CS620 Data Project

NFL Elo Rating Prediction Power

Bryan Kors

Demetrius Wright

Wes Hicks

Abstract

FiveThirtyEight is a well-known website that creates and shares statistical models on political opinion polls, economics, and sports betting amongst other popular topics. Up until 2022, FiveThirtyEight maintained sports models on projected game winners, season champions, and playoff contenders. What model primarily used was their Elo rating model.

The Elo rating system has its roots in chess. The rating system was created by a chess master named Arpad Elo as a way to measure a chess player's strength. The rating system rates each player (or team in the case of the NFL) based on their performance in previously rated games. Once a game or match is completed, the player's Elo rating is adjusted based on the outcome. If a higher rated player wins to a much lower rated player, this outcome is expected and thus a smaller overall impact to the players' Elo ratings. On the other hand, if the much lower rated player were to win the match against a much higher rated player, this would have a larger impact on the post-match Elo ratings.

In FiveThirtyEight's case, the organization created an Elo rating system for NFL teams based on their historical wins and losses. An "average" strength team is rated roughly 1500 while higher strength teams are above 1500 and lower strength teams below 1500. Similar to the Chess ratings, a team's Elo rating adjustment after a game is driven by the outcome of the game versus what the expected outcome was given the Elo rating difference.

The Elo rating system is a popular rating system that is used well past the NFL and chess. With this being a widely used system, we intend to explore the predictive power of FiveThirtyEight's rating system of NFL teams.

FiveThirtyEight has over 30,000 NFL games, their outcomes, and each team's Elo rating for each game. We will use this data to explore the following questions:

- How well does FiveThirtyEight's Elo rating predict winners?
- Can a previous season's Elo rating be used to predict next year's playoff contenders?
- Do certain NFL divisions historically have higher Elo ratings than others?
- Is there a relationship between an NFL team's Forbes list valuation and Elo rating?
- How are we expected to end up if we bet the money line on every game favorite?

An NFL team's Elo rating and their reported Forbes list valuation was explored and will be reported by Wes Hicks. The money line simulation of game favorites was explored and will be reported by Demetrius Wright. Finally, the predictive models using historical Elo ratings was explored and will be reported by Bryan Kors.

Overview of the Datasets

Main Dataset - FiveThirtyEight NFL Elo Ratings

FiveThirtyEight had created multiple datasets in CSV format on their github page (https://github.com/fivethirtyeight/data). We used their nfl_elo.csv dataset to explore their Elo ratings (https://github.com/fivethirtyeight/data/tree/master/nfl-elo). This dataset includes all NFL games dating back to the 1920s with scores, startings quarterbacks, home team/away team, playoff vs regular season, etc. This dataset also includes FiveThirtyEight's Elo rating of each team and their respective starting quarterback both pre-game and post-game (adjustments made based off game results.

This dataset is in a csv format. The data was extracted through the Pandas library. When pre-processing this dataset, we focused strictly on the 1970 season forward. In 1970, the NFL/AFL merger occurred which created one league with consistent rules, the NFL. Analyzing data prior to the merger risks inconsistencies with teams being in different leagues with different sets of leagues rules. Additionally, multiple NFL analysts report player or team stats as "since 1970" or "since the merger." The seasons we focused on are consistent with what has historically been analyzed.

Secondary Dataset - NFL Division/Conference Information

A large portion of FiveThirtyEight's NFL information is from Pro-Football Reference (https://www.pro-football-reference.com/). Pro-Football Reference houses a plethora of NFL figures and statistics dating back to the 1920 season. We used this website to gather divisional and conference information.

This dataset was pulled together using the BeautifulSoup and Requests libraries. There were a few important considerations as the dataset was put together. Pro-Football Reference places a 20 requests per minute limit on how many times a user can send requests to their sites. Any violation of this will "put you in jail for an hour."

Being aware of this limit and being courteous to avoid overwhelming their sites, we ran our script once to visit respective sites every four seconds (avoids exceeding 20 requests a minute) and saved the divisional information to a CSV on github. This also allowed us to easily access the information immediately through Pandas versus waiting on a four plus minute web scraping script to pull together the data needed.

Secondary Dataset - Forbes NFL Team Valuations

Forbes is well known for its valuations of people, companies, and even sports teams. Since at least 2012, Forbes has been compiling financial incomes and debt information for NFL teams and creating "valuations composed of the monetary worth of the sport, market, stadium deals, and brand." This company uses numerous variables to create an annual valuation for the teams. Organizations such as CBS, Sports Illustrated, and NFL have recognized the insight and contextual trends that the annual valuation reports have conveyed.

A consolidated table of each team's valuation from 2012 to 2021 was found on Wikipedia. Unlike Pro-Football Reference, Wikipedia does not have request limits. Additionally, this table is located on a single page and was extracted with BeautifulSoup in a relatively short period of time. This information was pulled on demand given the above factors.

Secondary Dataset - Historical NFL Game Money Lines

Sportsbook Reviews Online contains historical sports betting information including historical money lines for NFL games for the 2007 season through the 2021 season. The tables on the available seasons/pages are formatted by individual teams home/away status, date, team, quarter scores, final scores, and money line. This information was extracted using the BeautifulSoup library. While we did not have request limits similar to Pro-Football Reference, we did receive 404 responses which means "Not Found" when using requests.get(). After researching, we discovered that this error was driven by not passing a user-agent in the header parameter in the requests.get function. After passing one in this parameter, we were able to successfully scrap the necessary data through the BeautifulSoup library.

Data Wrangling and Pre-Processing

Noted previously, our primary dataset for NFL team's Elo ratings overtime was stored in a CSV format on FiveThirtyEight's github page. This dataset was loaded through pandas.read_csv() function. The initial data frame included seasons dating back to 1920 while we intended to use the 1970 season to the most recent season available. While we initially designed this project to explore just NFL team Elo ratings, it is important to note that FiveThirtyEight did not begin Elo ratings for starting quarterbacks until the 1950 season. If all NFL seasons were to be explored with quarterback Elo ratings being considered, the earliest season available would be 1950.

Below outlines pre-processing that was done to the Elo ratings dataset:

- Dropped all games for seasons prior to the 1970 season
- Filled "NaN" values in the playoff column with 'r' to represent regular season games.
  - Regular season game - 'r'
  - Wild card playoff game - 'w'
  - Divisional playoff game - 'd'
  - Conference championship game - 'c'
  - Super Bowl game - 's'
- Dropped 'quality', 'importance', and 'total_rating' columns from the dataframe (unnecessary for our project)
- Dropped unneeded Elo ratings including Elo probability, post-game Elo ratings, and quarterback ratings from the dataframe
- Updated column headers to be more readable (i.e. team1 = 'home', team2 = 'away', updated team1 & 2 ratings to home & away ratings, updated qb1 & 2 to home & away qb, updated score1 & 2 to home & away score)

The next dataset extracted and pre-processed was the conference and divisional assignments for each NFL team from the 1970 season to the 2022 season. The conference and divisional information is located on a single site for each season, 53 sites total. Recall from the previous section, Pro-Football Reference has a 20 requests per minute limit before you are "put in jail for an hour." To adhere to this rule, we used time.sleep(4) at the end of each page's web scrape to pause for at least 4 seconds before requesting to go to the next season's website. With this pause, we visited 20 websites in 1.33 minutes (greater than Pro-Football Reference's limit) and all 53 websites in roughly 3.5 minutes. To avoid having to wait 3.5 minutes each time we needed to populate this dataset, the dataset was saved to a CSV on Bryan's github repo (https://raw.githubusercontent.com/bryan255/NFLeloAnalytics/main/nfl_division_info.csv).

