Frisbee Data Dynamics: Boston Glory

Team Members: *Audrey Sellers, Mohamad Ali Saadeddine, Priyanka Chaudhari*

Project Management and Workflow

Background Research:

Ultimate Frisbee (UF) is a sport that is played all over the world, but we will be focusing on the Ultimate Frisbee Association (UFA) throughout this project. During an ultimate game, each team has seven players on the field with an objective of getting the frisbee to the opposing team's end zone.[1] There are several strategies that involve offensive plays and defensive zones that teams implement to win games.[1] In general, the key to winning games is to score goals with the best lineup possible. The positions of ultimate are Handlers, Cutters, Poppers, Wings, the Cup, Short Deeps, and Middle Flats.[2] Our Business Advisor has informed us that they know position is highly correlated with scoring goals. It has also been seen that teams who win have significantly fewer backward passes,[3] which aligns with our project's overall objective: to increase pass yardage and reduce players' turnover rates.

Along with goal scoring, it is imperative for ultimate teams to have a cohesive offensive lineup that executes assists and passes. There are different ways to decide which players should start and receive the most playing time. In fact, a research study conducted on a national UF team concluded that players exert more energy when playing higher-ranked opponents.[4] This indicates a need for higher offensive performance, more effort, and calculated lineup formations during games against lower-ranked teams. Similarly, through feature selection, It has been found that ultimate teams are more likely to win when players who have more complete passes and assists.[5] The same project also found that when players who have a high drop rate play, the team is less likely to win due to an increased turnover rate.[5]

---

[1] "Ultimate 101." 2018. WatchUFA. January 29, 2018. https://watchufa.com/league/ultimate101.
[2] "Ultimate Frisbee Positions - Which One Is Best for You?" n.d. Theuap.com. Accessed July 13, 2024. https://www.theuap.com/ultimate-frisbee-positions-which-one-is-best-for-you.
[3] Lam, Hilary, Otto Kolbinger, Martin Lames, and Tiago Guedes Russomanno. 2021. "State Transition Modeling in Ultimate Frisbee: Adaptation of a Promising Method for Performance Analysis in Invasion Sports." Frontiers in Psychology 12: 664511. https://doi.org/10.3389/fpsyg.2021.664511.
[4] Castillo, Daniel, Javier Raya-González, Aaron T. Scanlan, Marta Domínguez-Díez, and María C. Madueno. 2023. "Influence of Opponent Ranking on the Physical Demands Encountered during Ultimate Frisbee Match-Play." Sports Biomechanics 22 (7): 822–33. https://doi.org/10.1080/14763141.2020.1766101.
[5] Go, Caitlin. n.d. "Exploring Predictors of Team Success in Ultimate Frisbee: An Analysis of Game Statistics for Stanford Women's Ultimate."

Problem Statement:

Currently, Boston Glory does not sufficiently utilize the vast amount of data available to make strategic decisions. The primary objective is to enhance the performance of the Boston Glory frisbee team by leveraging data analytics to provide actionable insights. The focus is on identifying key areas of offensive and defensive strengths and weaknesses and suggesting on-field strategies that can lead to wins and championships within the UFA league.

Scope:

This project will encompass data collection and preprocessing, exploratory data analysis, feature engineering, and machine learning model development. The data will be sourced from the UFA API and processed through a Google Cloud infrastructure, including Google Big Query for data storage and analysis. The goal is to analyze various performance metrics, such as points per game, yards received, and defensive weaknesses, to provide comprehensive insights for improving the team's strategies.

Data Collection and Preprocessing:

Utilize the UFA API to collect relevant data from the year 2023-2024. Implement an ETL pipeline using Google Cloud Functions to automate data ingestion into Google Big Query. Perform data cleaning and preprocessing to ensure data quality like:
[i] Dropped irrelevant columns that did not seem essential for our analysis and subject matter.
[ii] Data Transformation: Converted categorical features to appropriate data types for analysis and modeling, including extracting players names and playerid's from the list form that they were initially in. This involved writing a function that could parse through the JSON list of strings and give us an individual set of values.

Exploratory Data Analysis (EDA):

We conducted exploratory data analysis to learn more about the data and ultimate frisbee in general. This allowed us to explore the distribution of the players' statistics, identify key performance metrics, and create visualizations to uncover trends and patterns.

