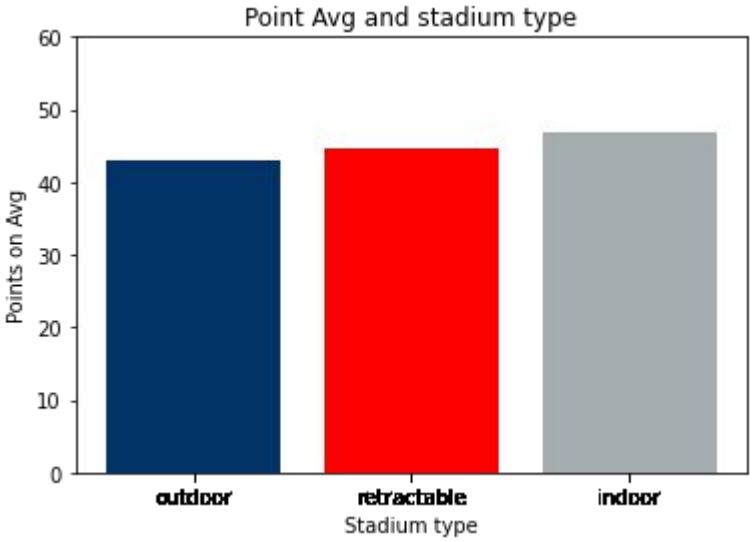


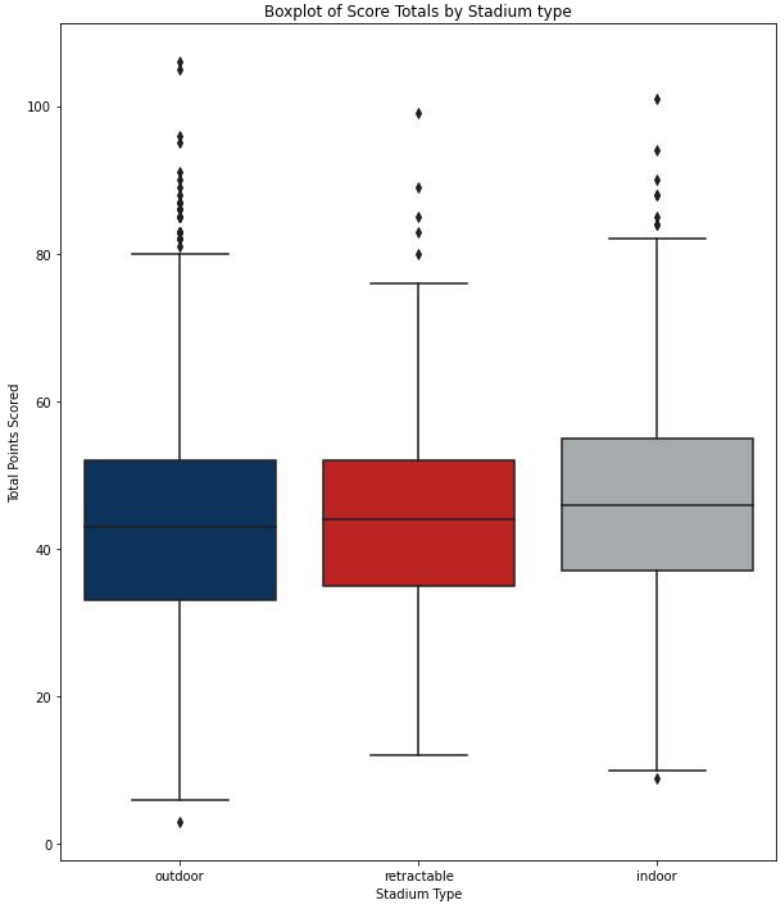
## Do different types of stadiums have more average points scored?

The scores on average are different but not by much which is surprising since I thought weather would be a big factor in football. Indoor scored more on average which is expected but not by much.