From 1970 and 2022, there have been new teams that joined the league (i.e. Carolina Panthers and Houston Texans) and other teams that have relocations (i.e. Oakland Raiders to Las Vegas Raiders, St. Louis Rams to Los Angeles Rams). Team names are denoted by their three letter acronym. For example, the New York Giants three letter acronym is NYG. Below outlines the challenges we faced and how worked through it:

- Three letter team acronyms in divisional/conference information not matching three letter team acronyms in Elo rating dataset (ex: Tampa Bay Buccineers is called TAM in the divisional/conference information but TB in Elo rating dataset)

    - Solution: all mismatch team names were identified, team names were iterated through in divisional/conference dataset to align with Elo rating dataset
- Teams relocated causing team names to change (ex: San Diego Chargers relocated to Los Angeles; name changed from SDG to LAC)
    - Solution: Since this project focuses on simply a team's Elo rating over time and not the impact of a geographic change, team names were were aligned to be consistent across all seasons and match the Elo rating dataset
    - Ex: The Chargers were in San Diego until the 2017 season when the team moved to Los Angeles. The team acronym was changed to LAC (Los Angeles Chargers) for all seasons to align with the Elo rating dataset

The money line dataset was extracted from Sportsbook Review Online using the BeautifulSoup library. We were able to extract a list of measures from the website including the date of the game, the home team, the away team, each team's money line, and who won the game. Below outlines how money line wagers occur:

- The team with the lowest money line is considered the favorite (i.e. -150 vs +200; the team with the -150 money line is considered the favorite)
- If the money line is negative, it can be interpreted as the wager needed to win $100 (i.e. -150 = place a wager of $150 to win $100)
- If the money line is positive, it can be interpreted as the winning amount based off a $100 wager (i.e. +200 = win $200 by placing $100 wager)

While we did extract a small amount of information from the website, we did face challenges. Below outlines our challenges and how we worked through them:

- The website returned a 404 code which means "Not Found" when using requests.get()
    - We passed a user-agent in the headers parameter. This returned a 200 code which means "OK" and were able to successfully scrape from each needed site
- The table layout on the website was not conducive to the layout we needed for our project. Specifically, each game had two rows instead of one for the home team and away team scores and money line.
    - To solve this issue, we used a "sliding window technique" to place the necessary measures for each game into a single array then pass it into the data frame to return a single row for each game

The final dataset was the Forbes list of NFL team valuations from 2012 to 2021. This dataset was listed in a table on a Wikipedia page that was pulled together using BeautifulSoup. The format of the site and table was well put together making scraping the information from the page simple and quick.

<u>Money Line Simulation and Analysis</u>

The money line bet, as mentioned earlier, is a straightforward wager placed on which team may win a game. Despite the bet's lack of complexity, many conclusions can be drawn from analyzing financial returns of teams and, by association, their wins and losses. In this approach within the project, we have developed a script to simulate what might be the profit and returns made for NFL franchises, as well as season favorites and underdogs. The main goals of the money line simulation are as follows:

- Determine the viability of sports betting as a form of financial investment.
- Uncover whether seasonality is apparent in particular franchises.
- Project potential future profits for sports betting persons.
- Gauge which franchises have most consistent returns on investment over the years

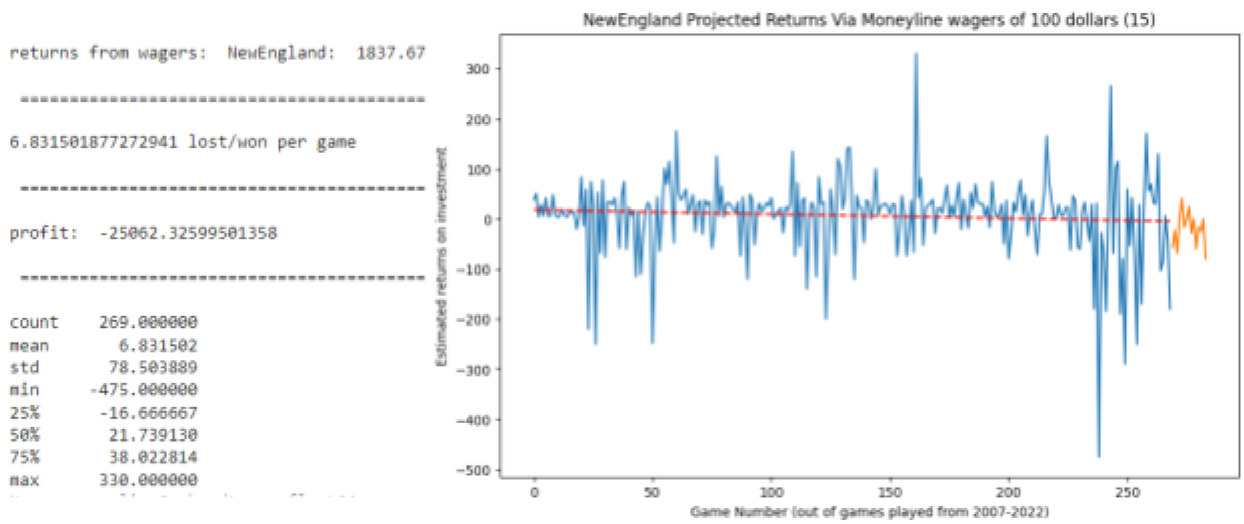
Upon cleaning and processing the data, it had become immediately apparent that there was lacking information from the, now, developed frames. Information such as, which team was considered the favorite or underdog, which team had won or lost a particular game, as well as the aforementioned calculation of money profited was all necessarily added to the frame in order to perform the forecasting of the placed bets.

Methods for forecasting the data frame come by way of time-series analysis using Holt Winters exponential smoothing model. This method was chosen for its ability to adjust weight ($\alpha = 0.12$) in favor of newer data. Implementing this model was done with the intention of replicating the changing of players, coaches and staff who work on a team for some time. As for the chosen weight, as a result of its predictions being very close to that of current events in football, it was found suitable to maintain this weight.

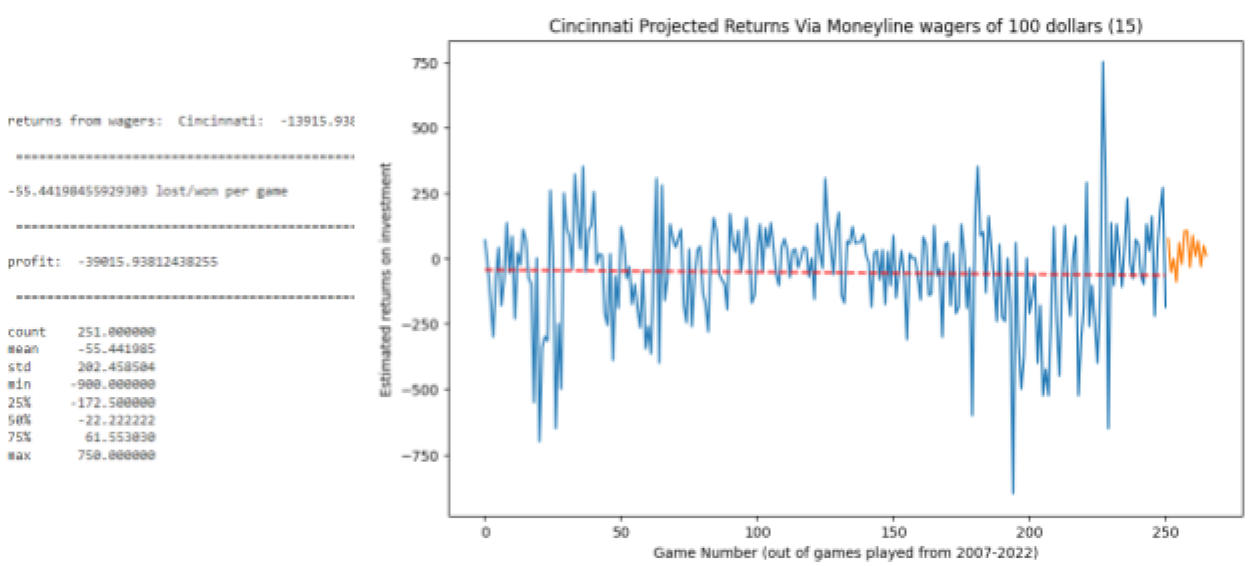
When analyzing the graphs, there are several important factors that are telling about a team's performance and how that may apply to one's betting strategy. Things that should be considered are:

- Linear Regression Line: In the circumstance in which the linear regression is below 0, this indicates an increased likelihood of a potential future loss. This means that teams are consistently losing money, as opposed to a team that exists largely above 0. In the case in which a team's standard deviation exists above 0, then they are a team with higher rates of return.
- Standard Deviation: This illustrates consistency within profits and losses. If a team has a small standard deviation, they typically do not have good returns, though they are likely to have a lower cost of losses. Additionally, teams who vary widely among their linear regression line, are likely to have consistent wins and losses. This would result in a team with more upsets (underdog winning) when both the favorite and underdog.
- Curves/ Trends in deviation: Visible "curves" in the graph may indicate signs of seasonality - variations that occur in the form of an interval or cycle. In this case, management and staff of a franchise may change every so often.

