

# **Exploratory Data Analysis of FiveThirtyEight Baseball Data**

By

Zach Watkins

11/20/20

MSDS 430: Python for Data Science

## **Introduction**

For this exploratory data analysis, I focused on baseball game outcome data. The data came from the website FiveThirtyEight.com and had every single baseball game dating back to the 1800's. I have always loved baseball as a sport so this analysis of how accurately the two prediction methods in the data actually predicted the outcome was very interesting to me.

Overall the goal of the project was to see whether or not the FiveThirtyEight rating method, which takes into account home/away team as well as starting pitcher, was more accurate than an elo (originally developed to rank chess players) rating prediction. My hypothesis was that because there is more rigor in the FiveThirtyEight rating methodology that it would more accurately predict the outcomes of the games.

## Data Preparation and Analysis

The initial data included many fields such as pitchers that were starting that day which were not needed for this analysis. In order to make the data more manageable to analyze these fields were removed and we were left with the following fields:

Field Name	Description	Sample Data
Season	Season the game was played	2015
Team 1	Name of the home team	CHC
Team 2	Name of the away team	STL
Elo1_pre	Elo rating of the home team	1500
Elo2_pre	Elo rating of the away team	1500
Elo1_prob	Elo probability of home team winning	0.50
Elo2_prob	Elo probability of away team winning	0.50
Rating1_pre	FiveThirtyEight home rating	1500
Rating2_pre	FiveThirtyEight away rating	1500
Rating_prob1	FiveThirtyEight home probability of winning	0.50
Rating_prob2	FiveThirtyEight away probability of winning	0.50
Score1	Final score of the home team	3
Score2	Final score of the away team	5

Table 1 Data Dictionary for Baseball Game Variables

After the data was slimmed down to the fields in Table 1, there were some games that were removed in order to have more consistent data to analyze. Playoff games were removed in order to focus on predictions throughout all regular seasons and because different seasons had different number of playoff games. Neutral games were also removed since the FiveThirtyEight rating system relies on home/away teams.

Now that the data was in a good starting point, the analysis began. The first step was to create an additional field for the decade of each season to provide a better view overtime of how the predictions performed. I also based whether a prediction was correct

based on that rating system's forecasted probability of a team winning. A team only had to have a probability higher than 0.50 to be considered the team predicted to win.

Initially I found that the overall Elo percentage correct was 57.49% whereas the overall FiveThirtyEight rating correct was 57.55%, a 6-basis point difference which was not incredibly significant. Next, I decided to look at performance by decade to see if there were any decades where one system outperformed the other.

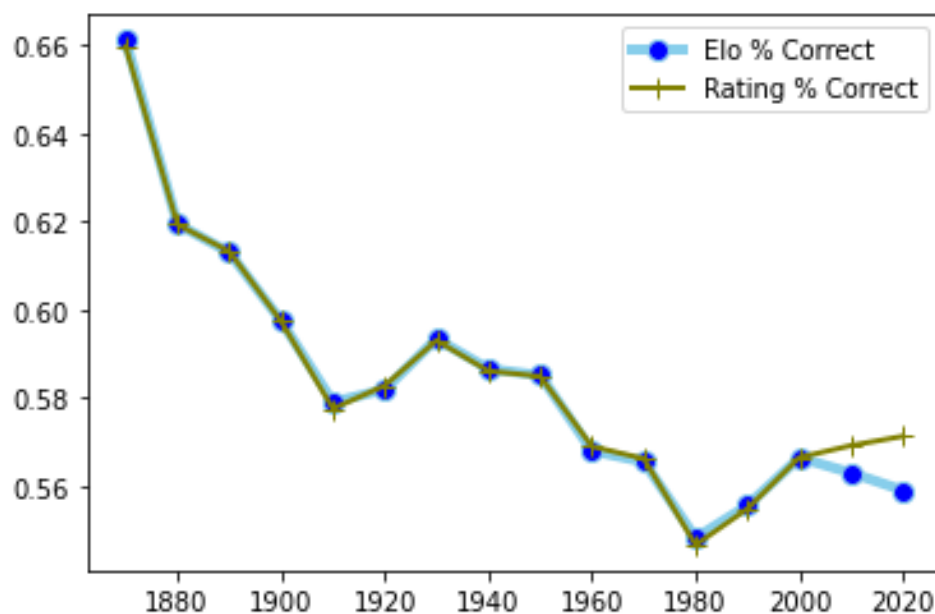


Chart 1 Prediction % correct by decade

As you can see in Chart 1, there was virtually no difference in the prediction ratings from the 1870's all the way through the 2000's. I started to see that beginning in the 2010 decade the FiveThirtyEight rating prediction started to outperform the elo rating by a considerable margin.

