

ANALYZING NFL INJURIES & IMPACT

By: Tim Pao, Ken Schoel, Nick Rasmuson

Table of Contents

- I. Introduction**
- II. Methodology**
- III. Analysis & Findings**
- IV. Implications**
- V. Conclusion**
- VI. References**

I. Introduction

This project highlights a critical issue that has far-reaching implications for player safety and the sport's future. By analyzing datasets, we aim to provide a clearer understanding of the frequency, severity, and long-term effects of concussions among NFL players.

Our journey into the data takes us beyond mere statistics; we explore the human aspect of these injuries, their impact on players' careers, and the steps the NFL has taken to mitigate risks. Through innovative visualizations, we unravel the complex dynamics of concussions, from on-field incidents to post-injury care and recovery processes. This analysis is not just about numbers; it is a narrative of resilience, health, and safety in one of America's most beloved sports.

As we dive deeper, we also examine how concussion trends have influenced NFL policies, training practices, and player equipment over the years. Our goal is to offer insights that not only inform but also inspire changes leading to a safer environment for current and aspiring players. This project is a call to action for coaches, medical professionals, players, and fans alike - to acknowledge the realities of concussions in the NFL and to work together towards a safer future in the sport. Through data, we seek to empower and educate, contributing to a more informed and health-conscious NFL community. Specifically, we want to answer the following business questions:

Q1: Based on historical data, which position is projected to have the highest rate of concussions in the upcoming season?

Q2: What was the average recovery time for players after experiencing a concussion?

Q3: During an NFL season, when is a player most likely to get injured?

Q4: Is there a correlation between a team's success and a player's susceptibility to concussions?

II. Methodology

For our group project, we explored several different datasets relating to the NFL and Injuries. After going through various datasets, we found the Concussion Injuries 2012-2014 data interesting because there was data that related to injuries in the NFL, the amount of time a player plays before and after being injured, and if the team was a winning team that year or not. We knew this data would be difficult to analyze because we would have to combine datasets and delete strings to convert variables to factors, but we chose it because we wanted to build a decision tree model to predict if a team was successful or not based on these variables.

Our merged dataset has 41 variables from three sources. Our dataset has a variety of data types with most being quantitative and discontinuous, such as height, age, year, number of injuries. There is also a suitable number of categorical variables, mostly Boolean/Binary variables. For example, whether the NFL player was injured pre-season (yes/no) or whether they were on the winning team (yes/no).

The original data was a folder and multiple datasets consisting of multiple datasets organized by the year and event. The folder had 3 datasets. To better analyze all the data, we combined all the datasets into one big dataset named “concussion.”

We also made sure the data was clean and complete. Checking for null and missing values. Where there were missing values, we found the mean value and imputed it into the null section. From there, we reduced the size of the dataset by removing some variables that we would not need for our analysis. For example, we deleted the “PlayerID” variable as it was a unique identifier for each player, but we were not analyzing a specific player. In the end, the most important variables that we had were “Winning Team,” “Position,” “AveragePlayTimeBefore,” “AveragePlayTimeAfter.” These four were the most helpful in drawing conclusions relating to injuries, their frequency, how to prevent them, and how they impacted team performance.

When doing advanced analysis like a decision tree, handling outliers is important. To maintain the accuracy of our results, we also reduced the number of records by handling outliers, only capturing values that were within three standard deviations of the mean. This reduced the number of records. The final dataset ended up with 20 variables.

To answer the questions related to concussion injuries, we utilized multiple analysis techniques such as data visualization, correlation analyses, and descriptive analyses. To determine concussion factors, we used decision tree analysis to determine what factors that influence a player's susceptibility to getting concussed and how this relates to performance. After we identified what some of the significant elements are, we used data visualization to map the frequency that these factors play out in the NFL and what variables may increase or decrease the impact of these variables. We used descriptive statistics to calculate concussions across the NFL, to get a baseline of what is the normal number of concussions in the league. We also used histograms, bar charts, and other forms of visualization to determine a baseline.