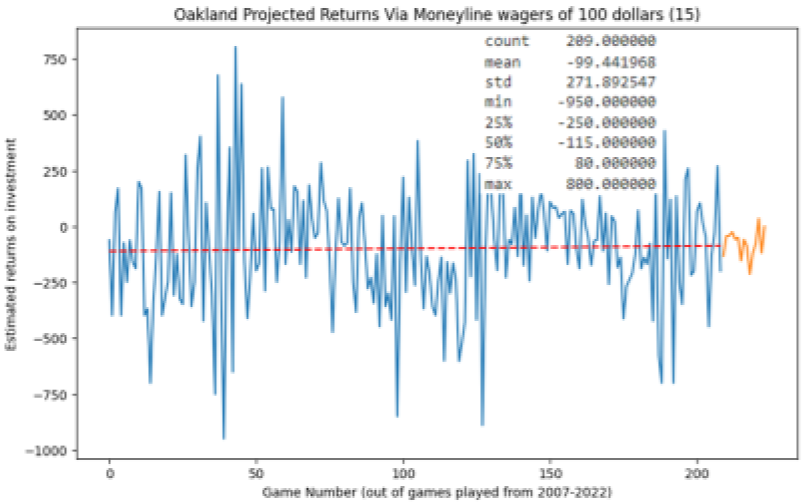
Regarding the output of the script, trends are most apparent in teams with consistent wins and profits. Teams such as the New England Patriots have a considerable record, having a majority of their returns above 0, something not terribly common in the league. This team has been noted in recent years as being one of the best in the league. Although they currently boast an overall return of 1837.67, the forecast was capable of determining their poor performance in the 2022-2023 season, which has been confirmed by their past and current standing in the American Football Conference (AFC) as one of the worst in the east.



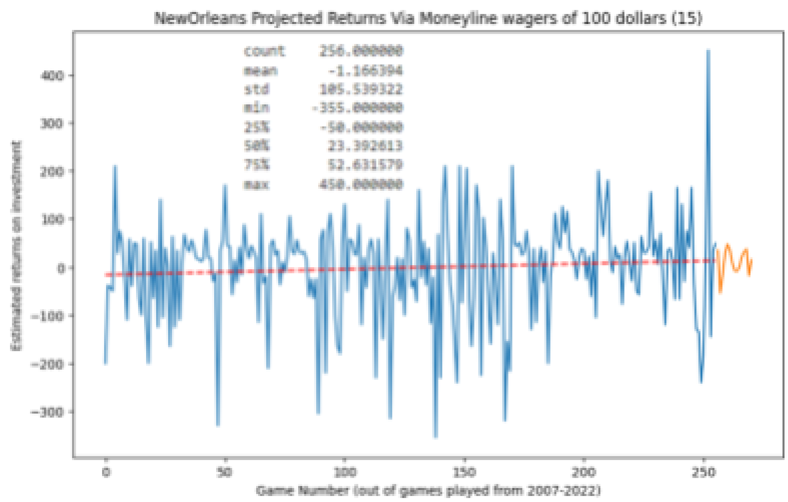
Another notable team is that of the Cincinnati Bengals. As shown in the graph above, the Cincinnati Bengals are a poor team, who, in recent years, have begun to improve from an investment perspective. Not only is this evident on the graph, but they were in actuality the best in their region as of 2022-2023 . As of the 2023-2024 football year, the Cincinnati Bengals have begun to decline in performance, which again is depicted in the graph, and evident in the real world.



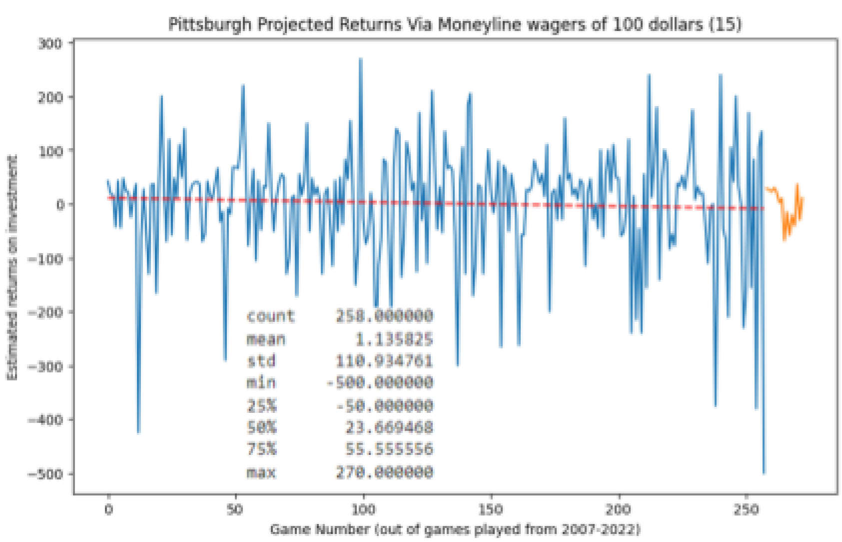
Brief Analysis of the performance of particularly unique teams are as follows:



The Oakland Raiders seem to increase in returns roughly every 3 years, the decrease every 3 years. This franchise is worth investigating to understand why they lack in consistency. This is especially evident when regarding their standard deviation of 271 - one of the highest in the league



The New Orleans Saints is projected to perform well in the future, gradually increasing in returns of wagers made on the money line. Regarding progress over the last 15 years, they are statistically one of the most improved, with a low standard deviation as compared to other teams. This team would be a considered a more reliable and consistent long term investment

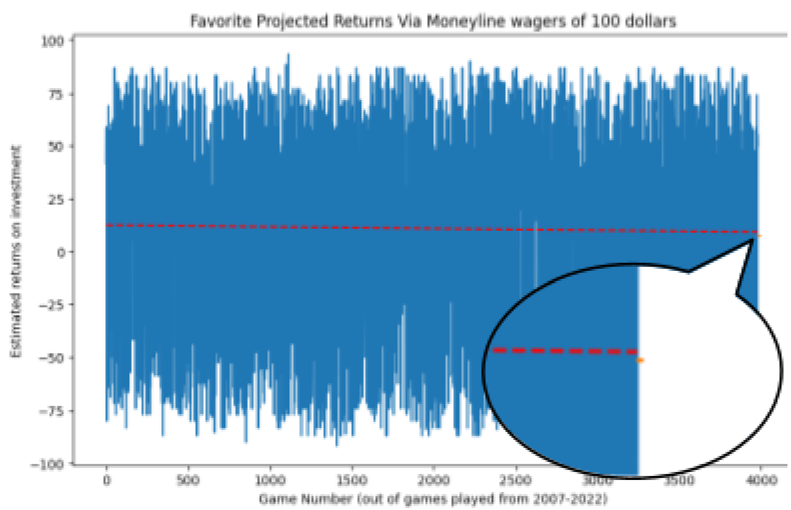


The Pittsburgh Steelers exhibit low amounts of deviation, yet low rates of return. If an individual would like to invest in a team such as this, they should expect consistency over the course of their years

Referring back to the original questions posed regarding money line simulation and analysis:

- The most notable franchise that exhibits seasonality are that of the Oakland Raiders.
- Sports betting can be a viable form of financial investment, given that individuals have a higher tolerance for risk. This can largely be reduced by adequately training forecasting models with multiple classes of data. Despite this, by only analyzing the average rates of return on the money line in the last 15 years, we are capable of determining future trends in games with reasonable certainty.
- We can gauge which franchises have consistent rates of return throughout the decade, notably the Pittsburgh Steelers and New Orlean Saints.

```
returns from wagers   43438.1322624376

 ====================================

10.919590815092409 lost/won per game

 ====================================

profit:  -354361.8677375624

 ====================================

count    3978.000000
mean       10.919591
std        49.700608
min       -91.743119
25%       -38.280116
50%        26.315789
75%        50.000000
max        93.457944
```



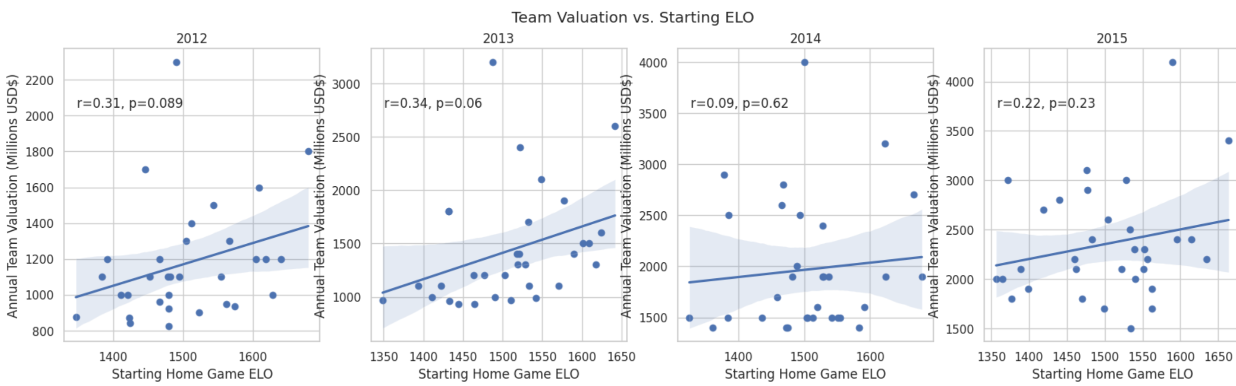
As a result of complications with competing plots, this diagram must remain in this view

Due to sporadic, inconsistent, and non-seasonal data as a result of large fluctuating wins and losses, forecasts of favorite and underdog profits are depicted linearly. Although not ideal, the information provided still grants us the ability to make estimations for future earnings. As seen above, placing a bet on either the underdog or the favorite is both very volatile, but very predictable. The consistent bet placed on the favorite yields a low risk low return. One may immediately assume that picking the favored team to win a game is financially opportunistic and fruitful, but clearly it is quite the opposite.