Although we wanted to focus on Boston Glory development, looking at full league data was imperative to draw conclusions about game play and player talent. After exploring the league statistics, we decided that higher performing players tended to have more than 5 points per game, which was an important decision when it came to filtering the dataset. Then, we compared the average PPG between Glory and other teams to see where we rank in the league. The initial data analysis revealed that Boston Glory scores fewer points per game (7.16 PPG) compared to the league average (7.43 PPG), indicating a need for offensive improvement. We also began to look at how playing time affected performance metrics with the intention of seeing if players performed better at certain times of the game. We found that players who played more frequently scored more goals, which was an obvious finding.

We also looked at defensive weakness, as it is not highly studied in the sport. We concluded that Boston Glory's defense is weaker than the rest of the league on average by allowing more goals against them. However, defensive weakness was very difficult to appropriately quantify due to lack of defensive metrics in the sport.

Next, we wanted to highlight key correlations between points per game and other performance metrics. We found that yards received, yards per reception, seconds played, and yards thrown had strong correlations with points per game. This further confirmed our initial hypothesis that seconds played was correlated with performance. After discussing our results with Boston Glory, they suggested further investigating one of the correlations, so we decided to hone in on yards per reception in the project.

Feature Engineering:

Further we developed features that capture critical aspects of gameplay, such as yardage, pass completion rates, and defensive stats.

Expected Findings:

After conducting exploratory data analysis and feature engineering, we had three main goals. Firstly, we wanted to develop a model that predicted yards per reception because of the positive correlation and Boston Glory's suggestion, so we created a linear regression model based on league data. Secondly, we wanted to develop a model that predicted a certain lineup as a win or loss. Lastly, we aimed to use linear optimization to create an ideal lineup based on certain opponent statistics. Accomplishing these goals is intended to identify top-performing players and distinguish their contributions to the team's success, as well as provide in-game strategies to enhance overall team performance.

Machine Learning Model Development:

- **Linear Regression:** The first model we worked on was a linear regression model to predict a player's average yards per reception based on several attributes. We chose to predict yards per reception due to the stronger positive correlation of the created PPG feature (0.23). The model demonstrates a mean absolute error (MAE) of 1.73, meaning the predictions are 1.73 points off from the actual value on average. The model's mean squared error (MSE) of 8.52 implies larger errors were present, while the median absolute error (MedAE) of 1.15 reflects a typical prediction error that is slightly more consistent. However, the model's performance is not ideal, as the R-squared value of 0.0764, suggests that 7.64% of the variance in the predicted average yards per reception is explained by the model. These results suggest that the model may be underfitting and needs further improvement.
- **Logistic Regression:** The next was developed with the goal to predict a lineup as a win for Boston Glory to play against certain teams using Logistic Regression which demonstrates moderate performance. With a precision of 54.19%, the model correctly predicts a win just over half the time, but there's still a significant risk of false positives. The recall of 49.62% suggests that the model misses out on identifying a substantial portion of actual wins, leading to false negatives. An accuracy of 57.74% indicates the model is slightly better than random guessing, but there's room for improvement. The F1 score of 0.5181 reflects a moderate balance between precision and recall, but both are

below ideal levels. The log loss of 0.6869 suggests that the model is not particularly confident in its predictions, which could mean the predicted probabilities aren't well-calibrated. The ROC AUC of 0.5815 indicates that the model is only marginally better than random guessing in distinguishing between wins and losses. Overall, while the model is functional, it would benefit from further tuning and enhancements to improve its predictive accuracy and reliability.

Row	precision	recall	accuracy	f1_score	log_loss	roc_auc
1	0.541966426858...	0.496158068057...	0.577386934673...	0.518051575931...	0.686926989493...	0.581488511488...

- **Linear Optimization:** The linear optimization model we developed dynamically selects Boston Glory's starting lineup by weighing both offensive and defensive effectiveness against the specific strengths and weaknesses of the opposing team. The model is designed to be interpretable by management, providing a clear rationale for player selection based on metrics such as goals, assists, completions, and blocks. By dynamically adjusting these weights according to the opponent's profile (e.g., prioritizing defensive players against a strong offensive team), the model empowers management to make informed decisions that are directly tied to the strategic needs of each game. This setup not only enhances decision-making on the field but also facilitates communication between analysts, coaches, and management, ensuring that lineup choices are data-driven and strategically sound.

```
Optimal Starting Lineup:
    firstName   lastName
0       Henry    Babcock
26   Benjamin      Sadok
30        Ray  Tetreault
51  Nathanial       Dick
53       Luka    Govedic
65      Henry    Laseter
74   Benjamin      Sadok
78        Ray  Tetreault
79       Ryan     Tirner
```

Future Recommendations:

- **Enhance Defensive Metrics and Analysis:** Expand the collection of defensive data by integrating advanced tracking technologies, allowing for more granular analysis of defensive plays and strategies.
- **Integrate Real-Time Data and Dynamic Adjustments:** Incorporate real-time data during games to dynamically adjust strategies, optimize player rotations, and account for player fatigue.

