ISyE 6740 – Spring 2025
Final Report

Team Member Names: Cooper Moore

Project Title: Baseball Swing Path Analysis

Problem Statement

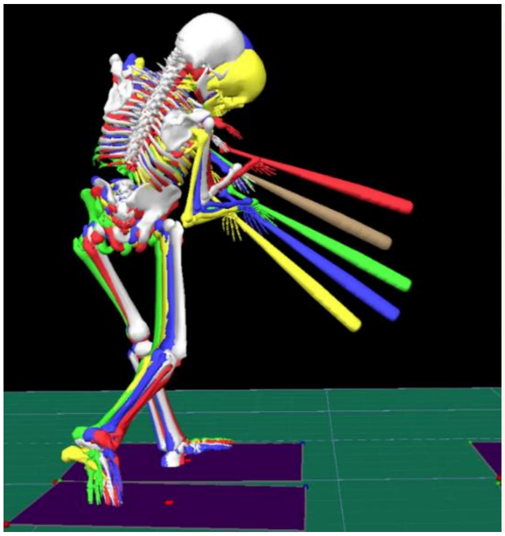
In the modern era of baseball, which is characterized by a data-driven revolution and an increasing emphasis on launch angle, exit velocity, and pitch design, optimizing offensive performance remains a constant pursuit. While traditional metrics offer a broad overview of player capabilities, they often fail to capture the nuanced interaction between a hitter's swing mechanics and the specific shapes of pitches faced. This research aims to bridge this gap by leveraging advanced metrics, Vertical Bat Angle (VBA) and Attack Angle (AA), to investigate the relationship between hitter swing paths and performance against specific pitch types. Specifically, this study addresses the following questions:

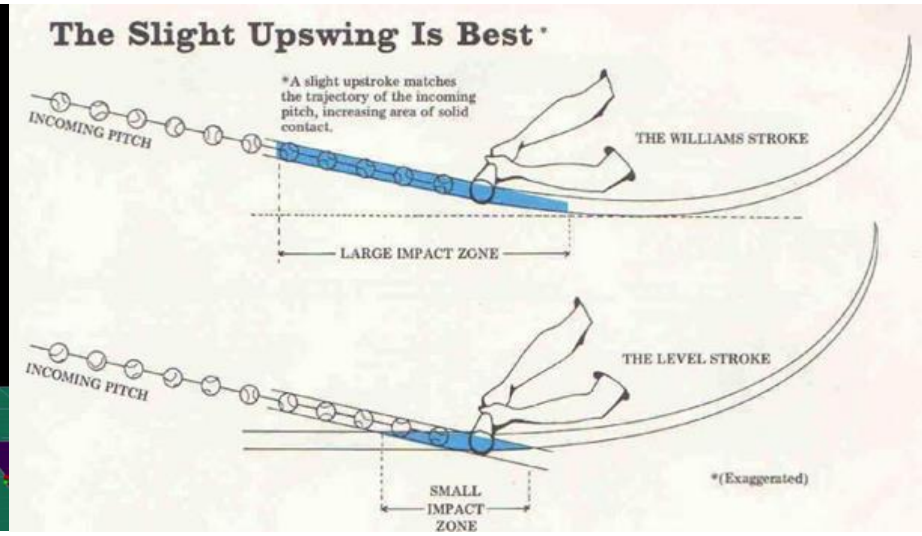
- Can hitters be effectively clustered into distinct groups based on their VBA and AA profiles, revealing underlying swing path archetypes?
- Do these swing path clusters exhibit statistically significant differences in performance against specific pitch types, such as four-seam fastballs high in the zone, sinkers running in on the hands, and curveballs with steep vertical movement?
- Can the performance differences observed be further refined by clustering pitchers based on their repertoire i.e. pitch usage enabling granular matchup analysis?