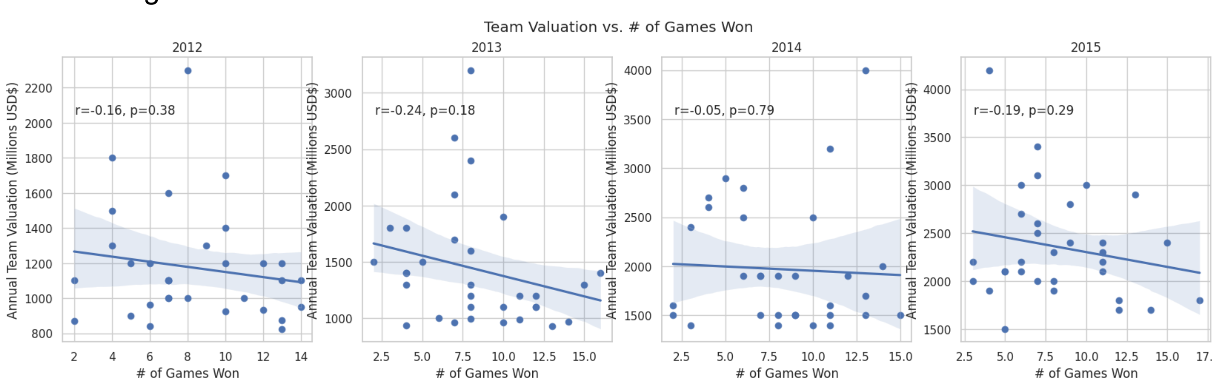
Forbes List NFL Team Valuations Linear Regression and Analysis:

As previously mentioned, we acquired the annual valuation of each NFL team, consisting of figures based on team performance, marketing and brand revenue, and stadium deals. Using a relatively simple BeautifulSoup script, we scraped a Wikipedia page containing a compiled table of all NFL team valuations from 2012 to 2021. This data was compiled directly into a Pandas.Dataframe() object.

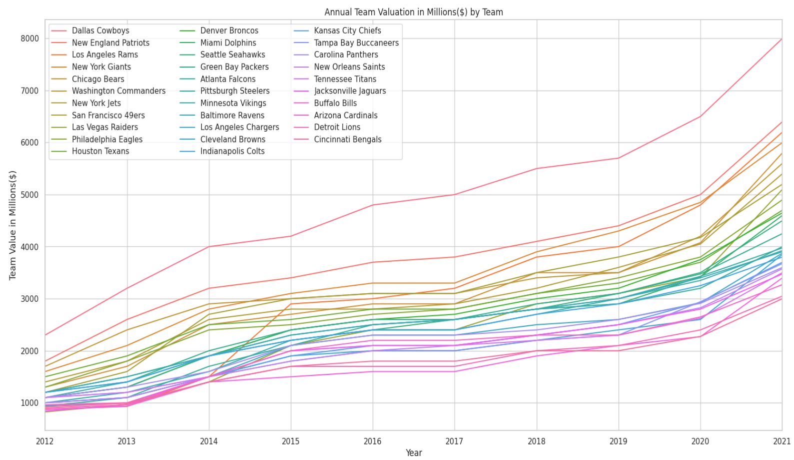
Due to circumstantial factors, such as injuries, changing of coaching staff, etc, we hypothesized that the monetary worth of each team would not likely be reflected in their average performance evaluation but possibly the initial value. Our first objective was to utilize a seaborn regression plot to determine the strength of correlation between the initial season Elo values of each team and their respective annual valuation. While exploring the R-values between these two factors and observing the resulting plot, we determined that the monetary worth of each team had a very little, if not insignificant, impact on the first Elo value of each season.

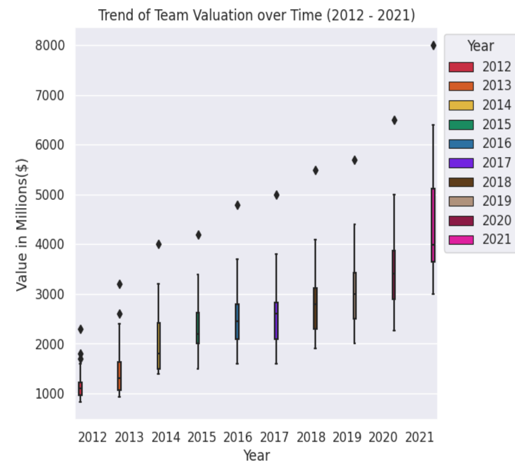


We pivoted and decided to explore if there was a different performance metric that had a stronger relationship with the annual valuation of each organization. We took advantage of the newly formed “Winner” column that was developed to serve the function of determining overall winner and playoff predictions, to see if annual valuation had an indirect impact on the number of wins per season. The regression plot was adapted to replace the ‘Starting Home Game Elo’ on the horizontal axis with the sum of seasonal wins.



Surprisingly, the adapted plot did not convey a stronger relationship, but instead showed a negative correlation in some seasons. The outcome of the two linear regressions performed on team valuations gave us the suspicion that team performance may not have a consistent nor predictable effect on organizational worth. Instead we determined that we would take a broader look into the trend of team valuations over time. The line chart below conveys the overall trend of organizational worth from 2012 to 2021, most of the teams follow a consistent positive trend. A supplemental box plot was also created to display the distribution of the organizational worth over time. From this plot, it is evident that the highest valued teams have consistently created a more pronounced separation from the lower valued teams over time.

