Predicting NFL Game Results with Elo Ratings

Pandas - June Lee, Victoria Midkiff, Alan Wang, Jessica Tang

Introduction and Data

Project Motivation

The history of the National Football League (NFL) can be traced back to 1920, when fourteen charter teams from thirteen "Original Towns" came together in September 1920 to form the American Professional Football Association ("NFL 100," n.d.). Over the NFL's century-old history, thousands of games and 52 Super Bowls have been played, with statistics recorded for each game. Today, many people enjoy betting on which football team will win the Super Bowl each year, using a variety of historical data and trends to compete with others. Elo ratings is a method that FiveThirtyEight uses to evaluate strengths of teams based on past performance (Reubenfb 2015). With football being so intertwined in our culture, we are particularly interested in looking at this data for games played in the year 1950 and later.

Research Question

How do differences between home and away team's various Elo values predict the probability of the home team winning an NFL game?

Overview of Data

The data were taken from FiveThirtyEight and has 17367 observations and 33 variables. The data were collected from the results of NFL games from 1920 to 2022. The observations are for a single NFL game and include information about both of the teams that are playing. It includes each team's Elo rating (measure of strength based on game-by-game results) before and after the game and the prior and post Elo rating for the quarterbacks playing in the game. It also includes the score of the game and the overall quality of the game, based on the harmonic mean of the Elo values of both teams before the game. FiveThirtyEight calculated all Elo values in the data set using data from Pro-Football-Reference.com.

The following key variables are used in this analysis. The data dictionary that contains descriptions of these variables can be found in the `README.md` file in the `data` folder of the project repository.

- `elo1_pre`: Home team's Elo rating before the game
- `elo2_pre`: Away team's Elo rating before the game
- `elo_diff`: Home team's Elo rating before the game minus the Away team's Elo rating before the game (`elo1_pre` - `elo2_pre`)
- `qb1_value_pre`: Home team starting quarterback's raw Elo value before the game
- `qb2_value_pre`: Away team starting quarterback's raw Elo value before the game
- `qb1_value_post`: Home team starting quarterback's raw Elo value after the game
- `qb2_value_post`: Away team starting quarterback's raw Elo value after the game
- `qb_diff_pre`: Home team starting quarterback's raw Elo value before the game minus Away team starting quarterback's raw Elo value before the game (`qb1_value_pre` - `qb2_value_pre`)
- `qb_game_diff`: Change in Home team starting quarterbacks' raw Elo less change in Away team starting quarterbacks' raw Elo. A measurement of by how much one quarterback outperformed the other in a given game (`qb1_value_post` - `qb1_value_pre`) - (`qb2_value_post` - `qb2_value_pre`).
- `score1`: Home team's score
- `score2`: Away team's score
- `playoff`: Whether game was in playoffs, and which playoff round if so
- `playoff_bin`: Whether or not the game was a playoff game (1: yes, 0: no)
- `result`: Whether or not the home team won (1: yes, 0: no)
- `neutral`: Whether or not the game took place at a neutral site (1: yes, 0: no)

Hypotheses

1. If an NFL game is a playoff game, the home team will have higher odds of winning the game.

2. Among the three Elo differences calculated (team Elo differences, quarterback Elo differences before the game, and differences in changes of the two quarterbacks raw Elo scores due to the game), differences in the team Elo values has a greater effect the odds of the home team winning compared to differences in the other Elo values.

Exploratory Data Analysis

Data Cleaning

These were the following steps we took to prepare our data for analysis:

1. Filter out the original dataset `nfl_baseline` by only including games for which Home team's Elo rating before the game (`elo1_pre`), Away team's Elo rating before the game (`elo2_pre`), Home starting quarterback's raw Elo value before the game (`qb1_value_pre`), Away starting quarterback's raw Elo value before the game (`qb2_value_pre`), Home starting quarterback's raw Elo value after game (`qb1_value_post`), and Away starting quarterback's raw Elo value after the game (`qb2_value_post`) are available, since these will be our main predictor variables. Since Elo metrics have been recorded starting with the 1950 season, dropping observations for NA values for any of these has the same effect as filtering for the games during and after the 1950 season. Additionally, the dataset contains games that are scheduled to take place in the 2022 season but have not taken place yet. As we do not know the winner of such games, we also filter these games out by dropping observations with NA `score1` values. We are also only interested in games that are played with a certain team hosting, so we are filtering the data to only have non-neutral games.