Original list of variables in the dataset and definition	Variables used for decision tree analysis
--	---

<p>Concussion Injuries 2012-2014.csv – (17 variables)</p> <ol style="list-style-type: none"> 1. ID – Summary of player name, the game, and the date 2. Player – Name of player 3. Team – Name of team the player played for 4. Game – Game title and date 5. Date – Date of the game and injury 6. Opposing_Team – The opposing that the player was playing 7. Position – The football position that the player holds 8. Pre_Season_Injury – Whether the player had an injury before the season started or not (yes/no) 9. Winning_Team – Whether the player was on the winning team that game or not (yes/no) 10. Week_of_Injury – Which week in the season the player was injured during (discrete) 11. Season – The year the season was played in 12. Weeks_Injured – How many weeks the player was injured (discrete) 13. Games_Missed – How many games the player missed (discrete) 14. Unknown_Injury – Whether the player had an unknown injury or not (yes/no) 15. Reported_Injury_Type – The type of injury/where it is located that the player experienced 16. Total_Snaps – Total number of snaps the player made during the game (discrete) 	<ol style="list-style-type: none"> 1. Team – Name of team the player played for 2. Game – Game title and date 3. Date – Date of the game and injury 4. Opposing_Team – The opposing that the player was playing 5. Position – The football position that the player holds 6. Pre_Season_Injury – Whether the player had an injury before the season started or not (yes/no) 7. Winning_Team – Whether the player was on the winning team that game or not (yes/no) 8. Week_of_Injury – Which week in the season the player was injured during (discrete) 9. Season – The year the season was played in 10. Weeks_Injured – How many weeks the player was injured (discrete) 11. Games_Missed – How many games the player missed (discrete) 12. Unknown_Injury – Whether the player had an unknown injury or not (yes/no) 13. Reported_Injury_Type – The type of injury/where it is located that the player experienced
--	---

<p>17. Play_Time_After_Injury - How many downs the player played for the rest of the season after the injury</p>	<p>14. Total_Snaps – Total number of snaps the player made during the game (discrete)</p> <p>15. Play_Time_After_Injury - How many downs the player played for the rest of the season after the injury</p> <p>16. Rnd – Which round an NFL player was drafted (discrete)</p> <p>17. Pick – Which number of pick an NFL player was drafted (discrete)</p> <p>18. Franchise – The full name of the team the player plays for</p> <p>19. Year From – The first year the player played in the NFL</p> <p>20. Year To – The last year the player played in the NFL</p>
<p>Feb 2020 SportsVizSunday.xlsx – (8 variables)</p> <p>18. Rnd – Which round an NFL player was drafted (discrete)</p> <p>19. Pick – Which number of pick an NFL player was drafted (discrete)</p> <p>20. Tm Abbr – Abbreviation of the team the player plays for</p> <p>21. Franchise – The full name of the team the player plays for</p> <p>22. Player – Name of the player drafted</p> <p>23. Pos. – The position that the player drafted plays</p> <p>24. Year From – The first year the player played in the NFL</p> <p>25. Year To – The last year the player played in the NFL</p>	