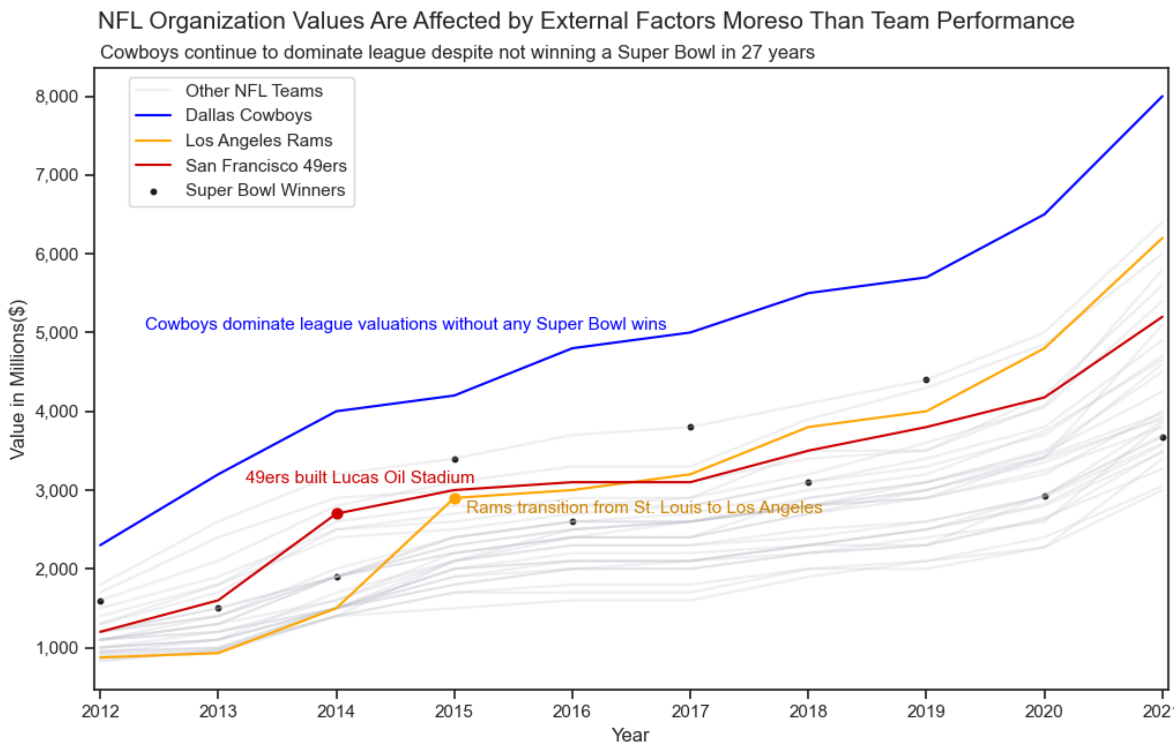




Within the 32 teams plotted on the line chart, three of them were particularly interesting and were highlighted in the line chart below. In this plot, you will see that the Rams, 49ers, and Cowboys have been selected from the other teams who have been grayed to reduce visual clutter. After concluding that their valuations were likely not influenced by their team's performance, we investigated what other factors may have spiked their valuation by over 1 billion dollars in under a year.

As indicated by the plot below, the Rams had an in valuation from 1.5 billion to just under 3 billion dollars in revenue and overall organizational worth. This increase occurred in the same year as their transition from St. Louis to Los Angeles, which we are confident played a large part in their valuation increase. A team from the same division, the 49ers, were valued at just over 1.5 billion dollars and increased to 2.7 billions dollars in 2014. We believe that this increase can be attributed to their newly built and opened stadium located in Santa Clara in the same year.

Finally, the Dallas Cowboys, a team who had not won a superbowl since 1996 (as indicated by the points plotted in black), has continually dominated the league in revenue and organizational worth. I believe that this is a fair supporting incident that further validates that team performance has a small if not insignificant effect on organizational value.



NFL Organization Values Are Affected by External Factors Moreso Than Team Performance
Cowboys continue to dominate league despite not winning a Super Bowl in 27 years
Other NFL Teams
Dallas Cowboys
Los Angeles Rams
San Francisco 49ers
Super Bowl Winners
Cowboys dominate league valuations without any Super Bowl wins
49ers built Lucas Oil Stadium
Rams transition from St. Louis to Los Angeles
Value in Millions($)
8,000
7,000
6,000
5,000
4,000
3,000
2,000
1,000
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
Year

NFL Elo Rating
Game Winner and Playoff Predictors:

There were two major predictor questions posed and one question over time in regard to FiveThirtyEight's NFL Elo rating:

1. How well does FiveThirtyEight's Elo rating predict winners?
2. Can a previous season's Elo rating be used to predict next year's playoff contenders?
3. Do certain NFL divisions historically have higher Elo ratings than others?

The final, processed Elo rating dataframe was merged with the divisional and conference information dataframe to create one single dataframe to work with. Once these dataframes were merged, each NFL team had a division and conference attribute for each season that could be grouped to explore historical Elo ratings at a higher aggregated level. This was done by using the groupby function to aggregate NFL teams' by division and conference and determining the mean for each season's attribute respectively.

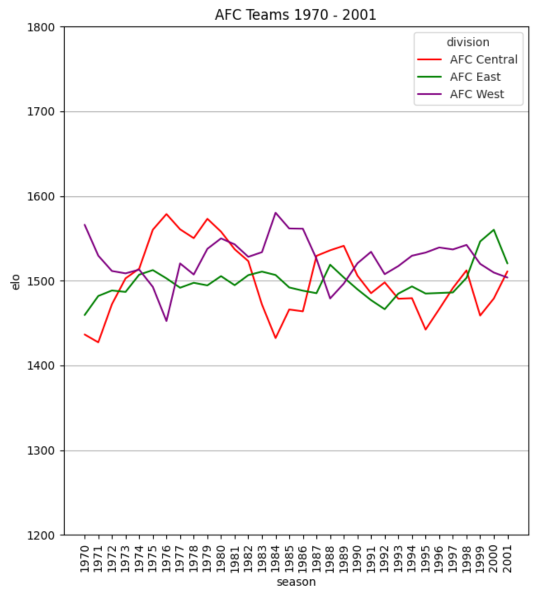
The below charts reflect AFC and NFC divisions for 1970 - 2001 and 2002 - 2022. The charts are split into two separate time periods due to the Central division in both conferences being split into North and South divisions with the Houston Texans joining the NFL. Interestingly, prior to the Texans joining the NFL there were 31 NFL teams. This led to uneven conferences and divisions as well as scheduling quirks such as one team getting an extra bye week due to these uneven team counts.

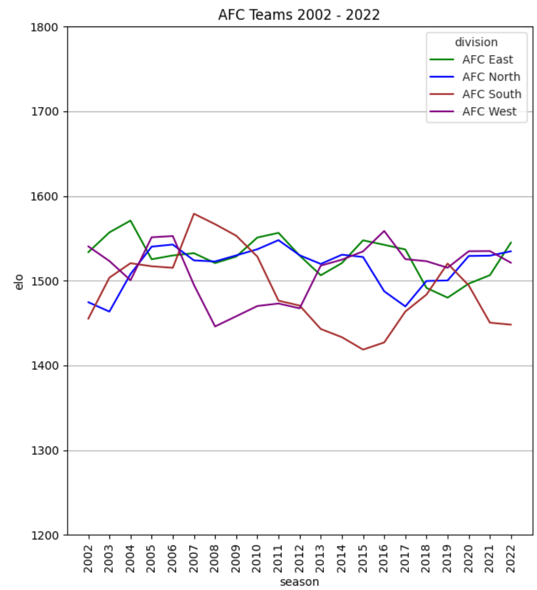
The average NFL team has an Elo rating of roughly 1500. Below recaps some highlights for each conference and division:

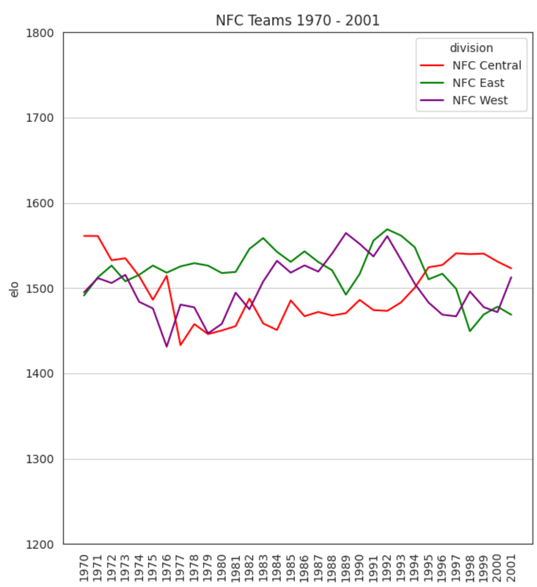
- AFC:
  - The AFC East had been a notably average team until the late 90s. This division included the Indianapolis Colts, New York Jets, Buffalo Bills, New England Patriots, and Miami Dolphins. The Colts were no longer part of the AFC East after 2002.
  - Prior to the AFC East rise, the AFC Central and West both swapped back and forth throughout the decades prior as the top divisions in the AFC. The AFC Central led largely in the mid-70s to early 80s and again in the late 80s while the AFC West led in the early to mid 80s and again throughout the most of the 90s.
  - After the split to four divisions, the AFC North and AFC East had relatively higher rated teams since the added division. The AFC South saw a brief moment in the late 2000s as higher rated but has since been rated much lower.
- NFC:
  - Throughout the 70s and 80s, the NFC East has been dominant leading the NFC divisions with higher Elo ratings per season. This division included the Philadelphia Eagles, New York Giants, Washington Redskins (now Commanders), and Arizona Cardinals. The Cardinals were no longer part of the NFC Easter after 2002.
  - The NFC Central was also a notably poorly rated division throughout most of the 70s, 80s, and 90s. This division did not see a rise in their ratings until the late 90s and up to

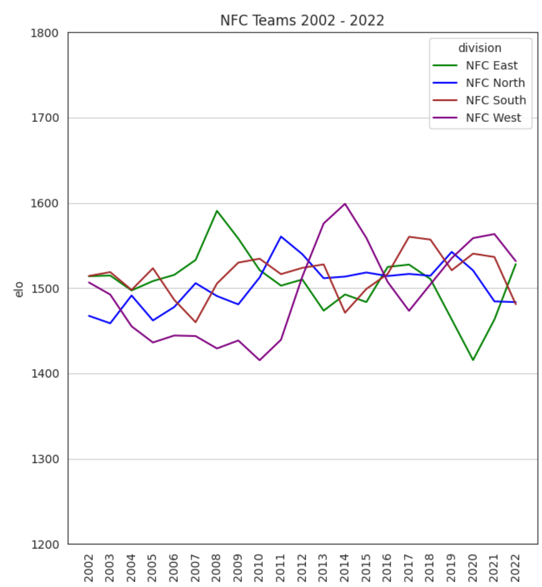
the division being split. Once split, the new North and South divisions have been largely average. There appears to be no dominant division over the past 20 years since the addition of the fourth division.

