Group 6
Project 4
05/28/2025

Predicting a “Good Hitter” Using MLB Statcast Data (2023-2025)

The Story

In 2023 Major League Baseball **(MLB)** installed a new camera system in all 30 of its ballparks to better track the biomechanical markers of its batters. The point was for clubs and fans to better understand what made a particular hitter “good”. In May 2025 they released further internal data metrics aimed at breaking down the physical characteristics of each batters swing. These include *attack angle, attack direction,* and *swing path **(tilt)***.

Our goal is to see if we build a model that can correctly predict a hitters offensive production just by using physical characteristic data

For more information on advanced batting statistics please see the following article https://technology.mlblogs.com/introducing-statcast-2023-high-frame-rate-bat-and-biomechanics-tracking-3844890264a6

The Data

To run our models we will be getting data from https://baseballsavant.mlb.com/ (MLB’s public domain website for advanced statistics). The time period for the data will be from Opening day 2023 to Current Date **(5/27)** 2025. Using baseball savants custom leaderboards feature we will begin by generating a csv of all batters with at least 10 plate appearances (grouped by years) and a mixture of results stats (i.e. hits, homeruns, ops, batting average) Expected stats (barrels, xWOBA, xBA) and physical stats (i.e. bat speed, exit velocity, swing path, attack angle)

The Models

To determine the value of the new swing path statistics we want to use the a Linear regression model to predict the **OPS** of player based on physical characteristics

Other Tools

Group 6
Project 4
05/28/2025

1. Matplotlib
    a. Strike zone heat map for wOBA by swing type (i.e. swing path, tilt. ect.)?
2. Pandas
    a. Clean data.
3. Tableau
    a. Visualization of the team created.
    b. Tell a story.