NFL Injuries

Francis Bunker, Luis Li, Siddharth Patel, Aryan Ramsinghani

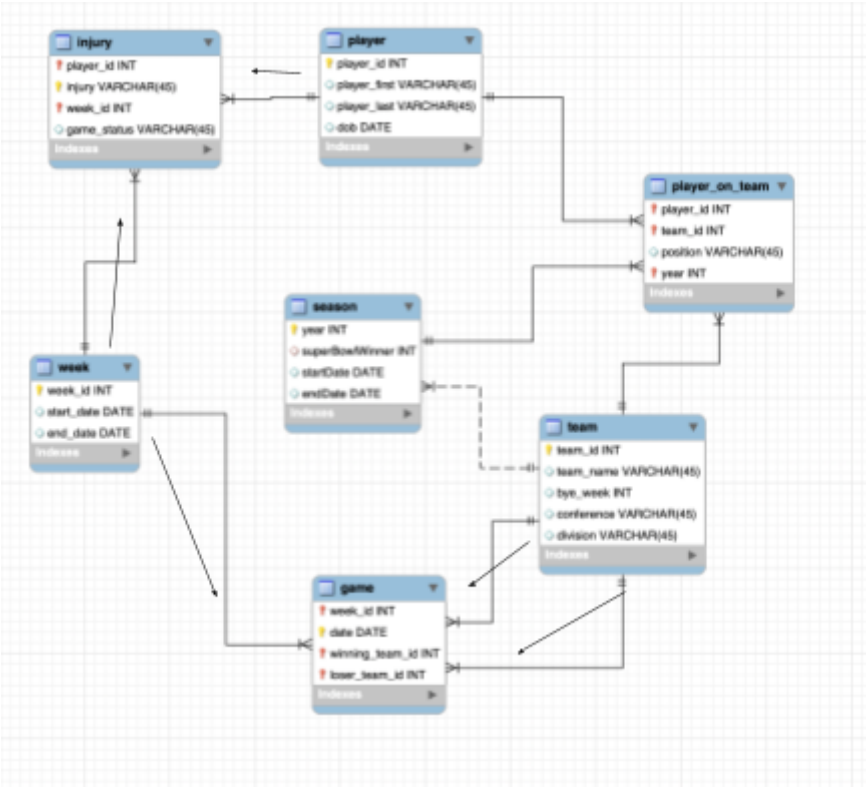
Northeastern University, Boston, MA, USA

CS 3200 Introduction to Databases

Introduction

This report discusses the design, implementation, and findings of a database designed to answer questions about injuries in the NFL for the 2024 season. The database holds information about each injury for each week, who it happened to, what team they are on, as well as information about the games each team played. With this, we can analyze which injuries come up the most and look for other factors that may be correlated with injuries. We hope that this data will help understand the impact of injuries on players and provide insight when looking at prioritizing which injury preventions and treatments can be improved. As sports enthusiasts, we see how much of a negative impact injuries can have on teams, so we hope our database can help answer questions that may come up.

Database Design



Our database is designed to store a weekly report of injuries for all teams and players in the NFL. The core of our design is the **injury** table. This table holds the week, a player, their injury, and their game status. Essentially, it reports who is injured for each week of the season. It has a foreign key for the **player** and for the **week**, referencing the **player** and **week** tables respectively. The **player** table stores the player's first name, last name, and date of birth. The **week** table stores the start and end date of that week. The rest of the tables store information about which team each player is on for some **season** and the **games** of the season. The **player_on_team** table has foreign keys to the **season**, **team**, and **player** tables. It also has a field for the player's position. The **game** table stores data on who won and lost each game using the team ids. We used foreign keys everywhere where it is appropriate in the design as shown in the diagram above. We created composite foreign keys for our game and **player_on_team** tables. Note that for the **week** table, each week must have a unique number, so if you were to include multiple seasons in this database, you would need to factor that in. A possible change that could be made if this database was used for multiple seasons is adding a season to the week table.

Data Sources and Methods

For our dataset, we decided to look at the 2024 season. We first created a few trivial tables that are needed for the database to function. These include week, team, game, and season. The week table shows the start and end dates of each week for the season. The team table lists each team and some basic information about them. The game table shows each game throughout the season and which team won and which team lost. The season table shows who won the Super Bowl. These tables were filled in manually from information available through the NFL website [1].