The graphs reflect aggregated average Elo ratings of NFL divisions since 1970. It is evident in the graphs that some divisions throughout the 70s, 80s, and some of the 90s have been able to maintain a higher Elo rating but it appears that since the fourth division was added to each conference in 2002, there has been no long run consistent high rated division.

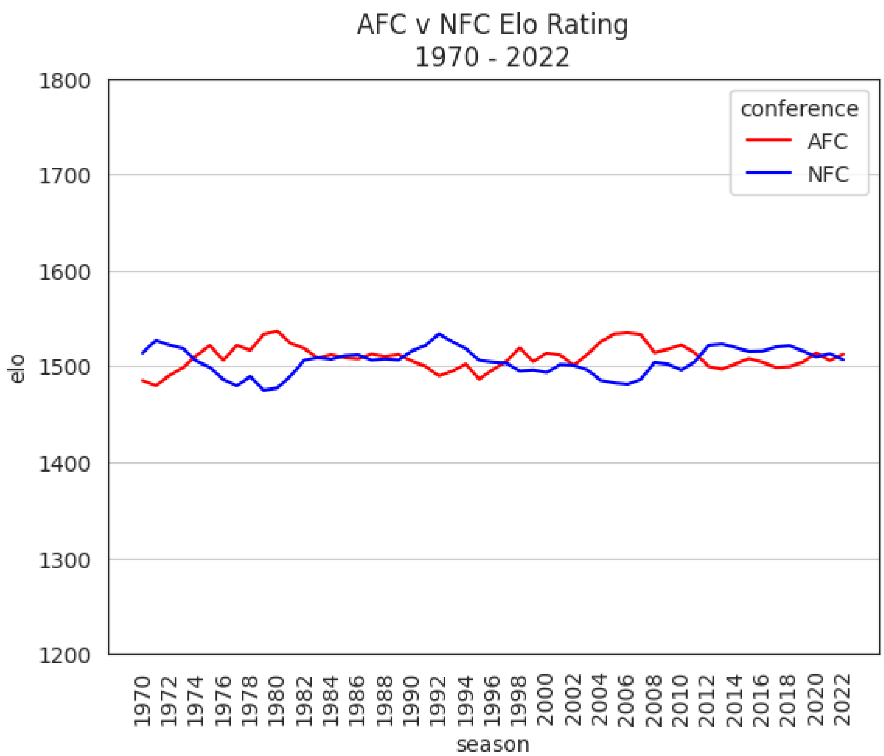


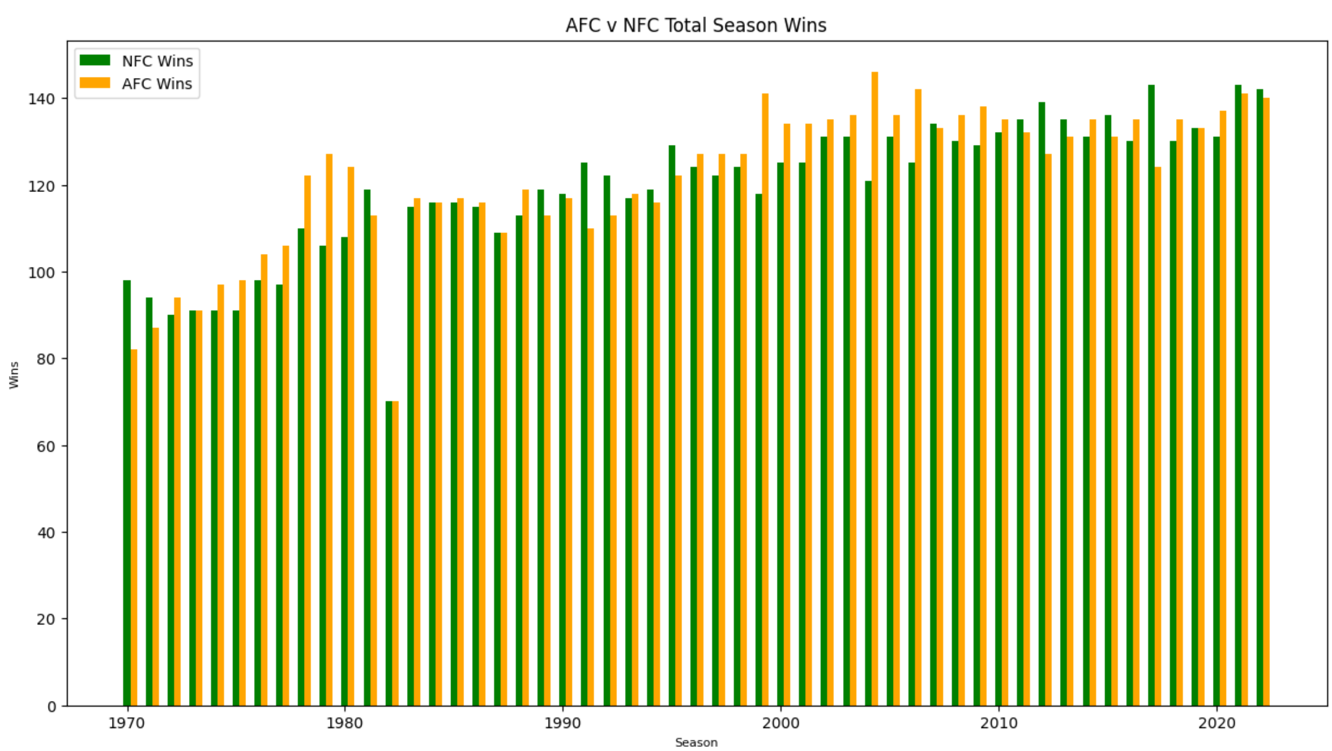






The charts below are an even higher aggregated average Elo rating for each conference in the NFL, AFC and NFC, and the total wins by conference for each season. This chart was used as a visual to see if one conference has been largely dominant over another. In the mid to late 70s, the AFC conference had higher rated teams versus the NFC conference which is also supported by the ~20 game win delta the AFC conference had over the NFC conference. Similar was true in the early 2000s with the AFC conference. The NFC conference was able to briefly have higher rated teams in the very early 70s and early 90s. Outside of these time periods, the two conferences have been rated relatively close.



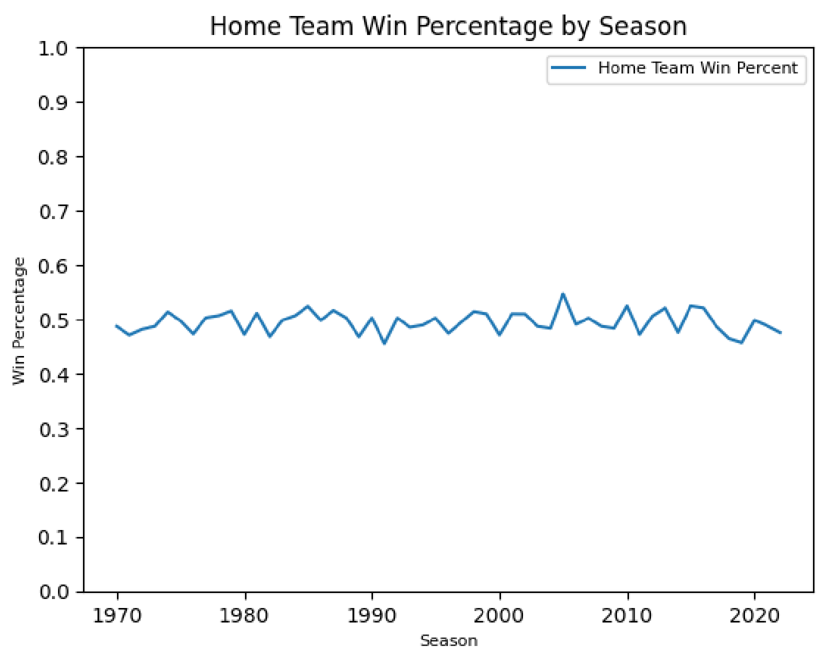


NFL Team Prediction on Winner a Game:

Two classifier algorithms were chosen to explore the predictive power of the Elo rating in relation to predicting game winners and teams to make the playoffs: Logistic Regression and Random Tree.

