NFL Potential Trade Analysis

Chris Haleas
Indiana University
Bloomington, Indiana
chaleas@iu.edu

Andrew Eby
Indiana University
Bloomington, Indiana
ebya@iu.edu

***Abstract*—This study explores the impact of NFL trades between offensive players (quarterbacks, wide receivers, running backs, and tight ends) using a Kaggle dataset of player stats from 2012-2023. By calculating normalized position-specific performance scores and aggregating them into team metrics, a heatmap is created to identify positional strengths and weaknesses. A linear regression model predicts how a player's performance may change when traded, helping users assess potential trade impacts. The research highlights the role of statistical modeling in enhancing decision-making in team management and fan engagement.**

***Key Terms*—NFL Player Trades, Offensive Player Performance, Trade Impact Analysis, Heatmap Visualization, Linear Regression Models, Sports Analytics**

I. Introduction

In recent years, sports analytics has gained traction as a powerful tool for enhancing strategic decision-making, not only among coaches and managers but also among fans and analysts. Trade events happen frequently throughout the year and are a crucial part of building a successful franchise. Teams consider many distinct factors when trading players, such as what positions they lack skill in, the value of a player, or the potential they possess. The variety of factors involved in trading players make it challenging to predict trade outcomes.

Currently, teams assess the possible outcomes of a trade by looking at historical performance data, team needs, and player statistics, helping them make informed decisions about which trades might contribute to future success. However, with the growth of machine learning and access to detailed statistics, it has become possible to build models that can predict the impact of a trade on a player's performance. These models help us understand whether a player would work well with another team and how their performance might change based on the trade.

This is exactly what we've done. By preprocessing historical NFL data, assigning scores to players by position, and aggregating them into team scores by position, we can visualize where teams are successful and where they lack skills in specific positions. We provide this information to our users, allowing them to make informed decisions when evaluating potential players. Our linear regression model then predicts how players will perform once they are traded for another player. This provides a significant solution to the problem of predicting the potential outcomes of player trades.

II. Methods

A. Data Preprocessing

- Finding our Dataset: We searched through Kaggle for a dataset with comprehensive data and found one containing 195 attributes related to individual player and team statistics.
- Removing Outliers: Outliers in our dataset were players with an insignificant number of games played. Since our focus was on starting players or those with significant playing time, we filtered the data to include only players who played at least 8 games in a season.
- Feature Selection: We selected key features to analyze the performance of each offensive position, including wide receivers, quarterbacks, running backs, and tight ends, based on common metrics used by the NFL for player rankings, as well as our own experimentation with attributes and knowledge of the game. The below are the features we chose for each position:
  - **Quarterbacks (QB)**:
    * pass_td_pct: Percentage of attempts that result in touchdowns.
    * ypa: Average yards gained per passing attempt.
    * comp_pct: Percentage of completed passes.
    * pass_ypg: Average number of passing yards per game.
    * passing_yards: Total passing yards accumulated.
    * passing_air_yards: Yards gained in the air before completion.
  - **Running Backs (RB)**:
    * rush_td_pct: Percentage of rush attempts that result in touchdowns.
    * ypc: Average yards gained per rush attempt.
    * yptouch: Average yards gained per touch (rushes + receptions).
    * rush_ypg: Average rushing yards per game.
    * rushing_yards: Total rushing yards accumulated.
    * receptions: Number of times the player catches the ball.
  - **Wide Receivers (WR) and Tight Ends (TE)**:
    * rec_td_pct: Percentage of receptions that result in touchdowns.
    * ypr: Average yards gained per reception.
    * target_share: Percentage of team targets directed toward the player.
    * air_yards_share: Percentage of total team air yards directed toward the player.
    * targets: Total number of times the player is targeted for a reception.
    * receptions: Number of times the player catches the ball.
    * receiving_yards: Total receiving yards accumulated.

Min-max Normalization: After selecting features, we applied Min-max normalization. This ensures that all features are scaled within the same range, so no single feature (like passing yards or rushing attempts) dominates due to having larger values, allowing all features to contribute equally to the analysis. Below, you will find the equation for Min-max.

$$X_{\text{norm}} = \frac{X - \min(X)}{\max(X) - \min(X)}$$

B. Equations

- Player Performance Score (PPS): We created an equation to assign a score to the performance of a player based on their position... PPS $= \frac{1}{n}\sum_{i=1}^{n} x_i'$, where n is the total number of normalized features and x_i' is the normalized value of each feature.
- Team Position Score (TPS): The TPS is calculated by averaging the PPS of all significant players in a given position... TPS $= \frac{1}{m}\sum_{i=1}^{m} \mathrm{P}_i$, where m is the number of significant players in the position and P_i is the PPS of the i-th player in the position.