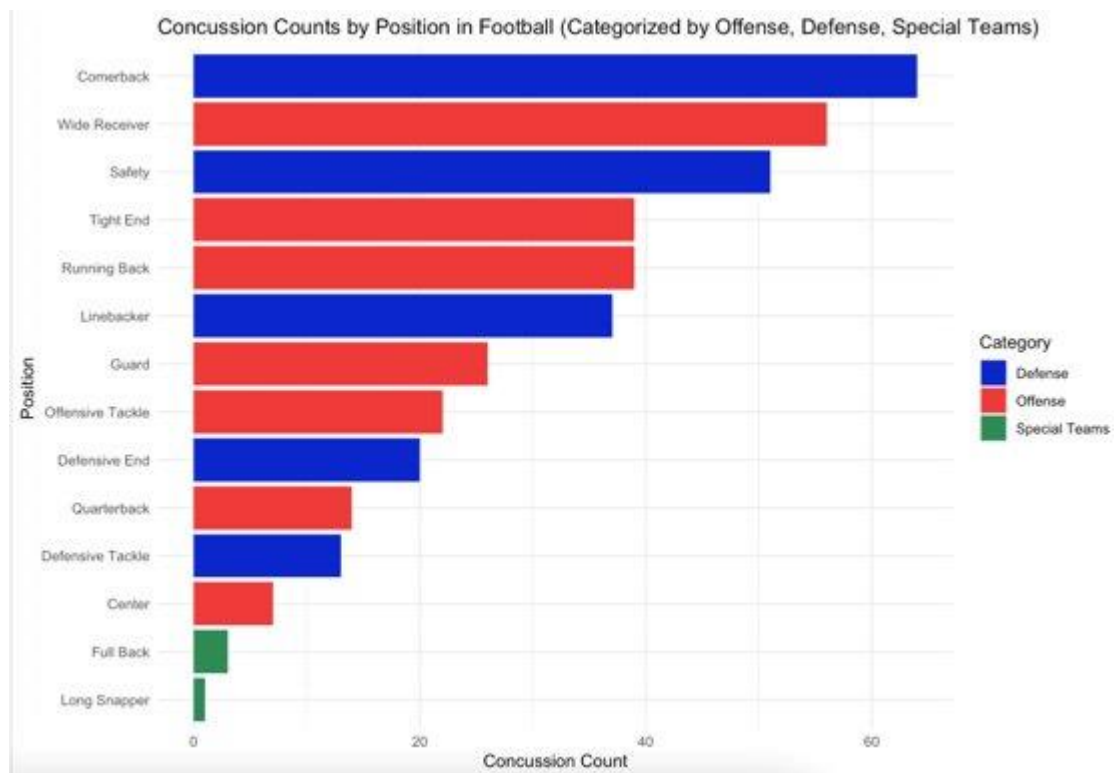
NFL Combine Data.xlsx – (16 variables)

- 26. Year – The year that the player did the NFL combine
- 27. Name – The name of the player
- 28. College – The college that the player plays for
- 29. POS – The position that the player plays
- 30. Height – The height of the player in inches (discrete)
- 31. Weight – The weight of the player in pounds (discrete)
- 32. Hand_Size – The hand size of the player in inches (discrete)
- 33. Arm_Length – The arm length of the player in inches (discrete)
- 34. Wonderlic - cognitive abilities test, scores range from 0 – 50 points (discrete)
- 35. 40_Yard – The time it takes to run 40 yards (discrete)
- 36. Bench_Press – How many reps of 225 lbs. on bench press the player was able to complete (discrete)
- 37. Vert_Leap – How high the player could vertically leap in inches (discrete)
- 38. Broad_Jump – How high the player could vertically leap in inches (discrete)
- 39. Shuttle – The time the player took to complete a shuttle in seconds (discrete)
- 40. 3Cone – The time the player took to complete a 3-cone drill in seconds (discrete)
- 41. 60Yd_Shuttle – The time the player took to complete a 60-yard shuttle in seconds (discrete)