The hypothesis is hitters with swing paths optimized for specific pitch characteristics will demonstrate superior performance against those pitches. For example, hitters with flatter swing paths (lower VBA and AA) may exhibit higher batting averages, slugging percentages, and exit velocities against high-velocity, high-spin four-seam fastballs. This research seeks to provide empirical evidence supporting or refuting this hypothesis, offering valuable insights for player development, game strategy, scouting, in-game adjustments, and sports betting.

A hitter's VBA varies from swing to swing as hitting a low pitch requires a steeper bat angle (Jones). The metric is stable year by year where a player's mean changes on average by only 1 degree. A player's average VBA ranges from 20 to 40 degrees. The first diagram at the end of this section shows a small angle in the red bat and a large angle in the yellow bat.

Attack Angle measures the bat path's angle relative to the ground. 0 degrees would mean the bat is moving parallel to the ground when the bat strikes the ball (What). Positive angle indicates the bat is being swung upward, which correlates with better performance. Professionals average 6-14 degrees of attack angle which lines up with the angle the ball arrives (Blast). The image below shows the difference between a flat swing and an upward swing.





Data Sources

- Swing Graphs: VBA and AA data, which are not publicly available, is acquired via membership with Swing Graphs. This source provides detailed swing path metrics for professional hitters (Swing). Data volume encompasses multiple seasons, providing a substantial dataset for analysis.
- Baseball Savant: Pitch data, including pitch usage rate, velocity, horizontal movement, vertical movement, spin rate, and pitch type, will be obtained from Baseball Savant. This platform also provides searchable play-by-play data, enabling the linking of swing metrics to specific pitch outcomes. Each search contains over 100 fields for every pitch (Statcast).
- FanGraphs: Traditional performance metrics, such as batting average, slugging percentage, strikeout rate, expected batting average (xBA), expected slugging percentage (xSLG), and xwOBA, is pulled from FanGraphs (Chamberlain). xwOBAcon is a particularly useful stat because it only includes balls in play and takes the expected outcome based on how well the ball was hit to reduce noise (Richards). It stabilizes significantly faster than most stats because it ignores defensive performance, weather, and park factors.
- Data Integration: Data from these sources will be integrated using player IDs and game dates as common keys. Some of this is already be implemented in the pybaseball (Python) package (Jldbc).

Methodology

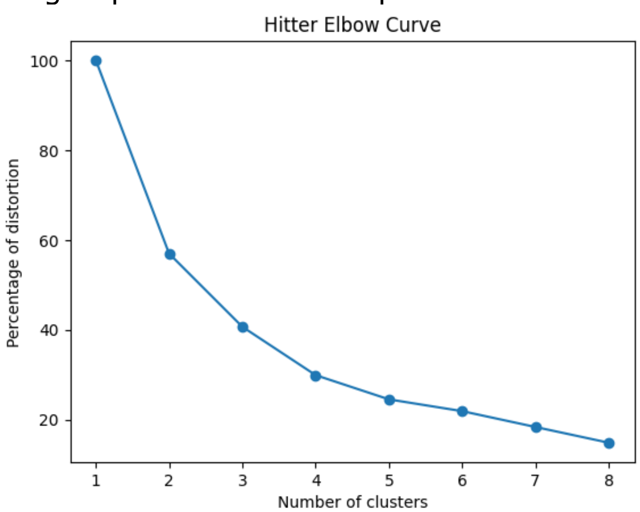
- Clustering:
  - K-means clustering is used to group hitters based on their similarity in swing paths. This is chosen for its computational efficiency and suitability for identifying distinct groups. The elbow method is used to determine the optimal number of clusters. Prior to clustering, features are standardized using z-score normalization. The Euclidean distance metric is used to determine similarity. The data was roughly normal with no outliers.
  - Baseball Savant clusters pitchers based on their results i.e. walks, strikeouts, and hard-hit balls allowed. Most open-source clusters of pitchers are determined by analyzing a pitcher's velocity, movement, and release. This project instead uses a