C. Data Visualization

- Heatmap: We used Seaborn, a Python data visualization library, to create heatmaps of the Team Position Scores for each position, organized by year and team. This provided a clear and intuitive visual representation for users of our analysis, highlighting which positions each team lacks strength in and where they excel.

D. Machine Learning Models

- Linear Regression: We implemented a Linear Regression model using Scikit-learn, with player statistics from the 2022-2023 season as input features and the average player scores from the 2023-2024 season as the actual outputs for comparison. Once the model was trained on data from players in the same position, it was used to predict the performance scores of two players selected by the user for trade evaluation. These predicted scores represent how each player is expected to perform in the 2023-2024 season, providing insights into their potential value and helping evaluate both sides of a trade.

E. Model Evaluation Metrics

- R^2 Score: This score calculates how well the model's predictions match the actual outcomes. An R^2 score of 1 means perfect predictions, while a score closer to 0 indicates poor predictive power.
- Root Mean Squared Error (RMSE): RMSE calculates the average magnitude of the errors between predicted and actual values, giving more weight to larger errors. Lower RMSE values indicate better model performance.
- Mean Absolute Error (MAE): MAE calculates the average of the absolute differences between predicted and actual values. It represents the average error in the predictions, with lower MAE values indicating better model accuracy.

III. Results

A. Heatmap Visualization Results

Below are the results of our heatmap visualizations, with one heatmap for each offensive position. The X-axis represents the year, while the Y-axis represents the team. The color intensity of each block corresponds to the strength of the Team Position Score, with darker shades indicating stronger performance. These visualizations allow readers to identify which positions teams excel in and where they may lack skill by year. The results indicate that teams often have very strong quarterbacks, followed by strong running backs and wide receivers, while many teams lack standout players in tight end positions.

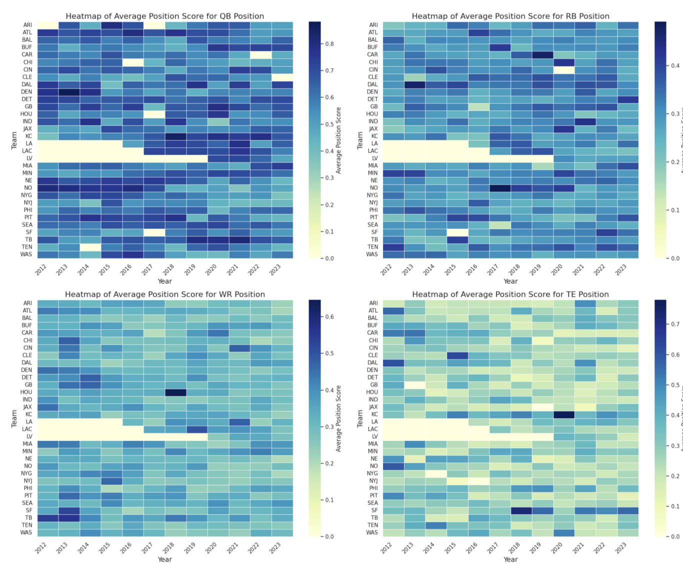

Fig. 1. Heatmap visualization of Team Position Scores across years for each offensive position. Darker shades indicate stronger performance.

As seen in the heatmap, we identified the strongest and weakest teams for each position in each season. The standout data points are found in Fig. 2.

	Position	Strongest Team	Score_Strongest	Weakest Team	Score_Weakest
0	QB	MIA	0.734256	NYG	0.375081
1	RB	DET	0.416176	LAC	0.213719
2	WR	SEA	0.386028	NE	0.197937
3	TE	SF	0.562972	CIN	0.131295

Fig. 2. Table of the strongest and weakest teams for each position in the 2023 season, based on their Team Position Scores.

B. Linear Regression Results

In Figure 3, I have tested the Linear Regression model on a trade scenario for each position, keeping the trades between positions for simplicity in the examples. The model is trained and tested 20 times, with the scores representing the average of these 20 runs. This extensive testing ensures that the performance scores are highly reliable.

	Player Position	Traded Player	Received Player	Traded Player Score	Received Player Score	Training R^2 Score	Training RMSE	Training MAE	Test R^2 Score	Test RMSE	Test MAE
0	QB	Daniel Jones	Tua Tagovailoa	16.9435	79.9979	0.2913	0.1471	0.1103	0.4459	0.1791	0.1627
1	QB	Tua Tagovailoa	Daniel Jones	79.9979	16.9435	0.4985	0.1358	0.1101	-0.2782	0.2215	0.146
2	RB	Breece Hall	Derrick Henry	151.1845	48.7529	0.3958	0.1463	0.1155	0.3409	0.1442	0.1176
3	RB	Derrick Henry	Breece Hall	48.7529	151.1845	0.4351	0.1452	0.1122	0.0949	0.1533	0.1216
4	TE	George Kittle	Travis Kelce	203.0103	212.8364	0.5728	0.131	0.0986	0.2342	0.1172	0.0989
5	TE	Travis Kelce	George Kittle	212.8364	203.0103	0.5239	0.1345	0.1036	0.7106	0.1025	0.0812
6	WR	Alec Pierce	Cooper Kupp	216.9345	188.1722	0.6144	0.1115	0.0857	0.7403	0.1161	0.085
7	WR	Cooper Kupp	Alec Pierce	188.1722	216.9345	0.6537	0.1131	0.0858	0.6952	0.1036	0.0796

