

Fall 2024 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: Quantifying a Catcher's Impact on Pitcher's Performance Over a Game using Bayesian Item Response Theory

For: Yale School of Management MGT 817 (Sports Analytics) Semester-Long Project

Abstract: This project introduces a novel application of Bayesian Item Response Theory to develop a metric that quantifies a catcher's impact on run prevention beyond traditional measures like framing, stealing, and blocking. While the multifaceted influence of catchers on pitching outcomes has been acknowledged, recent literature lacks a comprehensive assessment of their contributions in areas such as game-calling, pitch sequencing, and the dynamics of pitcher-catcher relationships. Traditional statistics often overlook these critical factors that affect pitcher performance and runs allowed. By utilizing a Bayesian response model, we aim to provide a more holistic evaluation of a catcher's influence on the game. The model predicts the expected number of runs a pitcher is likely to allow based on variables such as the opposing team's offensive capabilities, the pitcher's historical performance, and contextual game factors. By comparing these expected runs to the actual runs allowed when different catchers are behind the plate, the metric isolates the catcher's latent effect on run prevention. This approach seeks to deepen the understanding of a catcher's role by capturing their intangible contributions, aiding teams in optimizing pitcher-catcher pairings and enhancing overall performance.

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.