

## **Executive Summary of a Deep Dive Into Ultimate Frisbee**

### Introduction:

Ultimate Frisbee is a sport in which two teams play in a football field-like environment where they advance a frisbee up and down the field with the aim of scoring goals by throwing the frisbee to a receiving player in the end zone. The Ultimate Frisbee Association (UFA) is the premier professional Ultimate Frisbee league in the U.S.. Our primary goal was to analyze as much as we could into what aspects of Ultimate Frisbee determine playing style, winning, winning teams, and top players.

### Methods:

Methods used for the project include the use of Agglomerative Nesting (AGNES) and Divisive Analysis (DIANA) models for hierarchical clustering to determine optimal player clusters and styles of play based on player data from 2017-23 (excluding 2020) where variables Hucks, Huck Percentage, Pulls, Offensive Points Played, and Defensive Points Played are excluded, and 2021-23 data where previously excluded variables were included. We also conducted random forest modeling to get variable importance plots (VIPs) and One-Class Support Vector Machines (One-Class SVMs) to ascertain an idea as to which players are outliers in terms of completions of passes compared to hockey assists and receiving yards.

### Results:

Our dissimilarity matrices for our 2017-23 and 2021-23 models for Hierarchical Clustering produced clear differences in the observations by a wide margin using the euclidean dissimilarity method, with the 2021-23 model having more dissimilarity. When creating our models, Ward's Minimum Variance method with AGNES clustering was used given that it had the highest agglomerative coefficient of 0.9922 and 0.9949 for our 2017-23 and 2021-23 models respectively. Using the Elbow and Silhouette methods, we confirmed that two clusters were optimal for both models. The output generated from those two models gave us more players in a "creative sense" (assists, throwing yards, etc.), and more players in a "scoring" sense (goals, receiving yards, etc.), and these represent the handlers and cutters in Ultimate Frisbee. Our results for our random forest models to get VIPs and One-Class SVMs to show outlying points for players and teams are as follows: Both the 2017-23 and 2021-23 models have minutes played and offensive possessions as the most important variables for players, with variation below those two. The player who stood out the most from these analyses was Pawel Janas of the Los Angeles Aviators, making him arguably the best player in the league based on completions comparative to receiving yards and hockey assists. The team data gave us VIPs for break percentage and scores against for both models as the most important variables. The teams that stood out the most were the 2023 Atlanta Hustle and the 2021 DC Breeze based on points scored against versus opponent points off turnovers.

### Conclusions:

Our hierarchical clustering models produced two clusters for each model that reflect the two main positions, more or less, in Ultimate Frisbee. For players, the most important stats were found to be completions, assists, hockey assists, and receiving yards, where Pawel Janas was the best player based on these. Our team results having points scored against and break percentage as the main variables were not as strong due to the undefeated for the past three years, New York Empire, were not present within the data set for teams.