Fig. 3. Table summarizing the trade performance for NFL players across different positions in the 2023-2024 season. The table includes the R^2 scores, RMSE, MAE for both training and test datasets, along with the performance scores of traded and received players for each position: Quarterback (QB), Running Back (RB), Tight End (TE), and Wide Receiver (WR).

Below are the visualized results of each tested trade. The Linear Regression model was trained and tested 20 times for each position, with the results averaged for high accuracy. Each figure represents the performance of the trade for a specific position: Quarterback (QB), Running Back (RB), Tight End (TE), and Wide Receiver (WR).

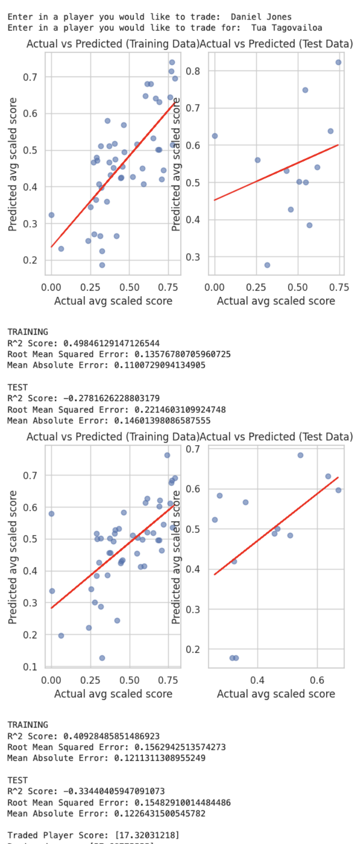

Fig. 4. Results of the Linear Regression model for the Quarterback (QB) trade. The results indicate that Tua Tagovailoa is expected to perform better for his new team compared to Daniel Jones' performance for his.

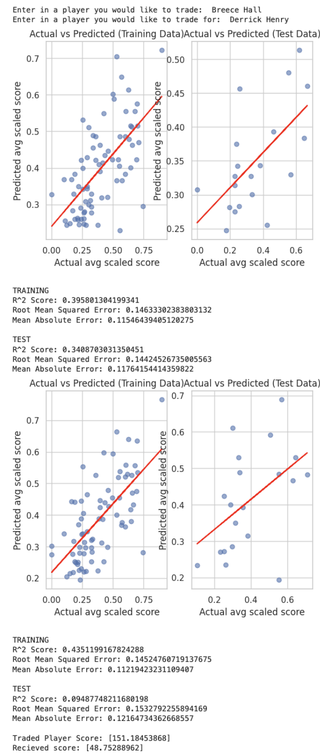

Fig. 5. Results of the Linear Regression model for the Running Back (RB) trade. The results indicate that Breece Hall is expected to perform better for his new team compared to Derrick Henry's performance for his.

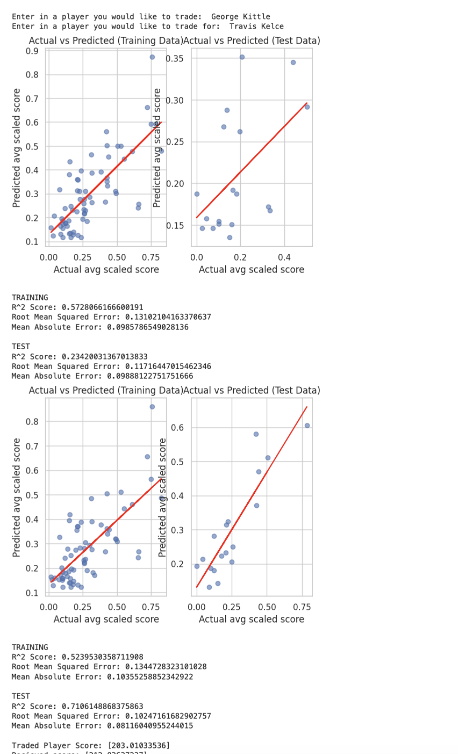

Fig. 6. Results of the Linear Regression model for the Tight End (TE) trade. The results indicate that Travis Kelce is expected to perform better for his new team compared to George Kittle's performance for his.