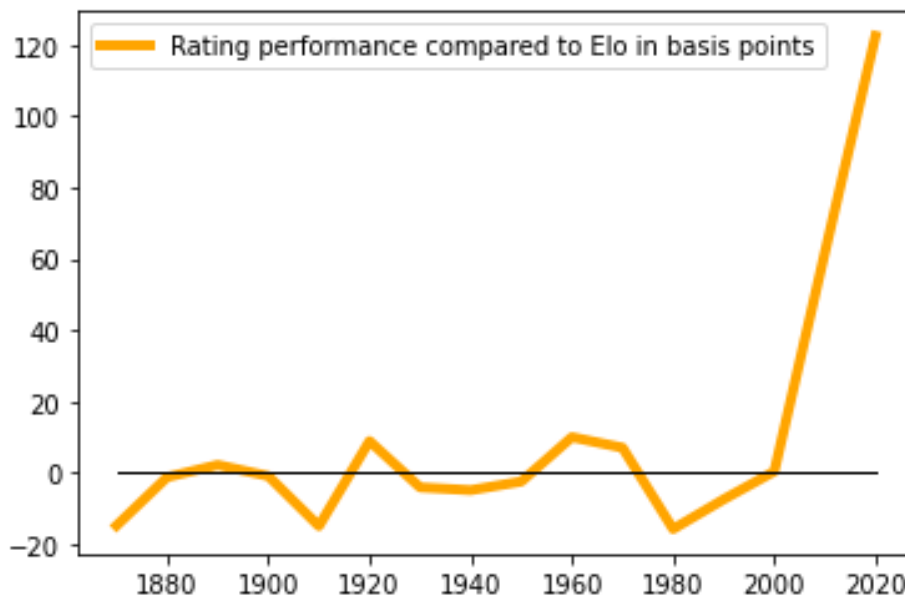


Chart 2 Basis point difference between prediction methods

As you can see in Chart 2, the basis rating difference is hovering right around 0 but in more recent decades include this past season (2020), the FiveThirtyEight rating method outperformed the elo rating method by 120 basis points, a pretty significant margin. In fact, when breaking down the basis point difference by season since 2000 we see there are seasons where the FiveThirtyEight rating outperformed elo by up to 200 basis points.

The final thing I wanted to look at was the performance of the two rating methods in games where they predicted different winners. In total, there were 7479 games in which the elo prediction and FiveThirtyEight prediction were different. In the games where this was the case, Elo was correct 49.15% compared to FiveThirtyEight being correct 50.76%, a 160-basis point difference. Breaking it down by decade there is not a telling story of which is more accurate. It was also hard to equally compare decades against each other

since the amount of games in which the outcomes differed significantly skewed towards more recent decades (there were 3000+ such instances in the 2010's of the 7479 total). Breaking down the difference by season since 2000, we see that starting somewhere around the 2012 season FiveThirtyEight has consistently significantly outperformed elo on games in which the outcomes were different, as seen in Chart 3.



Chart 2 Basis point difference between prediction methods

## Conclusion

Overall, I would say that the FiveThirtyEight rating methodology was better at accurately predicting the outcomes of the game, especially in more recent seasons since 2012. This was consistent with my hypothesis due to the fact that the FiveThirtyEight

methodology takes into account more variables and a variable such as starting pitcher can play an important role in the outcome of a baseball game.

Possible next steps on the analysis of this data would be looking at how accurate the predictions were based on different deciles of probability a team would win. For example, we would expect that if a rating system gave a team a 70% - 80% chance to win these predictions would be more accurate than all the games where the teams had a 50% - 60% chance to win. Looking at these deciles could further breakdown the performance between the two rating system and lead to more detailed analysis of which is better and how much is it better.