NFL Player & Team Performance Database Project Report

Lance Heinrich, Matias Avila
CSCI126 – Database Systems
Spring 2025 – Dr. David Ruby

Introduction

This project focuses on building a relational database that analyzes the relationship between NFL player performance and overall team success across multiple seasons. The NFL was chosen as the domain due to its rich history of statistical data and the opportunity to explore how different positional contributions impact team outcomes.

Our primary goal was to design a normalized, scalable database that integrates player-level and team-level performance data, supporting meaningful analysis through structured SQL queries. We aimed to answer questions such as:

- How strongly do quarterback statistics correlate with team win totals?
- How important are defensive statistics like points allowed?
- Do rushing or receiving stats better predict success?

Dataset Selection

We selected two complementary datasets from Kaggle that provide detailed information on NFL players and teams:

Dataset 1: NFL Player Stats

- Source: Kaggle
- Link: https://www.kaggle.com/datasets/zynicide/nfl-football-player-stats
- Size: ~30,000 rows, 50+ attributes
- Structure: Contains individual player statistics per game, including passing yards, rushing yards, receiving touchdowns, interceptions, tackles, field goals, and more.
- Modifications: We trimmed a lot of the player profile data like the school they went to and their date of birth, instead we just kept their name and position. We aggregated player game data into season totals per player to focus on season-level analysis using season totals. We changed all of the team names in our player data to match the team_id of our teams in the team table.

Dataset 2: NFL Team Stats (2003–2023)
- Source: Kaggle
- Link:
https://www.kaggle.com/datasets/nickcantalupa/nfl-team-data-2003-2023
- Size: ~640 rows (32 teams × 20 years), 20–30 attributes
- Structure: Yearly team performance metrics including wins, losses, points scored/allowed, offensive/defensive ranks, turnovers, and playoff appearances.
- Modifications: We allowed the team table to auto-increment and uploaded all of the team names into it.

Potential Insights Enabled by This Data:
- How quarterback performance correlates with team win totals.
- Whether rushing or receiving stats are better predictors of playoff appearances.
- Which positions have the greatest impact on team success.
- The role of defensive strength in winning seasons.

Database Design and Normalization

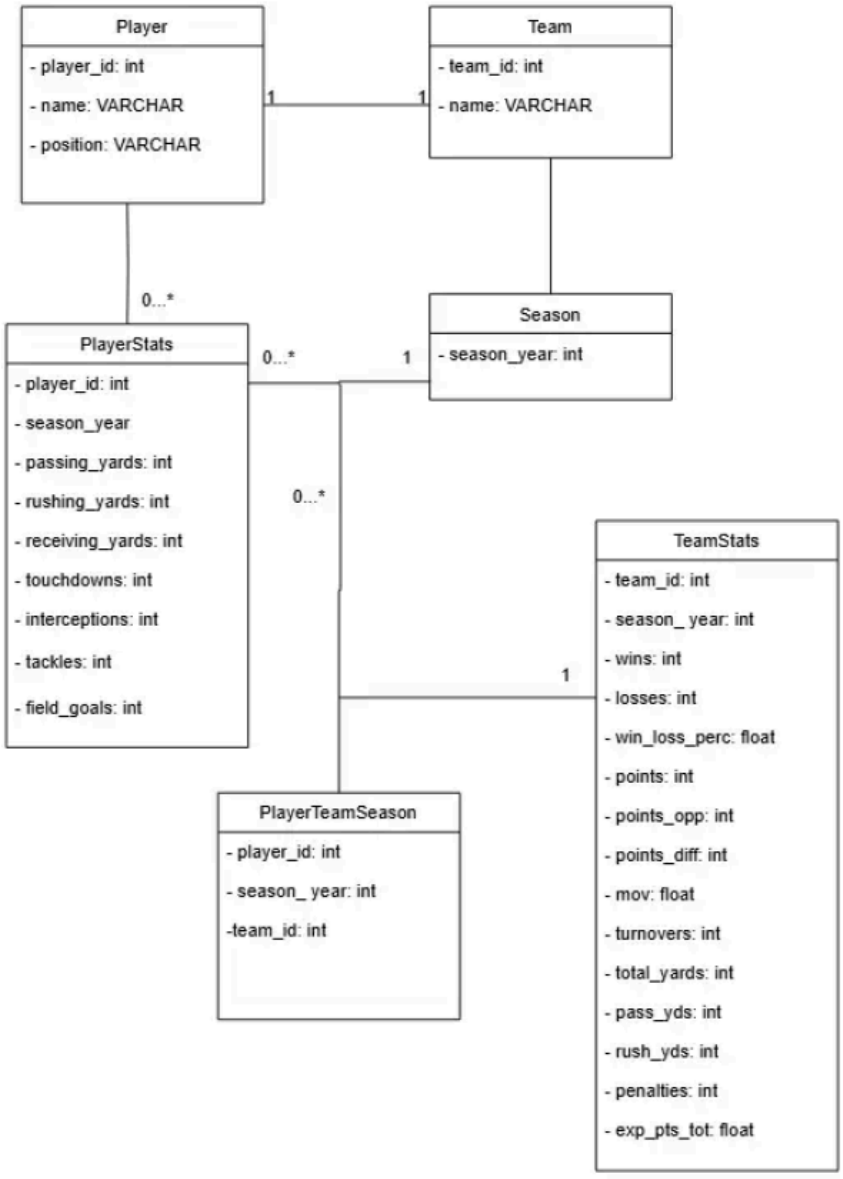
To support meaningful analysis, we designed a relational schema that follows Third Normal Form (3NF) principles. The schema separates static descriptive data from dynamic seasonal performance data.