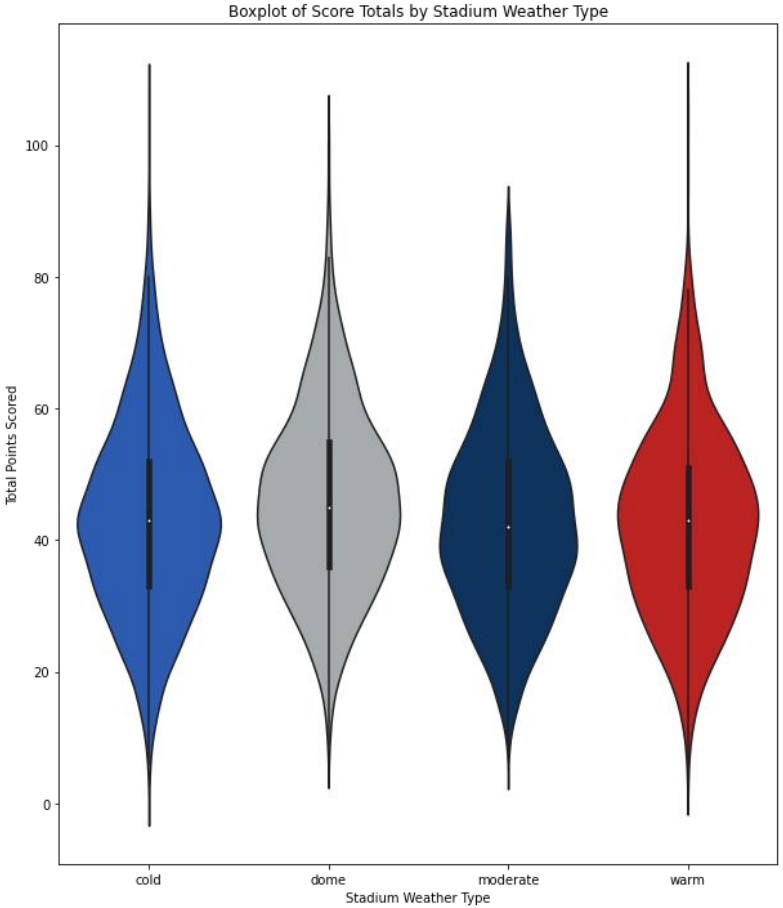


# Bar Graph and Box Plot



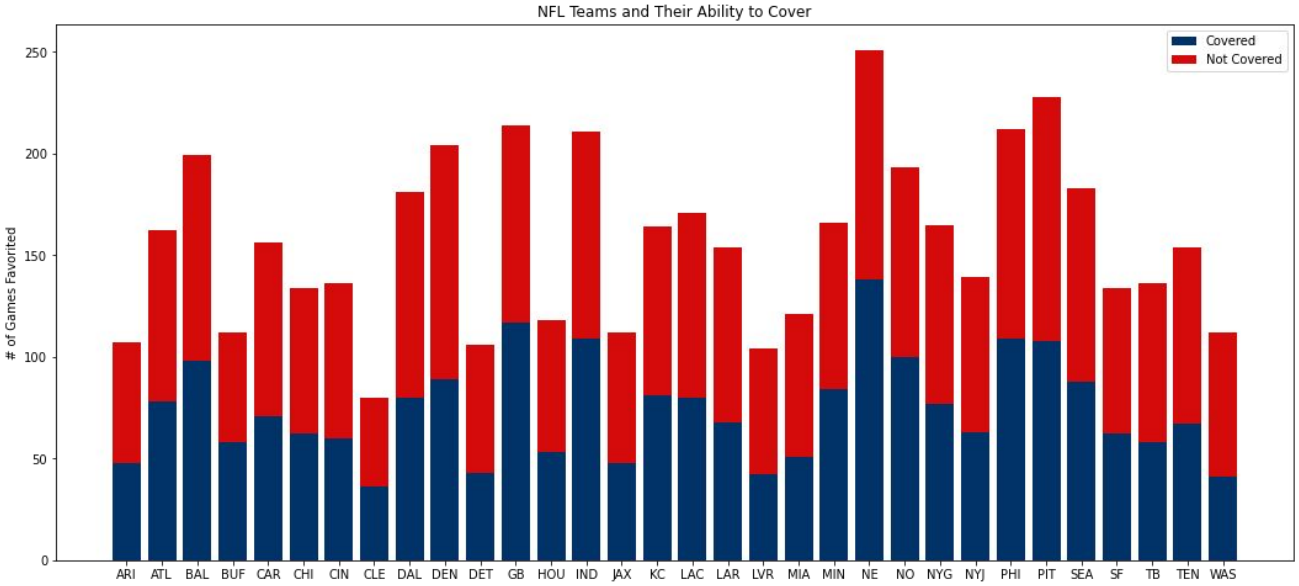



# Violin Plot


Boxplot of Score Totals by Stadium Weather Type
Total Points Scored
100
80
60
40
20
0
cold
dome
moderate
warm
Stadium Weather Type

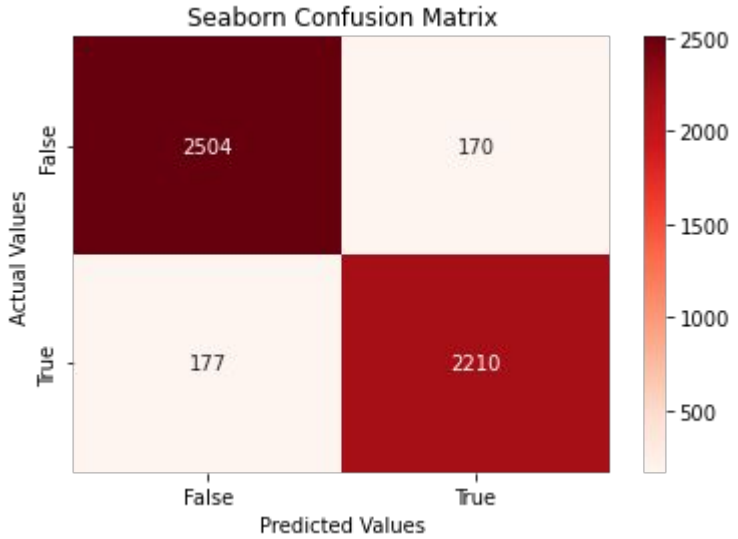
# Regression Analysis

The goal of our Regression Analysis was to see how likely it would be a for the favored team to cover the spread. We chose a Logistic Regression due to the ability to use multiple features.



# Confusion Matrix

Using RandomForestClassifier, we fed the predictor the schedule year, the spread, if the home team was the favored team, and who the favored team was. Analyzing this data, the regression model was able to accurately predict if a team would cover the spread 93% of the time.



# Usability

Through the use of get_dummies, the Team Favorite column was able to be converted to a binary value for each of the 32 NFL teams.

| | schedule_season | spread_favorite | home_favorite | team_favorite_id_ARI | team_favorite_id_ATL | team_favorite_id_BAL | team_favorite_id_BUF |
|---|---|---|---|---|---|---|---|
| 0 | 2000 | -6.5 | True | 0 | 1 | 0 | 0 |
| 1 | 2000 | -1.0 | True | 0 | 0 | 0 | 1 |
| 2 | 2000 | -10.5 | False | 0 | 0 | 0 | 0 |

Using this function and the RandomForestClassifier, it is possible to manipulate the previously mentioned features columns to predict whether a team would cover the spread.

For example, if the season is 2019, the favored team is the home team by 5 points, and the home team is the eventual Super Bowl winning Kansas City Chiefs, the model only gives the Chiefs a 35% chance to cover the spread. This roughly lines up with real life data, as the Chiefs only covered around 45% of the time when they were favored by around 5 points.

# Conclusions

How often do favored teams win?

- Favored teams win around 66.1% of the time. You are more likely to win money if you bet on the moneyline (betting on the favored team).

Do different types of stadiums have more average points scored?

- Yes, games in indoor stadiums on average score more total points, but the difference is marginal.

Regression Analysis

- You are more likely to win money if you bet against the favored team to cover the spread.

# Work cited

Kaggle nfl sport betting