2. Calculate the `result` variable for each observations based on the score difference. If the home team score is greater than the away team's score, that means that the home team won, so `result = 1`. Otherwise, `result = 0`.

3. Note that the NFL's playoff format has changed over time. Mainly, the NFL adopted its current playoff format consisting of Wildcard, Divisional, Championship, and Super Bowl rounds starting in 1970 after its merger with the American Football League. Since the vast majority of the games in the dataset were regular season games, with not enough samples of games to consider the effects of each round of the playoffs, we create a new binary variable `playoff_bin` such that it equals 0 for regular season games and 1 for all types of playoff games. Thus, this allows us to overcome the fact that there are too few observations to consider each round of the playoffs as well as the fact that there was a change in the NFL's playoff format.

4. We recognize that there is an issue with independence between observations in this dataset since a given team plays multiple games during each season and the Elo ratings for each team rely on performance from its previous games. Thus, we attempt to make our observations (individual NFL games) more independent by taking a random sample of the dataset. By taking a random sample of games from 1970 to 2022, we decrease the probability that our dataset contains several games from the same season containing a same team. Additionally, teams regularly overhaul their roster of players over time, we can assume that two observations containing same teams are independent of one another given ample difference in the seasons that they were played in (i.e. the performance of the 1992 Dallas Cowboys will be independent of the performance of the 2022 Cowboys).