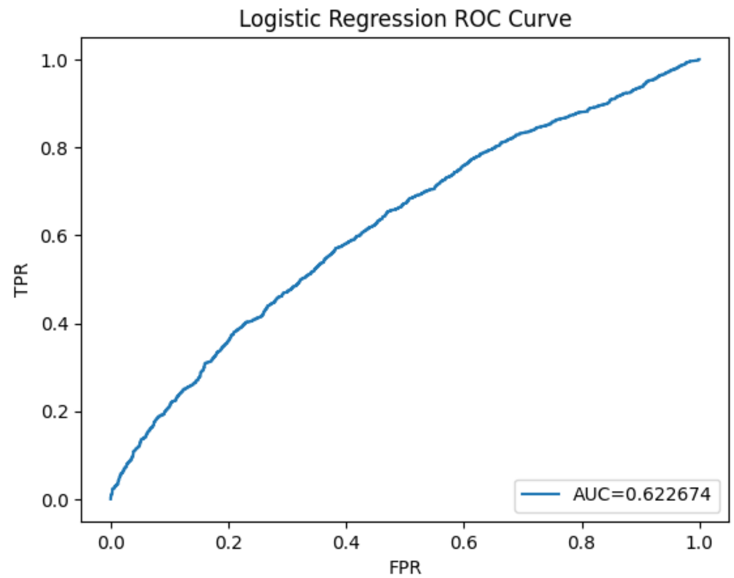
In preparation for creating a classifier for predicted game winner, a measure was created called “Elo spread.” This measure is simply the favored team Elo rating minus the underdog team Elo rating. The favored team is the team with the higher Elo rating. If there is a higher spread between two competing teams, the team with the higher rating is expected to win.

Although The model focused strictly on Elo spread. It is commonly believed across sports that there is a home field advantage. This translates to an increased likelihood of a team winning given they are the home team. The below does not support this indicating that from 1970 to 2022, the home team win percentage has been roughly 50%. Due to this, home/away status was not included in any of the models.



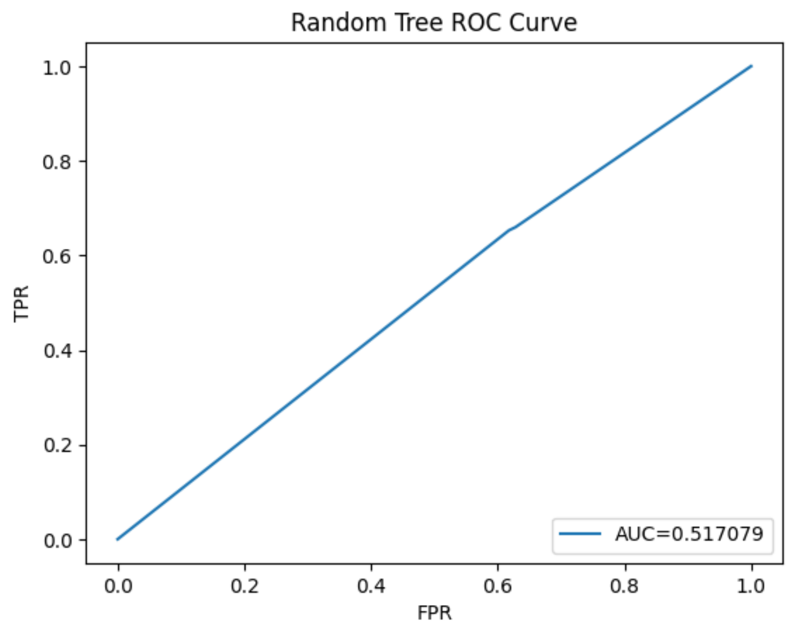
Note, in the output for each model TP = True Positive, FP = False Positive, FN = False Negative, and TN = True Negative

The LogisticRegression module was used from the scikit-learn library to create this classifier. There are 12,830 games being classified as predicted winner and predicted loser. The training set consisted of 70% of these games and the remaining 30% were the test set. Below are the outputs from the confusion matrix, key scores, and the respective ROC curve.




```
TP = 2419
FP = 1430
FN = 0
TN = 0
Accuracy: 0.6285
Precision: 0.6285
Recall: 1.0000
F1 Score: 0.7719
```

The DecisionTreeClassifier was also used from the scikit-learn library to create a Random Tree classifier to predict game winner and game loser. Similar to the Logistic Regression, 70% of the games were used in the training set and the remaining 30% in the test set. Below are the outputs from the confusion matrix, key scores, and the respective ROC curve.




```
TP = 1589
FP = 874
FN = 845
TN = 541
Accuracy: 0.5534
Precision: 0.6451
Recall: 0.6528
F1 Score: 0.6490
```

Below is a comparison and interpretation of the two models when predicting the game winner based off the Elo Spread:

- The Logistic Regression has a higher accuracy, recall, and F1 than the Random Tree
- The Random Tree has a slightly higher precision

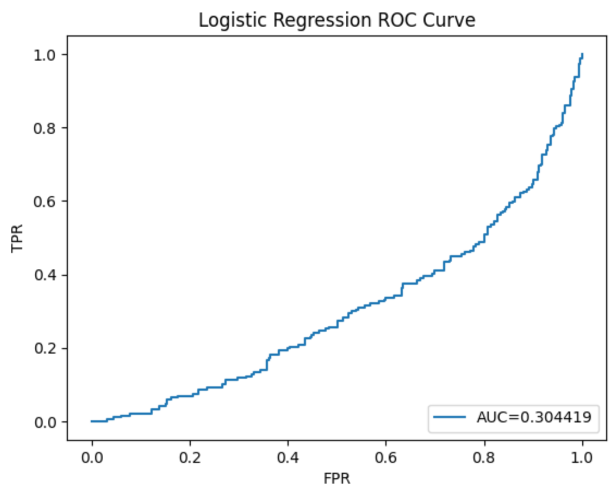
The accuracy of both models is a concern. While the Logistic Regression does have a higher overall accuracy (roughly 7 points higher) it is a concern that the performance against the test set returned all positives indicated by the 1.0 Recall score. This translates to the model predicting every team that has a higher Elo rating, regardless of how high, would win the game. The Random Tree accuracy is also ~55% which is slightly better than a coin toss.

NFL Team Prediction on Making the Playoffs:

Both classifiers were also used to predict whether a team will make the playoffs or not. The prediction is made based on an NFL team's Elo rating at the start of the regular season. This dataframe was narrowed down by finding the index of the first game of the season for each NFL team and returning what the pre-game Elo rating is. A binary column was also added to the dataframe indicating whether the team made the playoffs or not. There are 1,573 data points to classify (26 teams in 1970 to 32 teams as of 2022 totaled in between).

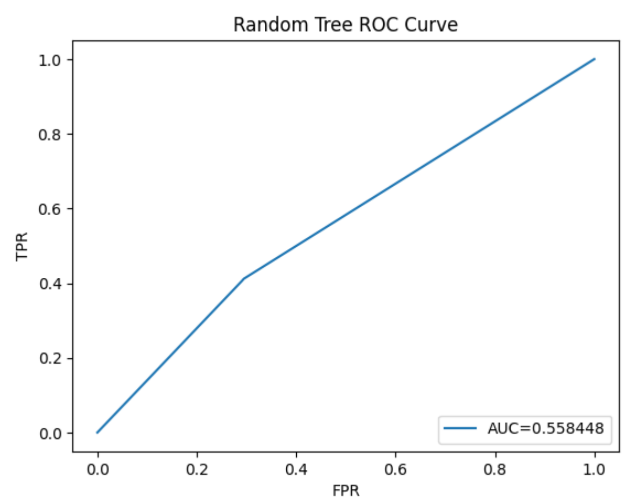
Similar to the game winner predictor, the data set was broken into 70% used for training and 30% used for testing. Below outlines the outputs for the Logistic Regression.

```
TP = 0
FP = 0
FN = 187
TN = 285
Accuracy: 0.6038
Precision: 0.0000
Recall: 0.0000
F1 Score: 0.0000
```



Below also outlines the outputs for the Random Tree classifier of NFL teams making the playoffs based on the start of season Elo rating.

```
TP = 76
FP = 83
FN = 111
TN = 202
Accuracy: 0.5890
Precision: 0.4780
Recall: 0.4064
F1 Score: 0.4393
```



The outputs for both classifiers are cause for concern:

- Logistic Regression
  - While the model did produce a 60% accuracy score, the model failed to produce any positive outcomes.
  - This in turn returned zeroes for precision, recall, and F1 score.
  - The AUC for the ROC curve is .3044 which is significantly lower than 0.5. This is indicating a worse than random performance.
- Random Tree classifier
  - The model produced a somewhat similar accuracy score of 59%.
  - The precision, recall, and F1 score are 47.8%, 40.6% and 43.9% respectively.
    - Precision is a concern indicating that less than 50% of what the Random Tree is classifying as positive is actually positive.
    - Recall is also concern with the model classifying 40% of teams making the playoffs correctly.