- **Explore Advanced Machine Learning Techniques:** Utilize ensemble learning methods and deep learning techniques to improve predictive accuracy, address underfitting, and model complex player interactions**.**
- **Develop Opponent-Specific Strategies:** Create detailed opponent profiles and counter-strategy models to tailor Boston Glory's game plans based on the strengths and weaknesses of specific teams.

Limitations:

A major limitation we had was our initial assumptions about the given data. One of our initial ideas was to explore how certain players performed in different times of the game. Further, with that result, we would be able to determine which players shine in certain quarters of the game. Initially we were going to use how many seconds were played and goal scoring, and we also needed the quarter in which the goal was scored. However, due to mismatched game IDs in the given data, we were unable to go forward with this idea.

Another limitation was the lack of work in defensive UF strategies and scarce defensive statistics in the sport. Due to this, we also wanted to explore defensive strategies, as there is a limited amount of research in it. However, due to said limited defensive metrics in the sport, as a whole, we were unable to go forward with this approach.

While the project introduced several advanced models, including logistic regression and linear optimization, there are notable limitations. The linear regression model for predicting yards per reception showed signs of underfitting, suggesting that more complex modeling techniques or feature engineering may be necessary. The logistic regression model, though functional, demonstrated moderate accuracy and recall, indicating potential for further improvement. The linear optimization model, while effective in creating dynamic lineups, does not account for player fatigue or real-time data, which could influence game outcomes. Additionally, the project's reliance on historical and static data limits its ability to adapt to real-time game conditions or unexpected player performance changes. Future work should focus on integrating more dynamic data sources, improving model calibration, and exploring more sophisticated machine learning techniques to enhance overall predictive accuracy and strategic value.

Conclusion:

After conducting exploratory data analysis, background research, and meetings with Boston Glory, we found that the ultimate goal of our Capstone Project was to identify and implement ways to win. With this in mind, we developed three machine learning models. While the current Logistic Regression model provides a basic predictive capability, its moderate performance metrics suggest there is significant room for improvement. Future steps which can be included are tuning hyperparameters for better model calibration and incorporating additional features like player synergy and game context. The linear regression model to predict average yards per reception by player did not perform as well as we expected. With an MAE of 1.73 and MSE of 8.52, this model is underfitting on the dataset. This could have been due to the small dataset provided, as we had to aggregate statistics by player due to duplicates which left us with less rows than originally intended. Due to this, future hyperparameter tuning, and perhaps, principal component analysis should be explored in the future so that the fit improves and the model captures nonlinearity. Lastly, the linear optimization model is able to select a starting lineup for Boston Glory by producing the players with the best offensive and defensive strengths that are best suited for opponents. This model can allow coaches and management to use new data to select a lineup that is likely to win against a given opponent. Although we will not be able to see the full effects of our work, we were eager to experience a real-world sports analytics project and work with professionals in the analytics industry.

Code:

The coding portion of our analysis can be found at: [Colab file]
GitHub: https://github.com/priyankachaudhari08/Frisbee-Data-Dynamics

References:

1. "Ultimate 101." 2018. WatchUFA. January 29, 2018. https://watchufa.com/league/ultimate101.
2. "Ultimate Frisbee Positions - Which One Is Best for You?" n.d. Theuap.com. Accessed July 13, 2024. https://www.theuap.com/ultimate-frisbee-positions-which-one-is-best-for-you.
3. Lam, Hilary, Otto Kolbinger, Martin Lames, and Tiago Guedes Russomanno. 2021. "State Transition Modeling in Ultimate Frisbee: Adaptation of a Promising Method for Performance Analysis in Invasion Sports." Frontiers in Psychology 12: 664511. https://doi.org/10.3389/fpsyg.2021.664511.
4. Castillo, Daniel, Javier Raya-González, Aaron T. Scanlan, Marta Domínguez-Díez, and María C. Madueno. 2023. "Influence of Opponent Ranking on the Physical Demands Encountered during Ultimate Frisbee Match-Play." Sports Biomechanics 22 (7): 822–33. https://doi.org/10.1080/14763141.2020.1766101.
5. Go, Caitlin. n.d. "Exploring Predictors of Team Success in Ultimate Frisbee: An Analysis of Game Statistics for Stanford Women's Ultimate."