Relational Schema Overview:
Player - Static player info (player_id, name, position)
Team - Static team info (team_id, name)
Season - Season years (season_year)
PlayerTeamSeason - Links player, team, and season (join table)
PlayerStats - Aggregated player stats per season
TeamStats - Team performance metrics per season

Key Design Choices:
- Fact tables (PlayerStats, TeamStats) for seasonal metrics.
- Identity tables (Player, Team, Season) to reduce data duplication.
- PlayerTeamSeason join table to track which player played for which team each season.

Normalization (3NF Justification):
- 1NF: All tables use atomic columns.
- 2NF: Non-key attributes fully depend on the primary key.
- 3NF: No transitive dependencies; attributes relate directly to the primary key.

UML Diagram:



Implementation

Technologies Used:

- Database System: MySQL
- Data Source Format: JSON for player data(Converted to CSV using python) CSV file for team data.
- Data Modification: Python (For aggregating/modifying the large datasets)
- Tools: MySQL Workbench

Schema Creation:

- Table creation with primary keys and foreign keys.
- Indexing and data integrity constraints.
- Data cleaning (standardizing team names, matching ids to team names).

Data Exploration and SQL Queries

We wrote 12 SQL queries and one view, focusing on key performance questions. Queries involved joins, subqueries, grouping, and aggregations.

Query Topics

1. Top 10 Highest Passing Yard QBs All Time with Team Wins
    - How impactful were these record seasons for their team wins?
2. Average Wins for Teams With vs. Without a 3000-Yard Quarterback
    - How important is a star Quarterback?
3. Avg Passing TDs of QBs on Playoff vs. Non-Playoff Teams
    - How important is a QBs ability to throw touchdowns?
4. Average Wins for Teams With vs. Without a 1000+ Rushing Yard RB
    - How important is a star running back?
5. Avg Rushing Yards of RBs on Teams With 10+ Wins vs. 10 or Fewer Wins
    - How important is a strong running game?
6. Average Wins for Teams With vs. Without a 1200+ Yard Wide Receiver
    - How important is a star wide receiver?
7. Average Wins for Teams With vs. Without 3000+ Total WR Receiving Yards
    - How important is a strong receiving corps?
8. Average Wins for Teams With vs. Without a 1000+ Receiving Yard TE
    - How important is a star TE?
9. Avg Receiving Yards of TEs on Teams With 10+ Wins vs. 10 or Fewer Wins
    - How important are receiving tight ends?
10. Avg Wins for Teams Allowing Fewer Than 300 Points vs. 300+ Points
    - How important is a strong defense?