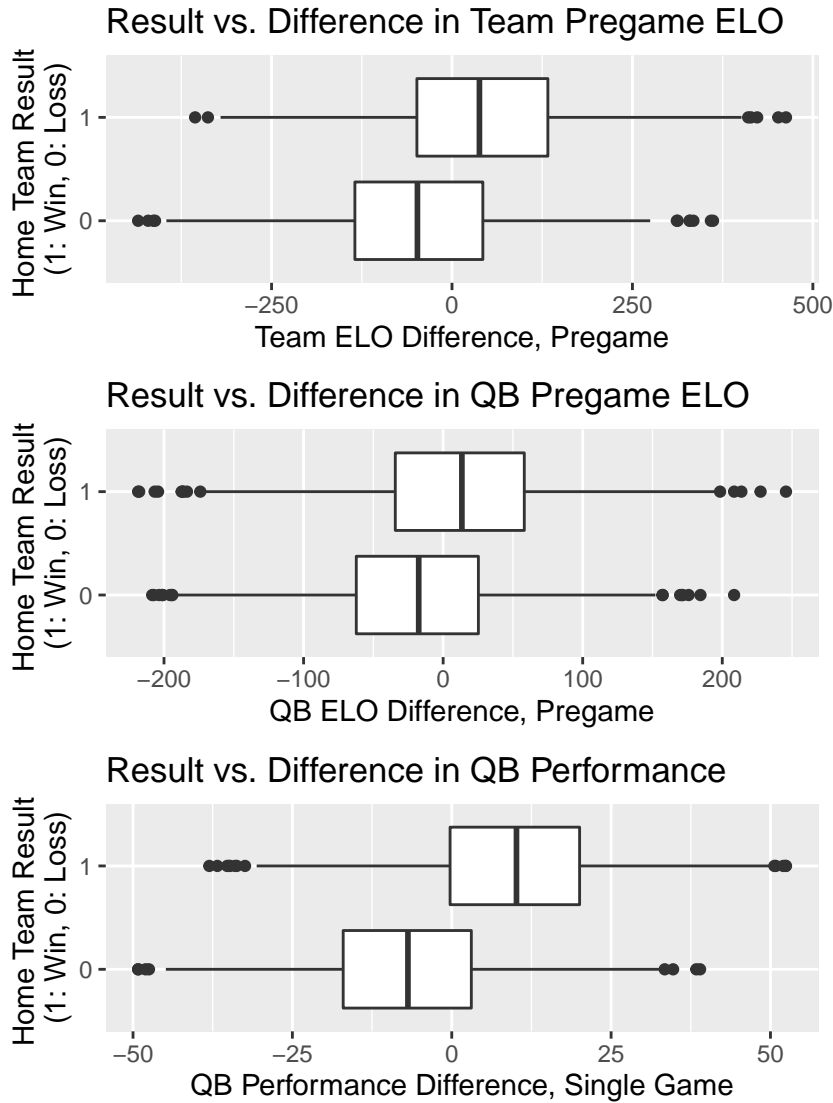**Home Team Wins/Losses Distribution**

Distribution of Home Team Wins/Losses
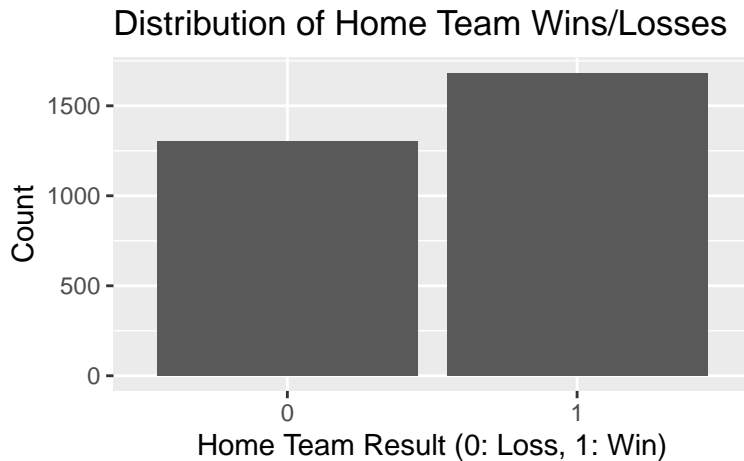

```
# A tibble: 2 x 2
  `result == 1`      n
  <lgl>          <int>
1 FALSE           1305
2 TRUE            1683


# A tibble: 2 x 2
  `playoff_bin == 1`      n
  <lgl>               <int>
1 FALSE                2871
2 TRUE                  117
```

Among the 2988 individual games that we have randomly sampled, 1683 resulted in a home team win for a 20% sample dataset's home team win probability of $\frac{1683}{2988} = 0.563$. This is also in line with the sample home team win probability of $\frac{8481}{14943} = 0.568$ for our 100% dataset. This shows that the 20% random sample is a good representation of the overall data set.

The fact that home teams tend to win slightly more than 50% of games makes logical sense. While every team plays the same number of games at home and at away in a given season regardless of the strength of the team (outside of playoff games, which only $\frac{117}{2871} = 4\%$ of the total number of games), yielding an expected probability of the home team winning of roughly 50%, there are certain benefits of playing at home such as having the support of its fans during the game as well as not having to travel that could account for the home team win probability being slightly greater than 50%.

Relationships between Response Variable and Predictors

Result vs. Difference in Team Pregame ELO

Result vs. Difference in QB Pregame ELO

Result vs. Difference in QB Performance

The Result vs Difference in Team Pregame ELO plot indicates that the median difference between the Home and Away team pregame ELO values among games in which the home team won is higher than the median difference among games in which the home team lost. Similarly, the Result vs Difference in QB Pregame ELO plot indicates that the median difference between the Home team quarterback and Away team quarterback pregame ELO values was higher among games in which the home team won than among games in which the home team lost. Finally, the Result vs Difference in QB Performance plot shows that the median difference in QB performance (Home QB performance - Away QB performance) among games resulting in a home team win was higher than that among among games result in a home team loss.