Appendix:

Summary for Promotional Purposes

Boston Glory, a professional Ultimate Frisbee team, aims to enhance its performance in the Ultimate Frisbee Association (UFA) through data analytics. The project identifies key offensive and defensive strengths and weaknesses to provide actionable insights for strategic decision-making.

Problem Statement:
Boston Glory has been underutilizing data in strategic decisions. This project focuses on maximizing pass yardage, minimizing turnovers, and optimizing lineup decisions to improve the team's performance and secure wins in the UFA league.

Project Scope:

- **Data Collection:** Gathered data using the UFA API.
- **Data Preprocessing:** Cleaned and prepared the data for analysis.
- **Exploratory Data Analysis (EDA):** Analyzed historical performance to identify patterns.
- **Feature Engineering:** Developed features like yardage and defensive stats to enhance model accuracy.
- **Machine Learning Models:** Utilized Google Cloud infrastructure to predict outcomes and optimize strategies.

Key Findings:

- Boston Glory scores an average of 7.16 points per game (below the league average of 7.43 PPG).
- Defensive analysis shows the team allows more goals than the league average.
- **Machine Learning Models:**
    - Linear Regression predicted average yards per reception with some underfitting (MAE 1.73).
    - Logistic Regression achieved 57.74% accuracy in predicting win probability.
    - Linear Optimization was used to dynamically select the starting lineup based on opponent analysis.

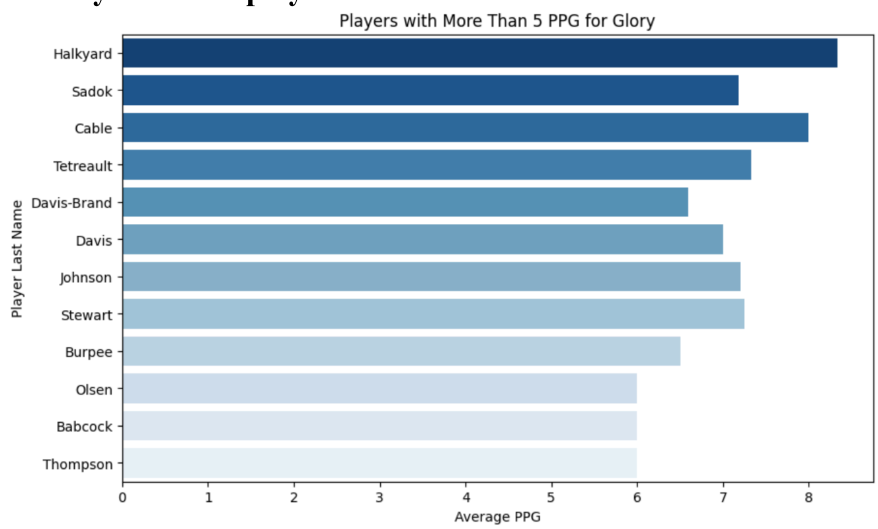
Future Recommendations:

- Enhance defensive metrics with advanced tracking technologies.
- Incorporate real-time data for dynamic strategy adjustments.
- Explore advanced machine learning techniques to improve predictive accuracy.

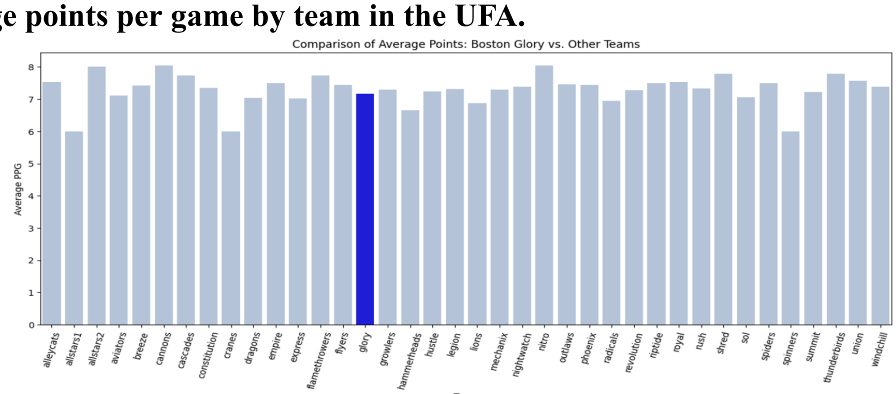
Conclusion:
The project identified key performance metrics and developed predictive models that provide actionable insights for Boston Glory. While promising, there is potential for further enhancement

through advanced techniques and real-time data integration, laying a foundation for data-driven strategies to help Boston Glory secure more wins in the UFA league.