III. Analysis & Findings

Finding 1: Position Matters

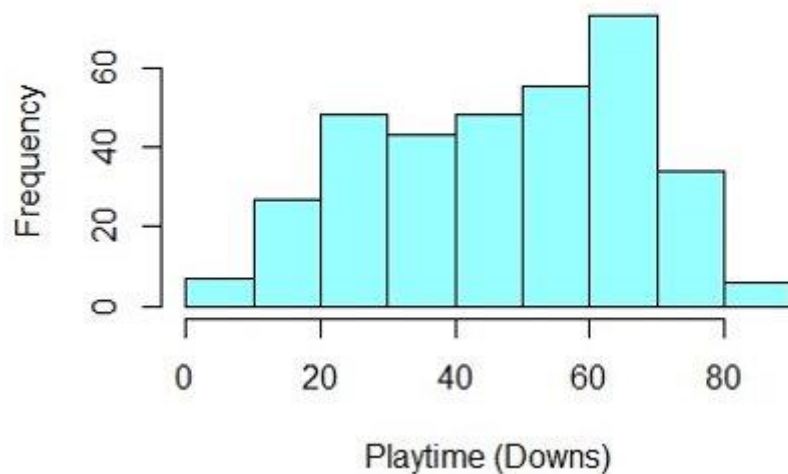
Intuitively, positions like quarterback or kicker, shielded by the team's formation, would appear safer. However, our bar chart analysis tells a different story, revealing a striking trend where cornerbacks incur the highest rate of concussions. The visual data clarifies those positions with greater exposure to open-field plays—where high-velocity impacts are more common—see a higher frequency of concussions. This pattern is substantiated by the data points on our chart, with defensive roles, especially cornerbacks, leading the count. This finding is a critical insight for coaches and players in strategizing for players' health and safety, and for medical staff in focusing prevention and treatment efforts. The data underscores the need for targeted safety measures for players most at risk in these high-impact roles.



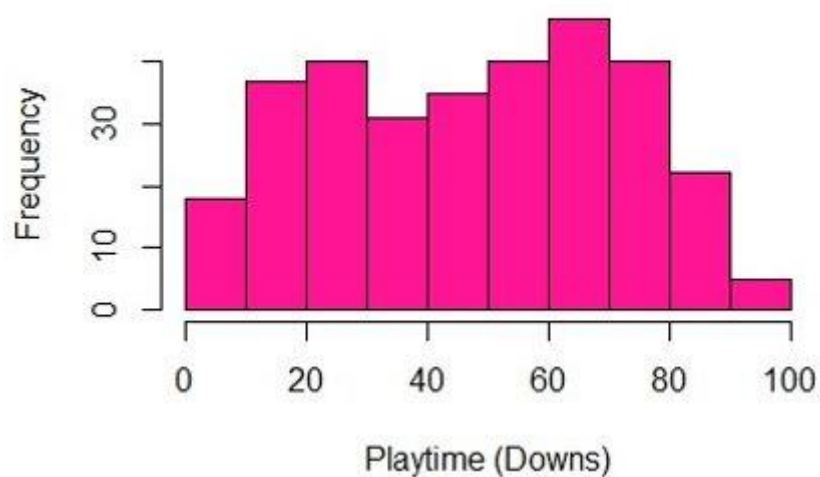
Finding 2: Playtime Decreases Post-Injury

We conducted a histogram comparison for playtime before injury and playtime after injury in our analysis in answering the question of “What was the average recovery time for players after experiencing a concussion?” The histogram of playtime before injury shows a slight skewed right distribution. The histogram of playtime after injury is less skewed. These histograms show that average playtime tends to decrease post-injury with a higher frequency of playtimes less than 20 downs.