Given the scores and confusion matrix for both classifiers, using the start of season Elo rating as the only measure to predict whether a team will make the playoffs or not is not a good measure. These models do not appear to produce desired results.

Conclusions:

The following questions were proposed based on FiveThirtyEight's dataset and Elo rating model of NFL teams:

- How well does FiveThirtyEight's Elo rating predict winners?
- Can a previous season's Elo rating be used to predict next year's playoff contenders?
- Do certain NFL divisions historically have higher Elo ratings than others?
- Is there a relationship between an NFL team's Forbes list valuation and Elo rating?
- How are we expected to end up if we bet the money line on every game favorite?

While exploring the data, we were able to draw conclusions on all the questions.

FiveThirtyEight's Elo rating predicting winners and playoff contenders:

A Logistic Regression model and Random Forest model was used to predict game winner based on Elo ratings and playoff contenders based on start of season Elo ratings. After reviewing the scores of these models, including accuracy, precision, recall, and F1, as well as each respective ROC curve, using Elo rating strictly as a predictor does not produce sound outcomes. Below recaps each model's performance:

- Predicting Game Winner based on Elo spread:
  - Logistic Regression:
    - 63% Accuracy and Precision
    - 100% Recall
    - ***Predicted all games to be winners (no false negatives or true negatives)***
  - Random Tree:
    - 55% Accuracy
    - 65% Precision
    - 65% Recall
    - ROC Curve AUC = .517
      - ***ROC Curve indicating roughly 50/50 chance of correctly predicting winner***
- Predicting Playoff Contenders based on start of season Elo rating:
  - Logistic Regression:
    - 60% Accuracy
    - 0% Precision and Recall
    - ***Predicted zero game winners; only produced false negatives or true negatives***
    - ROC Curve AUC = .3044
      - ***ROC Curve indicating that the model is worse than random chance***
  - Random Tree:
    - 58.9% Accuracy
    - 47.8% Precision

        - 40.6% Recall
        - ROC Curve AUC = .558
        - ***Suboptimal scores and AUC***

While the Elo rating alone does not have strong predictive power, including other parameters not available in the dataset may enhance these models' performance. Pro-Football Reference has countless team statistics available including numerous defensive statistics. Exploring other available measures to enhance these models could increase the respective scores.

Additionally, the playoff contender focused strictly on the start of season Elo rating. The Elo rating is adjusted based on the outcomes of previous games. While a team may have performed well in the previous season, and thus received a higher Elo rating at the end of the previous season, roster moves, personnel changes, and coaching changes all could play an impact in a team’s ability to make the playoffs. Similar to what was noted, exploring other statistics available or making predictions as the season goes along may enhance the models’ performance.

Do certain NFL divisions historically have higher Elo ratings than others?

NFL team division and conference assignments were scraped from Pro-Football References website and consolidated by season to account for team relocations or new teams added to the NFL. This was combined with over 12,000 NFL games to aggregate Elo ratings by season for each division and conference in the NFL.

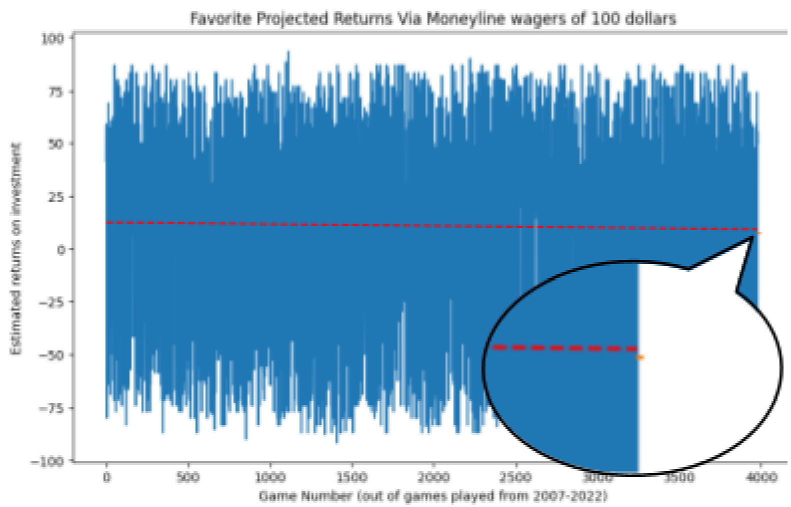
Reviewing charts of Elo ratings over time by AFC/NFC and their respective divisions, we were able to determine:

- There were two time periods where the AFC had noticeably higher rated teams than the NFC for several seasons (mid 70s, early 80s, and mid 2000s).
- NFC times still had a smaller amount of seasons where there were noticeably higher Elo ratings than AFC teams (early 70s & early 90s).
- There were time periods were a certain division in each conference had higher rated teams than other divisions in their respective conference:
    - AFC Central had much higher rated teams in the mid 70s then the AFC west dominated in the mid 80s
    - AFC South has more often than not been a bottom rated division in the AFC since it was established in 2002.
    - The NFC East was particularly dominate in Elo ratings from the 70s through the mid 80s and again in the late 2000s.
    - NFC West experienced brief instances where they were the highest rated division in the NFC and NFL. Notably, 2013 - 2015 they were significantly higher rated that the rest of the NFC divisions and the entire AFC.

How are we expected to end up if we bet the money line on every game favorite?

In short, consistent wagers placed on favored teams is not a viable way to make a profit when betting. In the event in which an individual were to place a $100 wager on all favored games between season 2007-08 - 2021-22, one would gross $43438.13, but one's net profit would be $-354361.87, due to the cost required to invest in the endeavor overall. This result is also apparent over nearly all other teams. Individuals would be at a loss by simply choosing one team to invest in.

```
returns from wagers  43438.132262437

 ====================================

10.919590815092409 lost/won per game

 ====================================

profit:  -354361.8677375624

 ====================================

count    3978.000000
mean       10.919591
std        49.700608
min       -91.743119
25%       -38.280116
50%        26.315789
75%        50.000000
max        93.457944
```



As a result of complications with competing plots, this diagram must remain in this view

Is there a relationship between an NFL team's Forbes list valuation and Elo rating?

There is certainly a relationship between Elo ratings and Forbes list valuation, however, it seems that the two are not codependent. Although team performance does have an impact on their annual valuation , there are auxiliary factors that evidently influence their financial status more than the performance metrics evaluated in this report.

Topics to be further explored may include developing a dataset normalized by the average income of each organization's respective state or division, in order to determine if there is a stronger correlation between the normalized valuations and the performance metrics described above.

Resources

1. *About Us. FiveThirtyEight.* https://fivethirtyeight.com/about-us/
2. *How Our NFL Predictions Work* (2023, January 9). *FiveThirtyEight*. https://fivethirtyeight.com/methodology/how-our-nfl-predictions-work/
3. *Elo Rating System. Chess.com.* https://www.chess.com/terms/elo-rating-chess
4. *nfl-elo. github.com.* https://github.com/fivethirtyeight/data/tree/master/nfl-elo
5. *Pro-Football Reference.* https://www.pro-football-reference.com/
6. *Forbes list of the most valuable NFL clubs. Wikipedia.com.* https://en.wikipedia.org/wiki/Forbes_list_of_the_most_valuable_NFL_clubs#cite_note-team-18
7. *NFL Scores and Odds Archive. SPORTSBOOK REVIEW ONLINE.* https://www.sportsbookreviewsonline.com/scoresoddsarchives/nfl/nfloddsarchives.htm
8. *NFLeloAnalytics. github.com.* https://raw.githubusercontent.com/bryan255/NFLeloAnalytics/main/nfl_division_info.csv