Our largest tables are player, player_on_team, and injury. Our data for these tables comes from both Pro Football Reference and the official NFL website [1][2]. Pro Football Reference was relatively easy to work with, as they have easy ways to export their data. We used their data for our team roster information, and we wrote a simple Java program that combines all the team rosters into a single table that includes players from all teams, as well as which team they are on. An important note on their roster data is that it includes all players who played on a team at least once throughout a season. This means there may be players on the original NFL team roster who do not appear on the Pro Football Reference list because they were out for the full season, and it may include players who were not on the original roster because they moved teams throughout the season. In our database, this is represented through our player_on_team and player tables. The player table lists distinct players, each with a unique ID. The player_on_team table lists each team a player played at least one game on.

To get our data for injuries, we used the official NFL website. For each week of each season, they have a webpage with a table for each team's injury report. For example, nfl.com/injuries/league/2024/reg1 includes 32 tables (one for each team) for Week 1 of the 2024 season. Each table lists the injured players for that team, their injury, and their game status (as well as some other information that is irrelevant for this project). We used a Google App Script to scrape this data, as it can put its output into a Google Sheet, which is easy to convert to data we can insert into SQL. The App Script can combine all the weekly report tables together into a single table, then iterate over the different weeks. The script also does a little bit of cleanup on the data so we can easily use it with our other tables. It separates out first and last names,

shortens the injury titles provided on the NFL website, and deals with situations where the NFL data includes multiple injuries in the same table column. We then wrote another simple Java program that replaces all the player names with the appropriate unique player IDs that were created earlier with the other Java program. Below is the JavaScript function we wrote in Google App Scripts to get data from the NFL website. We are then able to call this function for our Google Sheet to import the data.

```
function getAllNFLInjuries() {
  const maxWeek = 18; //18 weeks to look at
  let output = [["Player_first", "Player_last", "Injury", "Week", "Game Status"]];

  //loop though eacch week and get data
  for (let week = 1; week <= maxWeek; week++) {
    const baseUrl = `https://www.nfl.com/injuries/league/2024/reg${week}`;
    const html = UrlFetchApp.fetch(baseUrl).getContentText();

    // Extract all team tables
    const tables = html.match(/<table[\s\S]*?<\/table>/g);

    for (let t = 0; t < tables.length; t++) {
      const rows = tables[t].match(/<tr[\s\S]*?<\/tr>/g) || [];

      for (let r = 1; r < rows.length; r++) {
        const cells = rows[r].match(/<td[\s\S]*?<\/td>/g);

        const values = cells.map(c => decodeHTMLEntities(c.replace(/<[^>]*>/g, '').trim()));

        if (values.length >= 5 && values[4] && values[4].trim() !== "") {
          const player = values[0];
          const injuryRaw = values[2];
          const gameStatus = values[4];

          const [first, ...rest] = player.split(' ');
          const last = rest.join(' ').trim();

          const injuries = splitAndCleanInjuries(injuryRaw);

          for (const injury of injuries) {
            output.push([first, last, injury, `${week}`, gameStatus]);
          }
        }
      }
    }
  }

  return output;
}

```

Database Analysis Questions

There are several important questions our database can answer. Below are some of the ones we think are most important and the outputs.

How many injuries did each team have during the season?

SQL:
SELECT t.team_name, t.conference, t.division, COUNT(*) AS total_injuries
FROM injury i
JOIN player_on_team pot ON i.player_id = pot.player_id

```
JOIN team t ON pot.team_id = t.team_id
GROUP BY t.team_id, t.team_name, t.conference, t.division
ORDER BY total_injuries DESC;
```

