Introduction

Predicting the outcome of sports seasons has always been a topic of immense interest among statisticians, sports enthusiasts, and analysts. The National Football League (NFL) is no exception, its dynamic and highly competitive environment provides fertile ground for predictive modeling. The goal of this project is simple: Use data from the 2023 NFL season, to predict how many games each team will win in the 2024 season.

The backbone of our predictive model is the data from the 2023 NFL season, which includes the number of games each team won and their respective performance metrics, as evaluated by Pro Football Focus (PFF). PFF scores offer a comprehensive assessment of a team's offense, defense, and overall performance, and in my opinion, is the most quantified metric to assess the performance of any specific player, and any specific team.

By analyzing the relationship between these 2023 metrics and the number of games won, I developed a linear regression model to predict the outcomes for the 2024 season. My approach involved:

1. **Data Collection and Preparation:** We gathered and cleaned data for the 2023 season, ensuring all relevant metrics were numeric and ready for analysis.

2. **Model Development**: Using linear regression, we established a predictive model based on the 2023 season wins and performance metrics.

3. **Prediction and Visualization:** The model was then applied to the 2024 PFF metrics to forecast each team's win totals. The results were visualized to highlight the predicted standings and performance trends.

This writeup details the steps taken to build the model, the insights gained from the analysis, and the predicted outcomes for the 2024 NFL season. Our findings provide a data-driven perspective on how teams are expected to perform, offering valuable insights for fans, analysts, and stakeholders in the NFL community.

The Data:

The data used in this model was meticulously gathered from the Pro Football Focus (PFF) website. This dataset includes various columns, each integral to calculating my proprietary metric for predicting team performance. The data collection process was particularly challenging. I had to subscribe to PFF Pro, manually record scores for nearly every starting player in the NFL, and compile overall scores for offense, defense, and other metrics.

Web scraping techniques were not viable due to the dynamic nature of team rosters. To predict the 2024 season, it was crucial to account for roster changes that PFF had not yet updated from the 2023 season. This included tracking every rookie drafted, as well as players who were traded or signed during free agency. Consequently, I manually maintained an updated roster for each team, ensuring the accuracy of each player's PFF score. This labor-intensive process was essential for the reliability and precision of my predictive model.

Wanya Morris 55.6
Joe Thuney 74.9
Creed Humphrey 81.4
Trey Smith 74.6
Jawaan Taylor 49.8

Bernhard Raimann 82.7
Quenton Nelson 70.8
Ryan Kelly 77.2
Will Fries 61.2
Braden Smith 83.3

Brandon Coleman 60
Nick Allegretti 65.8
Tyler Biadasz 69.2
Sam Cosmi 80.6
Andrew Wylie 69.2

Tyler Guyton 67
Tyler Smith 74.4
Brock Hoffman 53.7
Zack Martin 68
Terence Steele 52.3

Terron Armstead 79.3
Isaiah Wynn 47.8
Aaron Brewer 76.6
Robert Jones 57.6
Austin Jackson 66.4

Jordan Mailata 84.8
Landon Dickerson 69.4
Cam Jurgens 61.4
Tyler Steen 53.2
Lane Johnson 80.1

Jake Matthews 71.2
Matthew Bergeron 59.1
Drew Dalman 82.3
Chris Lindstrom 87.6
Kaleb McGary 75.5

Andrew Thomas 74.1
Jermaine Eluemunor 68.7
John Michael Schmitz 41.4
Jon Runyan 56.5
Evan Neal 39.8

Cam Robinson 65.4
Ezra Cleveland 59.5
Mitch Morse 63.7
Brandon Scherff 67.3
Anton Harrison 53

Tyron Smith 83.7
John Simpson 56.3
Joe Tippmann 61
ALT 71.7
Morgan Moses 80.4

Taylor Decker 81.1
Graham Glasgow 74.9
Frank Ragnow 88.8
Kevin Zeitler 71.6
Penei Sewell 92.8