Relationships Between Predictors
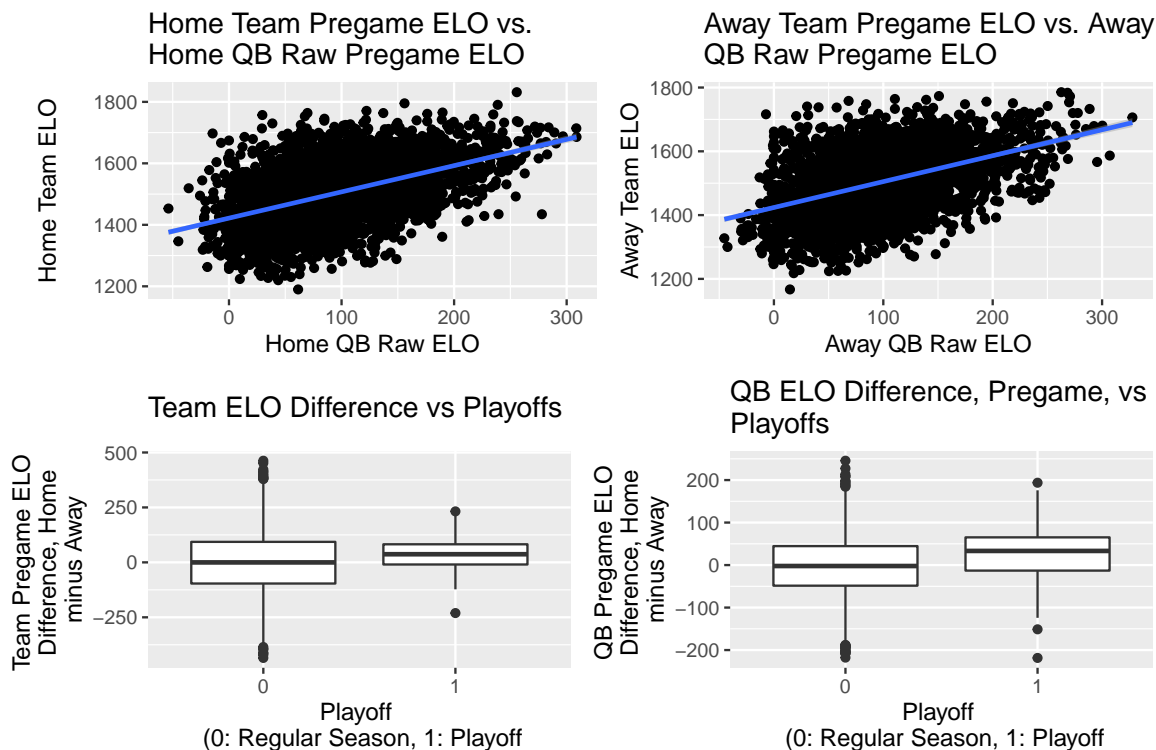
From the scatterplots, we can observe that the relationships between home team pregame ELO and home QB raw pregame ELO, as well as away team pregame ELO and away QB raw pregame ELO, appear to be moderate, positive linear relationships. For the two boxplots, there appears to be a greater spread in the Elo differences (home minus away) for non-playoff games and playoff games. We can also observe from the two boxplots that the median Elo differences (home minus away) for playoff games is higher than non-playoff games. This is most likely because teams that qualify for the playoffs are strong overall, unlike in regular season games, where a really good team could play a not so strong team.

Methodology

The response variable we are interested in predicting is the result of the game, either the home team wins or the away team wins. Our predictors include Home team's Elo rating before the game minus the Away team's Elo rating before the game (`elo1_pre` - `elo2_pre`), Home team starting quarterback's raw Elo value before the game minus Away team starting quarterback's raw Elo value before the game (`qb1_value_pre` - `qb2_value_pre`), Change in Home team starting quarterbacks' raw Elo less change in Away team starting quarterbacks' raw Elo (`qb_game_diff`), Whether or not the game was a playoff game (1: yes, 0: no), and

an interaction term between Whether or not the game was a playoff game (1: yes, 0: no) and Change in Home team starting quarterbacks' raw Elo less change in Away team starting quarterbacks' raw Elo (`qb_game_diff`). We will use logistic regression to assess how the Elo ratings for home and away teams will predict the probability that the home team will win.

Before we begin creating a model, we split the data into training data and testing data, using a proportion of 75%. We also set a model specification for a logistic regression model.

We create a recipe for the regression model by centering all numeric predictors, creating dummy variables for whether or not a game was a playoff game, creating an interaction term between between Whether or not the game was a playoff game (1: yes, 0: no) and Change in Home team starting quarterbacks' raw Elo less change in Away team starting quarterbacks' raw Elo (`qb_game_diff`), and removing all zero variance predictors. We then created a workflow for the model and fit the workflow to the training data.