Result:

team_name	conference	division	total_injuries
Carolina	National	South	197
New England	American	East	167
Tampa Bay	National	South	165
New Orleans	National	South	164
San Francisco	National	West	152
Dallas	National	East	142
Miami	American	East	138
Los Angeles C	American	West	135
New York J	American	East	129
New York G	National	East	125
Arizona	National	West	118
Cleveland	American	North	115
Las Vegas	American	West	115
Baltimore	American	North	112
Buffalo	American	East	110
Washington	National	East	109
Chicago	National	North	109
Green Bay	National	North	108
Tennessee	American	South	107
Houston	American	South	102
Seattle	National	West	98
Detroit	National	North	97
Jacksonville	American	South	94
Indianapolis	American	South	93
Pittsburgh	American	North	93
Philadelphia	National	East	90
Cincinnati	American	North	89
Los Angeles R	National	West	84
Minnesota	National	North	66
Atlanta	National	South	64
Kansas City	American	West	59
Denver	American	West	54

Findings: there is a significant difference in the number of injuries between different teams. In the next few queries, we will explore possible explanations for this.

Is there any clear correlation between the number of games a team wins and the number of injuries their players have?

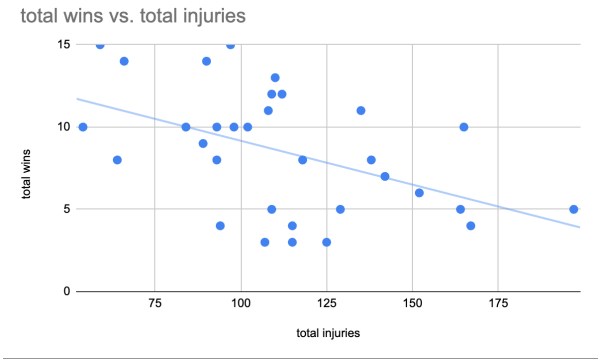
First, we can get the total wins for each team:

```
SELECT t.team_name, COUNT(g.winning_team_id) AS total_wins
FROM team t
LEFT JOIN game g ON t.team_id = g.winning_team_id
GROUP BY t.team_id, t.team_name
ORDER BY total_wins DESC;
```

Result:

team_name	total_wins
Kansas City	15
Detroit	15
Minnesota	14
Philadelphia	14
Buffalo	13
Baltimore	12
Washington	12
Green Bay	11
Los Angeles C	11
Tampa Bay	10
Denver	10
Los Angeles R	10
Seattle	10
Pittsburgh	10
Houston	10
Cincinnati	9
Indianapolis	8
Arizona	8
Miami	8
Atlanta	8
Dallas	7
San Francisco	6
New York J	5
Carolina	5
Chicago	5
New Orleans	5
Jacksonville	4
New England	4
Las Vegas	4
New York G	3
Cleveland	3
Tennessee	3

If we graph the total wins and number of injures for each team, we get:



Based on this, it appears that there might be some negative correlation between the number of total injuries and wins.

If we do not factor in repeating injuries (the same player being injured across multiple weeks), do the injury numbers across teams get closer?

SELECT t.team_name, t.conference, t.division, COUNT(DISTINCT i.player_id) AS distinct_injured_players
FROM injury i
JOIN player_on_team pot ON i.player_id = pot.player_id

JOIN team t ON pot.team_id = t.team_id
GROUP BY t.team_id, t.team_name, t.conference, t.division
ORDER BY distinct_injured_players DESC;

Result:

team_name	conference	division	distinct_injured_players
Carolina	National	South	47
New Orleans	National	South	44
New England	American	East	44
San Francisco	National	West	42
New York G	National	East	42
Miami	American	East	42
Arizona	National	West	42
Tampa Bay	National	South	42
Cleveland	American	North	41
Los Angeles C	American	West	41
Houston	American	South	40
Detroit	National	North	39
Baltimore	American	North	39
Dallas	National	East	38
New York J	American	East	38
Tennessee	American	South	36
Jacksonville	American	South	36
Las Vegas	American	West	36
Chicago	National	North	35
Washington	National	East	34
Kansas City	American	West	34
Seattle	National	West	34
Buffalo	American	East	32
Cincinnati	American	North	32
Los Angeles R	National	West	32
Indianapolis	American	South	31
Green Bay	National	North	31
Pittsburgh	American	North	31
Philadelphia	National	East	30
Minnesota	National	North	29
Atlanta	National	South	26
Denver	American	West	22

This data still varies a good amount, but more of the data falls towards the middle, which makes sense.

Which type of injury is the most common to come up in all reports?