11. Average Turnovers by Teams With 10+ Wins vs. Under 10 Wins
    - Do winning teams turn the ball over less?
12. Average Wins for Teams With a Player Recording 100+ Tackles
    - Does having a high-tackle defender correlate with more wins?
13. Average Tackles per Player on Teams With 10+ Wins vs. Under 10 Wins
    - Do players on successful teams tend to rack up more tackles?
14. Avg interceptions by QBs on Teams With 10+ Wins vs. Under 10 Wins
    - Do winning teams have QBs that throw fewer interceptions?

Results and Insights

1. While some of the top QB performances resulted in high win totals, a few did not lead to playoff-level success This highlights that even elite QB yardage alone doesn't always guarantee wins, suggesting that other factors like defense or turnovers also play a key role.
2. Teams with a QB throwing for over 3000 yards averaged about 9.2 wins, while those under 3000 averaged only about 6.3 wins. This shows that having a productive quarterback is a strong predictor of team success.
3. Playoff teams' QBs averaged nearly 14 passing touchdowns, compared to about 8.8 for non-playoff teams. QB scoring ability (touchdowns) is a key driver of team victories and playoff appearances.
4. Teams with a 1000+ yard rusher averaged 8.85 wins, while those without averaged only 7.22 wins. This suggests that while rushing success does correlate with wins, it may be slightly less impactful than QB performance.
5. Playoff teams' RBs averaged about 405 rushing yards, versus 312 for non-playoff teams. A stronger running game can contribute to team success, though the effect size here appears moderate.
6.Teams with a star wide receiver (1200+ yards) averaged 9.36 wins, while others averaged 7.33 wins. Elite receiving performance positively impacts team success.
7.Teams with a strong receiving corps (3000+ yards from all WRs) averaged 10 wins, while others averaged 7.45.Depth at the WR position may be even more important than having just one star receiver.
8.Teams with a 1000+ yard TE averaged about 9 wins, compared to 7.6 for those without. While tight end performance does show some correlation with success, it may be a secondary factor.

9.Winning teams averaged about 241 receiving yards from their TEs, versus 189 for others. Contributions from the TE position can support a strong offense but are not typically the primary driver.
10.Teams allowing fewer than 300 points averaged nearly 11 wins, while those allowing more averaged only 7.5. This confirms that defense, especially limiting points allowed, is one of the strongest predictors of success.
11.Winning teams averaged about 21 turnovers, compared to nearly 27 for losing teams. Avoiding turnovers is strongly correlated with winning, reinforcing the importance of disciplined offense and defense.
12.Teams with a high-tackle defender averaged 8.1 wins, only slightly higher than those without (8 wins). While tackling leaders may contribute defensively, raw tackle numbers alone may not directly translate to more wins.
13.Players on winning teams averaged about 8.25 tackles, compared to 7.5 on losing teams. Higher tackle rates could suggest better overall defensive involvement, but again, not the strongest success factor compared to points allowed or turnovers.
14. QBs on winning teams threw slightly fewer interceptions (6.15) than QBs on losing teams (6.56).While the difference is small, limiting interceptions still appears to play a supportive role in achieving winning seasons.

Team Contributions

Lance Heinrich - Dataset Description, Schema design, Data Aggregation/Cleaning, SQL queries & Results
Matias Avila - UML diagram, schema justification, Written Report, Presentation Slides

Conclusion

This project successfully integrated player and team data into a relational database, enabling analysis of how individual stats relate to team success. The results show that every position with individual players performing well improves team win percentage but really highlights the importance of quarterbacks and defense specifically in winning seasons and demonstrates the value of structured data analysis in sports.