With this new model, we wanted to make sure that it satisfied the conditions for logistic regression. First, we wanted to check the linearity condition by creating empirical logit plots for each of our predictors. Based on the empirical logit plots, it appears that there generally is a moderately linear relationship between the log odds and each of the predictors, so the linearity condition is satisfied.
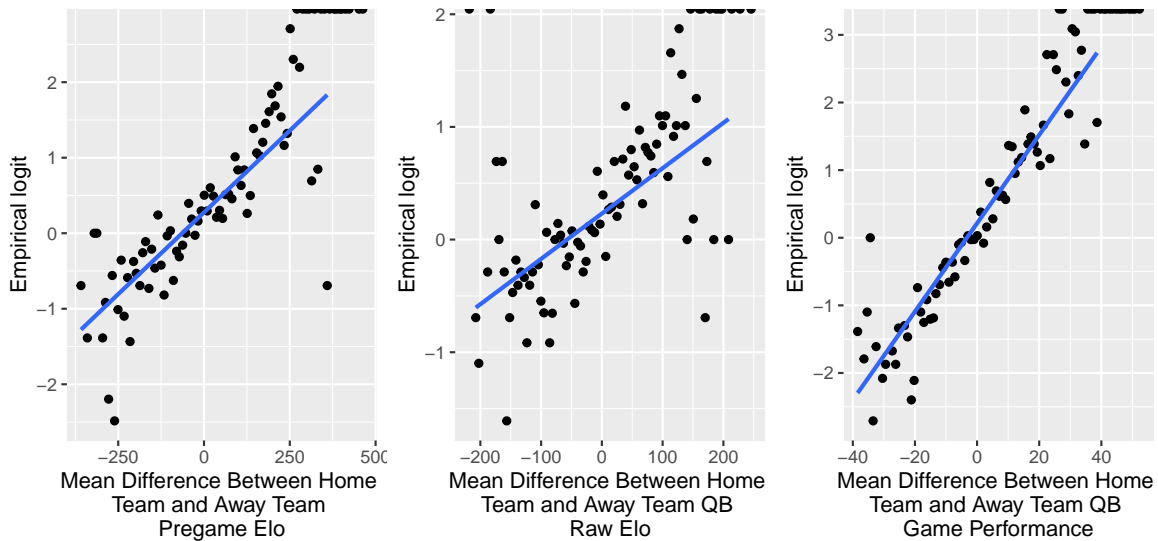
We collected a random sample of our dataset as outlined previously, thus it is reasonable for us to say that even if two observations contain the same teams, they are independent of one another given ample difference in the seasons that they were played in. Thus, we have also satisfied the randomness and independence conditions.

We lastly wanted to test our model to see if there are any issues with multicollinearity, but none of the VIF values for each of the predictors were greater than 10.

```
                        elo_diff                   qb_diff_pre
                        1.353819                      1.400738
                    qb_game_diff                 playoff_bin_X1
                        1.165331                      1.774254
    qb_game_diff_x_playoff_bin_X1
                        1.800119
```

Results

term	estimate	std.error	statistic	p.value
(Intercept)	0.33229	0.05466	6.07902	0.00000
elo_diff	0.00478	0.00048	9.96209	0.00000
qb_diff_pre	0.00537	0.00094	5.72965	0.00000
qb_game_diff	0.08720	0.00420	20.77460	0.00000
playoff_bin_X1	1.08872	0.40064	2.71743	0.00658
qb_game_diff_x_playoff_bin_X1	0.05336	0.03281	1.62632	0.10388

$$\log(\frac{\hat{\pi}}{1-\hat{\pi}}) = 0.33229 + 0.00478 \times elo\_diff + 0.00537 \times qb\_diff\_pre +$$

$0.008720 \times qb\_game\_diff + 1.08872 \times playoff\_bin\_X1 + 0.05336 \times qb\_game\_diff \times playoff\_bin\_X1$

For a game where the difference between the Home team's pregame Elo rating and the Away team's pregame Elo ratings is 0, the difference between the Home team starting quarterback's raw Elo value before the game and the Away team starting quarterback's raw Elo value before the game is 0, the difference between the change in the Home team starting quarterbacks' raw Elo and the change in the Away team starting quarterbacks' raw Elo due to the particular game is 0, the game is not a playoff game, we expect that the odds of the home team winning is 1.39416 (exp(0.33229)).

For each additional point increase in the difference between the Home team's pregame Elo rating and the Away team's pregame Elo ratings, the odds that the home team wins are expected to multiply by a factor of 1.00479 (exp(0.00478)), holding all other variables constant.

For each additional point increase in the difference between the Home team starting quarterback's raw Elo value before the game and the Away team starting quarterback's raw Elo value before the game (`qb1_value_pre` - `qb2_value_pre`), the odds that the home team wins are expected to multiply by a factor of 1.00538 (exp(0.00537)), holding all other variables constant.