Average Playtime Before Injury



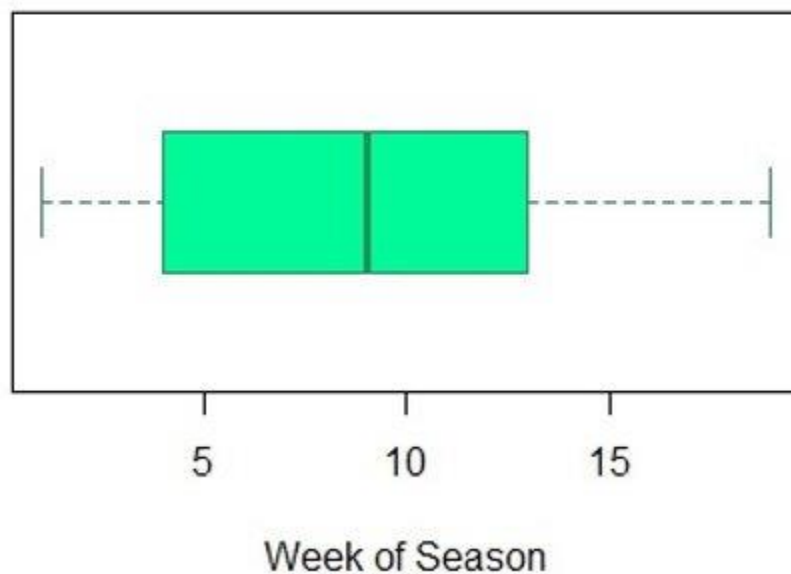
Average Playtime After Injury



Finding 3: Most Likely to Get Injured Halfway Through Season

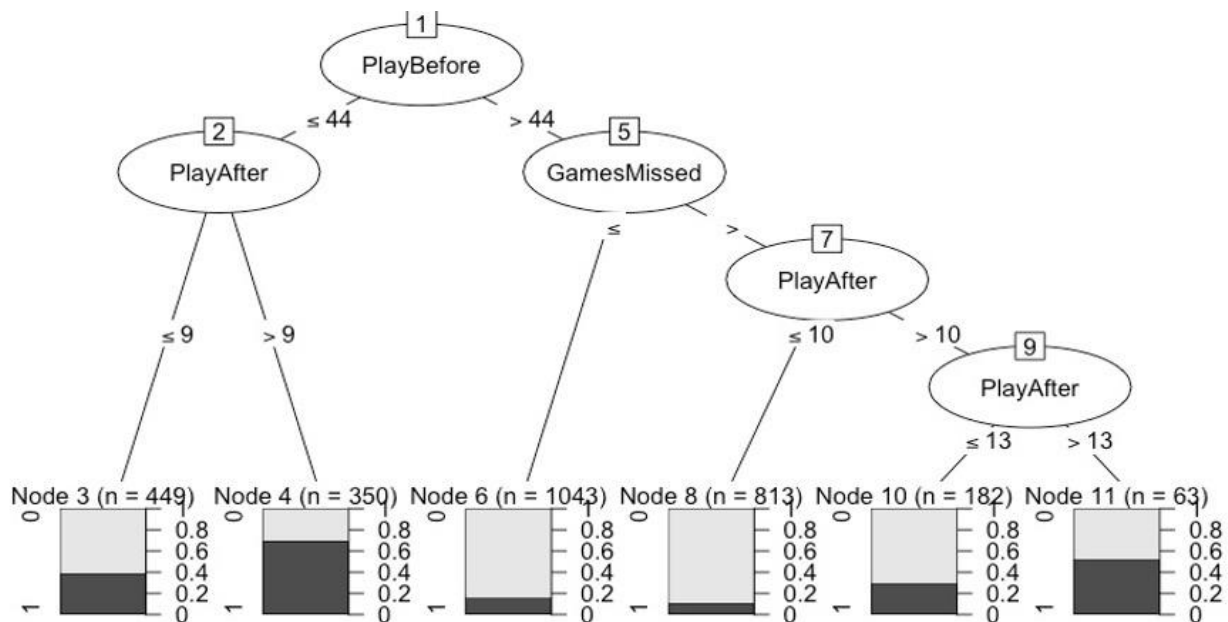
During an NFL season, when is a player most likely to get injured? An NFL football season is 18 weeks (about 4 months). Based on the boxplot analysis, we can see that the median is right under 10 weeks (about 2 and a half months), hovering around 9 weeks (about 2 months). This is halfway through the NFL football season. The box plot is also right skewed. This insight shows that it is more likely for an NFL player to get injured in the 1st half of the season rather than the latter half of the season.