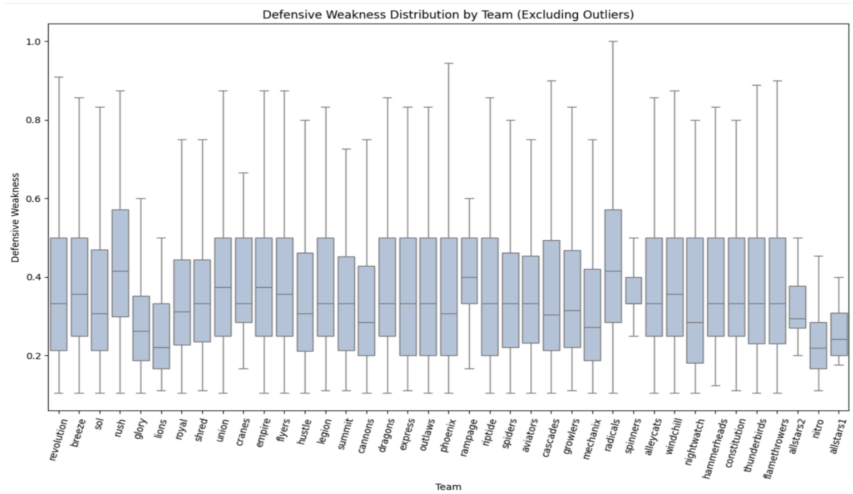
A. Top Boston Glory offensive players.



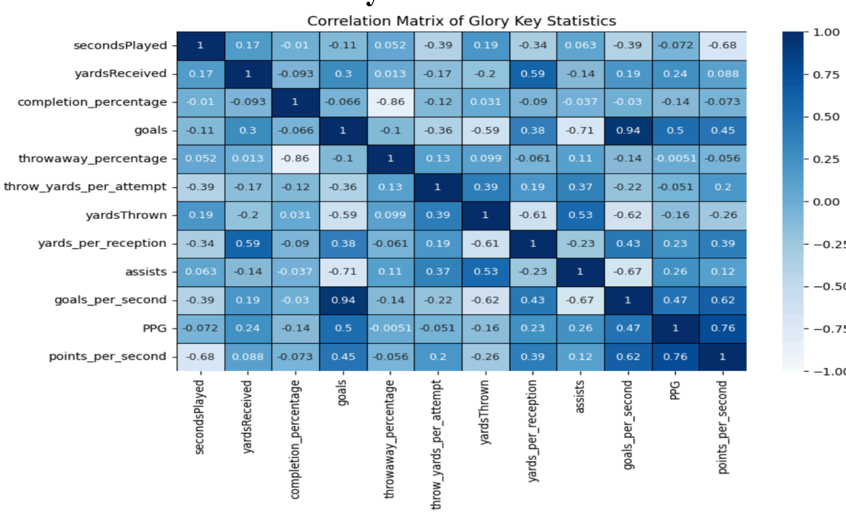
B. Average points per game by team in the UFA.



C. Defensive weakness by team in the UFA



D. Correlation matrix of UFA key attributes.



E. Linear regression model evaluation via BigQuery SQL.

Row	mean_absolute_erro	mean_squared_error	mean_squared_log_e	median_absolute_err	r2_score	explained_variance
1	1.731725362278...	8.518394331397...	0.475648590043...	1.145429944020...	0.076393972191...	0.130078274506...

Contribution:

Name	**Coding**	**Theory**
Audrey Sellers	BigQuery SQL and Exploratory Data Analysis, Linear Regression Model	Poster, Powerpoint Presentation, Report, Analysis in Colab
Mohamad Ali Saadeddine	BigQuery SQL and Linear Optimization Model	Poster, Powerpoint Presentation, Report
Priyanka Chaudhari (Project Manager)	ETL Pipeline(data collection and processing), BigQuery SQL and Logistic Regression Model	Poster, Powerpoint Presentation, Report