SQL query:
SELECT injury, COUNT(*) AS occurrence_count
FROM injury
GROUP BY injury
ORDER BY occurrence_count DESC
LIMIT 10;

Result:

injury	occurrence_count
Knee	592
Ankle	495
Hamstring	353
Shoulder	264
Concussion	256
Groin	164
Calf	120
Foot	118
Back	102
Hip	101

This shows that knee, ankle, and hamstring injuries are the most common by a good amount.

What is the total number of injuries for each week (is there any trend over time)?

SQL query:
```
SELECT w.week_id, COUNT(i.player_id) AS total_injuries
FROM week w
LEFT JOIN injury i ON w.week_id = i.week_id
GROUP BY w.week_id
ORDER BY w.week_id;
```

Result:

week_id	total_injuries
1	114
2	155
3	163
4	180
5	210
6	209
7	208
8	214
9	221
10	188
11	176
12	166
13	193
14	177
15	228
16	221
17	219
18	198

There is no clear trend here, which suggests that the number of injuries per week does not change very much throughout the season. This might be explained by the fact that people can become injured in pre-season training.

At which age do the most injuries occur?

SQL query:
```
SELECT TIMESTAMPDIFF(YEAR, p.dob, w.start_date) AS age, COUNT(*) AS injury_count
FROM injury i JOIN player p USING (player_id) JOIN week w USING (week_id)
GROUP BY age
ORDER BY injury_count DESC
LIMIT 5;
```

Result:

age	injury_count
25	511
24	392
26	388
23	383
27	372

There are a lot of confounding factors that could impact the number of injuries at different ages, but overall we do not see a huge trend here, as there are a higher number of NFL players who are 25 years old compared to the other ages, which explains them getting more injuries.

Which positions have the most injuries?

SQL query:
```
SELECT pt.position, COUNT(*) AS injury_count
FROM player_on_team pt JOIN injury i USING (player_id)
GROUP BY pt.position
ORDER BY injury_count DESC
LIMIT 5;
```

Result:

position	injury_count
WR	479
CB	319
RB	272
LB	271
TE	190

There is a huge difference here, and it confirms that some positions are much more prone to injuries than others.

We can also look at the game status information, as there can exist players who have an injury but are not out. Here is a query to get the proportions of players with different statuses who are injured.

SQL query:
```
SELECT game_status, COUNT(*) AS count,
ROUND(COUNT(*) / (SELECT COUNT(*) FROM injury), 4) AS proportion
FROM injury
GROUP BY game_status
```

ORDER BY proportion DESC;

game_status	count	proportion
Questionable	1836	0.5337
Out	1378	0.4006
Doubtful	226	0.0657

This shows that roughly half of injured players are out of the game when they are injured.

Conclusions

This database provides some powerful insights on player injuries. Our data suggests that there may be a correlation between low injuries and winning games, that this is a fairly large range in the number of injuries that different teams get, and that by a lot, the most common injuries are knee and ankle injuries. With this information, there are a number of things that can be done for the future of the NFL and how to decrease the harm caused by injuries. Looking into why some teams were able to have fewer injuries than others may give insights on how to prevent them. The fact that teams with less injuries tend to be more successful shows that injury prevention may lead to a more successful season.

Group Contributions

Throughout the project we worked closely on every major phase of the project. While all of us contributed to all the tasks we divided most of the tasks among us equally. Francis and Luis focused on collecting the data and building the database. They cleaned, organized and integrated the data from the NFL and ProFootballReference websites. Sid and Aryan focused on coming up with interesting questions and developing the SQL queries for those questions. All of us collectively wrote the report and created the presentation. We also worked together to come up with our analysis and come up with a suitable conclusion for our findings. This collaborative approach allowed each member to use their strengths to the fullest and contribute meaningfully towards the end goal of this project.

[1] National Football League, “Official Latest NFL Injury Report for Players - Week 13 of the 2025 Season,” NFL.com, 2025. [Online]. Available: https://www.nfl.com/injuries/. [Accessed: Nov. 26, 2025].

[2] Sports Reference LLC, “Pro Football Stats, History, Scores, Standings, Playoffs, Schedule & Records,” Pro-Football-Reference.com, 2025. [Online]. Available: https://www.pro-football-reference.com/. [Accessed: Nov. 26, 2025].