Rasheed Walker 66.3
Elgton Jenkins 63.8
Josh Myers 54.7
Sean Rhyan 49.3
Zach Tom 79.7

Ikem Ekwonu 67.4
Damien Lewis 59.6
Austin Corbett 47.9
Robert Hunt 76.4
Taylor Moton 74.6

Chukwuma Okorafor 60.4
Cole Strange 64.6
Jake Andrews 53
Sidy Sow 64.4
Mike Onwenu 71.5

Kolton Miller 80.2
Dylan Parham 60.4
Andre James 74.6
JPJ 68.
Thayer Munford 74.1

Alaric Jackson 66.6
Jonah Jackson 61
Steve Avila 60.5
Kevin Dotson 84.4
Rob Havenstein 79.8

Ronnie Stanley 68.1
Andrew Vorhees 55
Tyler Linderbaum 78.5
Ben Cleveland 65.8
Roger Rosengarten 60

Ryan Ramczyk 73.5
Taliese Fuaga 69
Nick Saldiveri 52.1
Erik McCoy 79.4
Cesar Ruiz 51.2

Team	QB	RB	WR	DL
49ers	75.1	82.1	82	68.2
Bears	62.6	77.1	63.7	69.5
Bengals	91	80.9	73.6	64.6
Bills	87.2	76.8	71.3	60.6
Broncos	64.5	64.3	69.2	67.9
Browns	51	90.7	68.7	73.3
Bucs	82.2	66.5	69	61.8
Cardinals	65.1	76.1	59.8	67.9
Chargers	76.7	82.6	76.5	70
Chiefs	90.1	74.2	75	76.9
Colts	70	71.5	60.4	68
Commanders	67	81	70	61.7
Cowboys	68.2	90	71.4	68.3
Dolphins	84.4	74.4	68.9	65.3
Eagles	83.8	83.2	77	77.8
Falcons	67	74	73	78.2
Giants	72.9	84.9	64.2	64.7
Jaguars	74.1	82.9	65.5	66.4
Jets	75.9	76.4	68.1	67.6
Lions	71.6	69.4	74.1	71.7
Packers	72.9	90.1	67.9	68.7
Panthers	75	81.8	69.6	63.2
Patriots	68.7	81.5	68.8	66
Raiders	69.5	91.9	73.2	69.5
Rams	67.6	67	68.9	65.5
Ravens	76.4	80	70.2	74.1
Saints	65.4	77.7	74.2	61.2
Seahawks	75.4	84.8	70.6	65.2
Steelers	73	75.2	69.8	68.4
Texans	73	84.8	65	60.2
Titans	71.9	86.7	68.6	58.6
Vikings	77.3	84.2	67.7	71.7

Previous Season's Wins: Although the model will only be predicting on the amount of games a team will win in the regular season, I am inputting even the playoff wins, because it's a good indicator of how the team will perform for the future season. For example, the *Green Bay Packers,* the youngest team in the postseason, only won 9 games during the regular season, just finishing with a winning record, but they defeated the Dallas Cowboys in the first round of the playoffs, which was the largest playoff upset last postseason. I believe that win is imperative for the model to know, it is a strong indicator of what is to come of a better, more experienced Packers team.

2023 PFF Offense: To determine the offensive PFF score of each team, I took a weighted average of the four main offensive position groups. QB (40%), RB (15%) WR (20%) and OL (25%).

To find the PFF score for the **quarterback** position, I just found the starters PFF passing grade, but if I determined that he's a "mobile" quarterback, I took a weighted average of his passing grade (80%) and rushing grade (20%).

To find the PFF score for the **running back** position, I found the starters PFF rushing grade, but if I determined that the team used a committee at the running back position, then I used a weighted average of the RB1 (75%) and RB2 (25%)

To find the PFF score for the **wide receiver** position, I took a straight average of the three starting receiver's PFF receiving grade, the starting running back's receiving grade, and the tight end's receiving grade

To find the PFF score for the **offensive line** position, I took a straight average of the five starting offensive lineman's overall blocking grade.