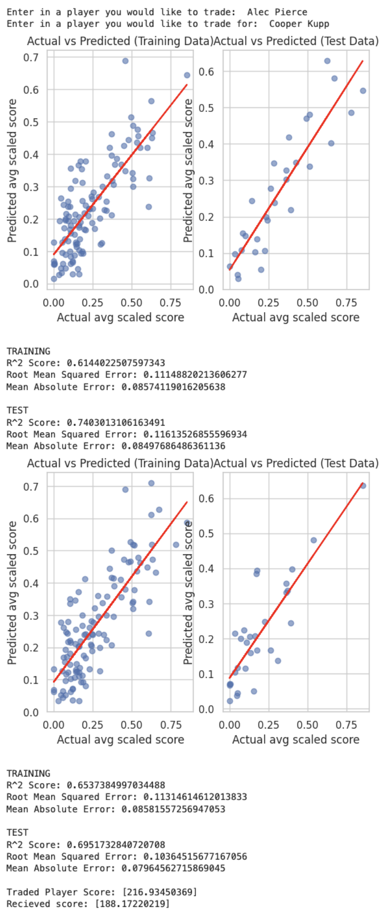

Fig. 7. Results of the Linear Regression model for the Wide Receiver (WR) trade. The results indicate that Alec Pierce is expected to perform better for his new team compared to Cooper Kupp's performance for his.

IV. Discussion

The model's performance varies across different trades. The model performs well for wide receiver trades, with high R^2 scores (0.7403 and 0.6952) and low RMSE values (0.1161 and 0.1036). It also shows good performance for tight end trades, with a high R^2 score (0.7106) and low RMSE (0.1025). However, the model performs poorly for quarterback trades, where the R^2 score is negative (-0.2782) and the RMSE is the highest (0.2215). This indicates that the model's predictions for quarterback trades are less reliable. Running back trades show moderate performance, with R^2 scores of 0.3409 and 0.0949, and RMSE values of 0.1442 and 0.1533. The inconsistency in performance across different positions is a limitation of the current approach. This suggests that the model struggles to generalize across all positions, especially for quarterback and running back trades. To address this, future research could explore more complex models, such as decision trees, which may better capture the underlying patterns in the data.

Overall, this project has deepened our understanding of football, particularly in identifying the attributes that determine a player's value and performance in trades. It has also enhanced our skills in working with datasets and interpreting data, making us more knowledgeable when selecting models and fine-tuning their parameters. This work brings us a step closer to predicting trade outcomes using data science, offering valuable insights for future applications in sports analytics.

V. Author Contribution Statement

Chris Haleas was responsible for preprocessing and visualizing the data, as well as organizing the final research paper. Andrew Eby focused on implementing the linear regression model and explaining its application in the document.

VI. References

[1] Monarch, Funk. (2024). NFL Stats 2012-2023. Kaggle. $https : //www.kaggle.com/datasets/philiphyde1/nfl - stats - 1999 - 2022/data?select = weekly_{p}layer_{d}ata.csv$

[2] Dabbura, I. (2018). K-means Clustering: Algorithm, Applications, Evaluation Methods, and Drawbacks. Towards Data Science. $https : //towardsdatascience.com/k - means - clustering - algorithm - applications - evaluation - methods - and - drawbacks - aa03e644b48a$

[3] Fahad, A., et al. (2014). A Survey of Clustering Algorithms for Big Data: Taxonomy and Empirical Analysis. *IEEE Trans. Emerging Topics Comput.*, 2(3), 267-279. $https : //doi.org/10.1109/TETC.2014.2330519$

[4] Ye, Y. Data, Data Representation, and Data Preprocessing (2). Indiana University Bloomington.

[5] Ye, Y. Cluster Analysis and K-means. Indiana University Bloomington.

[6] NFL.com. (2023). QB Index: Ranking All 66 Starting Quarterbacks from the 2023 NFL Season. $https : //www.nfl.com/news/nfl - qb - index - ranking - all - 66 - starting - quarterbacks - from - the - 2023 - nfl - season$

[7] NFL.com. (2023). RB Index: Ranking All 72 Starting Running Backs from the 2023 NFL Season. $https : //www.nfl.com/news/rb - index - ranking - all - 72 - starting - running - backs - from - the - 2023 - nfl - season$

[8] Analytics Vidhya. (2021). MAE, MSE, RMSE, Coefficient of Determination, Adjusted R-Squared: Which Metric is Better? $https : //medium.com/analytics - vidhya/mae - mse - rmse - coefficient - of - determination - adjusted - r - squared - which - metric - is - better - cd0326a5697e$

[9] RegenerativeToday. (2020). Multiple Linear Regression in Python - sklearn. YouTube. $https : //www.youtube.com/watch?v = wH_{e}zgftiy0$