Distribution of Week of Injury



Finding 4: Team Success is Directly Correlated with Injury

To answer the fourth question which evaluates whether we can predict team performance based off variables like AveragePlayTimeBeforeInjury and AveragePlayTimeAfter injury, we used a decision tree.



When creating our decision tree, we realized the factor that determines a team's success the most is PlayBefore. This variable is how many plays a player participates in before they are injured. This could be the most significant variable because of the relationship between player fatigue and injury. The more a player plays, the greater their chances of getting injured and damaging the team's performance. The next largest factor we discovered was PlayAfter. This is the number of plays that a player participates in after they are found to have an injury. Many players are continually reentering the game after they are injured, and consequently harming themselves and whether their team is successful or not. The last key factor in our decision tree is GamesMissed. If players adhere to policies in the NFL and wait for the proper amount of time they should recover from injury, their teams tend to be more successful.

IV. Implications

When looking at potential implications for the NFL and its players, the significance goes beyond which teams are practicing safe concussion practices. Our analysis will look at how the increased frequency of concussions can impact the length of a player's career and making sure that the players who are injured are fully healed before returning to the game. Often, organizations and athletes try to rush back into the playing field before they are ready and can get reinjured and end their career. We can find which materials like turf or grass protect players and cause less concussions and we will also be able to determine the most at-risk positions for getting concussions. All these factors will increase the awareness of safe practices in the NFL and make the related field safer and more focused on protecting its players. We conclude our study with four major implications:

- Referees need to be on high alert for player safety in the middle of the season.
- Skill positions are at the highest risk for injury.
- Players can educate themselves on safe tackling practices and avoid big hits.
- Too many players are remaining on the field after an injury.

V. Conclusions

Concussion management in the National Football League (NFL) has been a longstanding concern, prompting increased scrutiny and the implementation of various policies aimed at safeguarding the well-being of players. While progress has been made in enhancing concussion protocols, it is imperative to acknowledge that challenges persist.

The NFL has undeniably made strides in addressing the issue of concussions. The league has implemented comprehensive policies, including baseline testing, stricter return-to-play protocols, and increased education on recognizing and reporting symptoms. However, despite these advancements, it is evident that there is still much work to be done. The effectiveness of these policies relies heavily on their consistent implementation and enforcement, which remains an area requiring continuous attention.

Traditionally, the focus on concussions has often centered around positions like linemen and players known for their frequent and forceful hits. However, recent data suggests that concussions are not exclusive to these positions. The nature of the game exposes players in various roles to potential head injuries, necessitating a comprehensive approach to concussion prevention and management.

Enforcing concussion protocols consistently is paramount to ensuring player safety and team success. Referees play a pivotal role in this process, as they are responsible for identifying potential concussions during games. The league must prioritize the training of referees to recognize the signs of a potential concussion and empower them to intervene decisively when necessary.

In conclusion, analyzing concussion trends in the NFL reveals both progress and challenges. While the league has taken positive steps in addressing the issue, there is an ongoing need for refinement and a proactive approach to evolving concussion policies. Acknowledging that concussions can affect players across positions and emphasizing the strict enforcement of protocols, including enhanced penalties for violations, are critical components of a comprehensive strategy for player safety in the NFL. Only through a concerted effort from players, referees, and league officials can the NFL continue to make meaningful strides towards minimizing the impact of concussions on its athletes and helping teams be successful.

VI. References

Concussions in the NFL (2012- 2014),

<https://www.kaggle.com/datasets/rishidamarla/concussions-in-the-nfl-20122014>. Accessed 5 Dec. 2023.

Feb2020 SportsVizSunday,

www.nflcombinerresults.com/nflcombinedata.php?year=&pos=&college=. Accessed 5 Dec. 2023.

NFL Combine Results,

www.nflcombinerresults.com/nflcombinedata.php?year=&pos=&college=. Accessed 5 Dec. 2023.