Baseball Analytics Projects Abstracts

Project #1: Creating a Frontier League Player Projection System
For: Yale University S&DS 425 (Statistical Case Studies) Semester-Long Capstone Project

Abstract: This project is developing two models using Yakkertech tracking data of the Frontier League, in conjunction with supplemental Statcast data from the Florida State League (FSL). The first model is a logistic regression designed to determine the probability that players in the Frontier League will advance to higher levels of affiliated baseball leagues—Single-A, Double-A, Triple-A, or even Major League Baseball. The second model will project expected performance statistics for both pitchers and hitters at those higher levels. By analyzing data such as exit velocity, launch angle, pitch velocity, and spin rate, this approach uniquely integrates independent league data with historical advanced performance analytics in the FSL to provide new insights into player development and potential career trajectories.

Project #2: Beyond the Glove: Quantifying the Hidden Value of Catchers in Major League Baseball
For: Yale School of Management MGT 817 (Sports Analytics) Semester-Long Project

Abstract: This study quantifies the hidden value of Major League Baseball (MLB) catchers beyond traditional defensive metrics such as framing, blocking, and throwing. Utilizing pitch-by-pitch Statcast data and advanced statistical methods, including logistic regression, mixed-effects models, and Bayesian hierarchical models, we evaluate latent traits like game-calling and pitcher-catcher synergy. Our findings reveal year-over-year consistency in catchers' latent contributions to strikeouts and their influence on metrics like xFIP-, highlighting traits not captured by standard statistics. Additionally, team context and organizational strategies play critical roles in shaping catcher performance, with results suggesting that advanced analytics and game-planning can enhance or diminish a catcher's value. By bridging gaps in existing research and quantifying unmeasured attributes, this work provides a comprehensive framework for evaluating catchers' contributions, emphasizing the interplay between individual skills and team dynamics. These insights have significant implications for player evaluation, compensation, and roster construction in modern baseball.

Project #3: Data Challenge: Exploring Pitcher-Batter Dynamics Using New Bat Speed and Swing Length Data
For: Connecticut Sports Analytics Symposium (April 11-12 2025)

Abstract: This project aims to analyze newly available Major League Baseball data on bat speed and swing length to explore various aspects of the pitcher-batter interaction during an at-bat. Utilizing pitch-level data from Baseball Savant covering 346,250 plate appearances between April 2, 2024, and June 30, 2024, the study will incorporate relevant Statcast metrics alongside the new swing measurements. The objective is to investigate how bat speed and swing length

influence batter performance, pitcher strategies, and their mutual interactions. Potential analyses include examining the relationship between swing characteristics and plate discipline, assessing how these metrics affect a batter's decision to swing, and determining whether pitchers adjust their pitching approach based on a batter's swing profile. By tapping into this new data, the project hopes to uncover fresh insights into the game, helping us better understand player behavior and decision-making on the field.