For a non-playoff game, for each additional point increase in the difference between the change in the Home team starting quarterbacks' raw Elo and the change in the Away team starting quarterbacks' raw Elo due to the particular game, the odds that the home team wins are expected to multiply by a factor of 1.09112 (exp(0.08720)), holding all other variables constant.

Compared to a non-playoff game, we can expect that for a game that is a playoff game, the odds that the home team wins are expected to multiply by a factor of 2.9705 (exp(1.08872)) greater, holding all other variables constant.

For a playoff game, we can expect that for each additional point increase in the difference between the change in the Home team starting quarterbacks' raw Elo and the change in the Away team starting quarterbacks' raw Elo due to the particular game, the odds of a home team's margin of winning are expected to multiply by a factor of 1.15092 (exp(0.05336 + 0.08720)), holding all other variables constant. However, because the p-value of this coefficient in the model is greater than 0.05, this coefficient of the model is not statistically significant at the $\alpha = 0.05$ significance level.
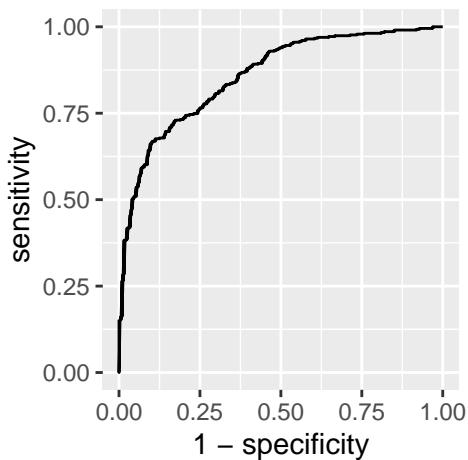

```
# A tibble: 1 x 3
  .metric .estimator .estimate
  <chr>   <chr>          <dbl>
1 roc_auc binary         0.858
```

The AUC is 0.858. Based on this, the model is a good fit for the training data. An AUC of 0.858 means that given a particular game, there is a 85.8% chance that the model correctly predicts if a game is won by the home team or lost by the home team.

Discussion and Conclusion

The results from the logistic regression model show that our first hypothesis is correct. Based on the model, we can observe that compared to a non-playoff game, we can expect that for a game that is a playoff game, the odds that the home team wins are expected to multiply by a factor of 2.9705 (exp(1.08872)) greater, holding all other variables constant. This shows that if the game is a playoff game, compared to a regular season game, there is a significant difference in the effect on the odds the home team wins.

However, the model shows that our second hypothesis is incorrect. Based on the model, we can observe that the difference between the change in the Home team starting quarterbacks' raw Elo and the change in the Away team starting quarterbacks' raw Elo due to the particular game was the predictor variable that had the greatest effect on the odds that the home team wins, out of all of the Elo difference predictor variables. Even if a team heads into a game as an underdog in terms of team pregame Elo ratings and QB pregame Elo ratings, an outperformance by the team's quarterback could make up for the difference in the Elo ratings.

One of the main issues we encountered with the dataset was that it violated the independence conditions necessary for logistic regression modeling, because the Elo ratings for each team for each observation in the dataset relied on previous games, which are also other observations in the dataset. Because the data were also not randomly selected, we attempted to address both the randomness and independence issues by taking a random sample of the dataset. With this random sampling, we assume that two observations containing same teams are independent of one another given ample difference in the seasons that they were played in. In addition, another limitation we encountered with the dataset was that there were a variety of iterations of the QB Elo ratings for each team, so we could only use a select few in our regression model.

Some ideas for future work include using residual plots to examine which teams over-perform and which teams under-perform and examining which factors lead to under or over performance (winning more or less games than expected), which could be an alternative method to addressing issues with independence for this dataset. Additionally, the components of the NFL teams could be examined separately. For example, separate Elo ratings of the offense and defense could be analyzed as predictors for winning an NFL game. This could provide more insight into which aspects of a particular team are contributing to their wins and losses.

References

"NFL 100." n.d. NFL. https://www.nfl.com/100/original-towns/.

Reubenfb. 2015. "The Complete History of the NFL." https://projects.fivethirtyeight.com/complete-history-of-the-nfl/#lar.