Once I found all of the individual position grades, I computed the weighted average mentioned above, and each team was given an offensive grade.

2023 PFF Defense: To determine the defensive PFF score of each team, I took a weighted average of the four main defensive position groups. Edge Rusher (25%), IL (15%), LB (24%), CB (20%), S (16%).

To find the PFF score for the **Edge Rusher** position, I took a straight average of the two starting edge rushers.

To find the PFF score for the **Interior Lineman** position, I took a straight average of the two starting interior lineman.

To find the PFF score of the **linebacker** position, I took a straight average of the three starting linebackers.

To find the PFF score for the **cornerback** position, I took a straight average of the two starting cornerbacks, and the nickel cornerback.

To find the PFF score for the **safety** position, I took a straight average of the two starting safeties.

Once I found all of the individual position grades, I computed the weighted average mentioned above, and each team was given a defensive grade.

2023 PFF Average: I took a straight average of the scores given for 2023 PFF Offense and 2023 PFF Defense for every team.

2023 Opponent PFF Average: Now that I have the average PFF score for every team, I can find the average PFF score of everyone's opponents. For example, let's say the Pittsburgh Steelers played three games this season. They play the Jets, Giants, and Panthers. I would take an average of the 2023 PFF score of those three teams. Whatever that number turns out to be, will be the 2023 Opponent PFF Average. So for all 32 teams, I am finding the average PFF score of all 17 of their opponents, and averaging all of those scores.

2023 Metric: The metric is simple, yet it is the most important part of this project. This is what I am predicting the 2024 season outcome on. The metric is just the average PFF score of any team, minus the average PFF score of all of their opponents. The formula is **2023 PFF Average - 2023 Opponent PFF Average.** This will give me an output for every team that I will use to predict the amount of games every team will win in 2024.

2024: The process for the 2024 columns are the same as 2023.

Now that we know how the data works, let's talk about the model.

<u>The Model:</u>

A linear regression model is being used. We are predicting on the 2024 metric, and we are predicting the amount of games a team will win in 2024. We look at the 2023 metric, and how many games each team won in 2023, then we will use those same data points to predict the 2024 results. The model was coded in R, and I hope to code it in Python in the near future.

<u>Areas of Improvement:</u>

The biggest issue I see with this project is the lack of data. If I am trying to predict how many games a team will win, I should be looking at much more data than just one season's worth. The issue is, gathering one season's worth of data takes a tremendously long time, especially when I am working a 9 to 5 everyday. Also, the NFL recently switched from a 16 game season to a 17 game season, meaning the data will be skewed once we reach the 16 game seasons. I plan on improving this for future seasons by making the 2024 data into training data for the 2025 season, and then making the 2025 data into training data for the 2026 season, etc etc.

Determining the Credibility of the Model:

To determine whether the results from my model are valid, I am comparing them to the DraftKings team wins over/under line. If my model is within half a game of the DraftKings line I will give myself 3 points, if my model is within 1.5 games of the DraftKings line, I will give myself 1 point.

Team	Model	DraftKings	Points Received
49ers	12	11.5	3
Bears	9	8.5	3
Bengals	8	10.5	0
Bills	9	10.5	1
Broncos	6	5.5	3
Browns	9	8.5	3
Buccaneers	7	7.5	3
Cardinals	6	6.5	3
Chargers	8	8.5	3
Chiefs	10	11.5	1
Colts	7	8.5	1
Commanders	7	6.5	3
Cowboys	12	10.5	1
Dolphins	11	9.5	1
Eagles	9	10.5	1
Falcons	10	9.5	3
Giants	6	6.5	3
Jaguars	9	8.5	3
Jets	11	9.5	1
Lions	11	10.5	3
Packers	8	9.5	1
Panthers	7	5.5	1
Patriots	8	4.5	0
Raiders	8	6.5	1

Rams	9	8.5	3
Ravens	12	10.5	1
Saints	10	7.5	0
Seahawks	10	7.5	0
Steelers	9	8.5	3
Texans	8	9.5	1
Titans	7	6.5	3
Vikings	8	6.5	1

75% of my model's predictions were within at least 1.5 games of Draftkings' predictions, scoring at least 1 point. 47% of my model's predictions were within half a game of DraftKings' predictions, scoring 3 points. Out of the 96 points I could have scored, I scored 58, achieving 60% accuracy.

This is much better than I expected. The folks at draftKings have access to billions of more data points than I do. They are constantly working on their predictive models, and they are all far more experienced than me, it is literally their job to do this. I am a chump compared to these guys. I had access to one data set (which I created myself), I used one of the most basic types of models, and I actually paid to work on this, only after I got home from my 9 to 5 as an actuary. My model being 60% as "correct" as the DraftKings model is a testament to the creativity and sharpness of the model.

<u>Acknowledgements:</u>

I would like to thank my good friend Jonah Bierig for his help gathering the data and PFF scores necessary for the completion of the project. I would not have been able to finish it without his help. [Jonah's LinkedIn]

<u>Code:</u>

````
---
title: "NFL Win Predictor"
author: "Joshua Eisner"
date: "07/16/2024"
---

```{r}
# Load necessary libraries
library(readr)
library(ggplot2)

# Load the dataset
````
````

```
nfl_data = read.csv("~/NFL Data CSV.csv")

# Display column names to check for exact matches
print(names(nfl_data))

# Rename columns to ensure there are no issues with spaces or special characters
names(nfl_data) = make.names(names(nfl_data))

# View the renamed columns
print(names(nfl_data))

# Convert relevant columns to numeric
nfl_data$X2023.Season.Wins = as.numeric(nfl_data$X2023.Season.Wins)
nfl_data$X2023.Metric = as.numeric(nfl_data$X2023.Metric)
nfl_data$X2024.Metric = as.numeric(nfl_data$X2024.Metric)

# Create the linear model using 2023 data
model_2023 = lm(X2023.Season.Wins ~ X2023.Metric, data = nfl_data)

# Summary of the model
summary(model_2023)

# Predicting the wins for the 2024 season using 2024 metric
nfl_data$Predicted_Wins_2024 = predict(model_2023, newdata = data.frame(X2023.Metric =
nfl_data$X2024.Metric))

# Round the predicted wins to 2 decimal places
nfl_data$Predicted_Wins_2024 = round(nfl_data$Predicted_Wins_2024, 2)

# View the results
print(nfl_data)

# Plot the data and the regression line using 2023 data
ggplot(nfl_data, aes(x = X2023.Metric, y = X2023.Season.Wins)) +
  geom_point(color = "blue", label = "2023 Actual Wins") +  # Scatter plot for 2023
  geom_smooth(method = "lm", col = "blue") +  # Regression line for 2023
  geom_point(aes(x = X2024.Metric, y = Predicted_Wins_2024), color = "red", label = "2024
Predicted Wins") +  # Scatter plot for 2024 predictions
  #geom_text(aes(label = Team, x = X2024.Metric, y = Predicted_Wins_2024), vjust = 1, color =
"black", size = 3) +  # Labels for 2024 predictions
  labs(title = "2023 Season Wins vs 2023 Metric with 2024 Predictions",
       x = "Metric",
       y = "Wins") +
  theme_minimal() +
```

```
  scale_color_manual(values = c("blue", "red")) +
  theme(legend.position = "top")

# Output the predicted wins for each team in 2024 to the console, sorted by
Predicted_Wins_2024 in descending order
predicted_wins = nfl_data[, c("Team", "Predicted_Wins_2024")]
predicted_wins = predicted_wins[order(-predicted_wins$Predicted_Wins_2024), ]
print(predicted_wins)
```