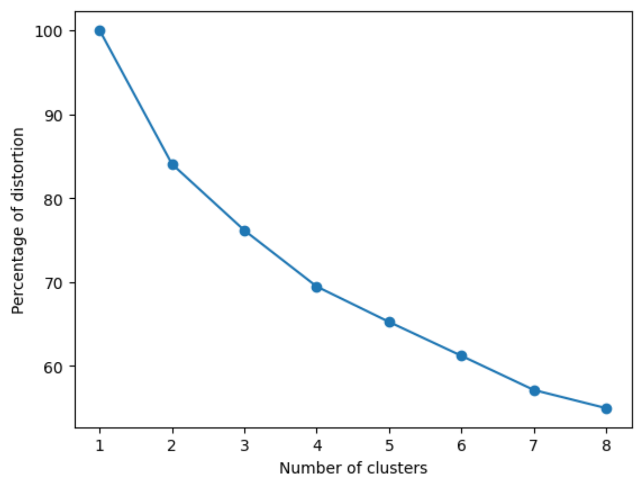
pitcher’s zone location to group players based on how frequently they pitch to each portion of the strike zone. This criterion was chosen because the first portion of this project suggests zone location is more responsible for batted ball quality compared to pitch movement and velocity. The number of clusters was also decided by the elbow method.

- Statistical Testing:
    - Mean performance metrics i.e. xwOBAcon is calculated for each player against each pitch type and pitch location. This is compared to each player’s season average xwOBAcon and aggregated with all members of each cluster to determine if the cluster performs better (or worse) against a specific pitch type compared to all pitches.
    - ANOVA is used to compare the cluster performances. It is a method of analyzing the means and variances of the different clusters to determine if they are statistically different. This study uses an alpha level of .05 to test for statistical significance.

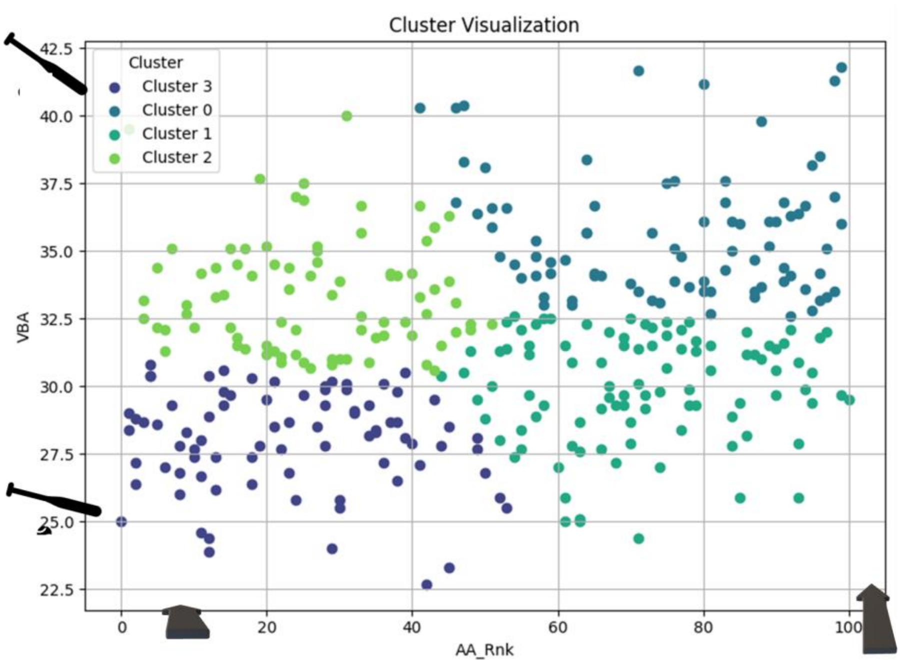
Evaluation and Final Results

- Cluster Quality
    - The quality of clusters is assessed using metrics such as silhouette score and Davies-Bouldin Index. The scatter plot of VBA vs. AA is also useful to determine the cluster quality graphically. Hitters were split into 4 clusters, and pitchers were grouped into 6. Elbow plots and results are below:



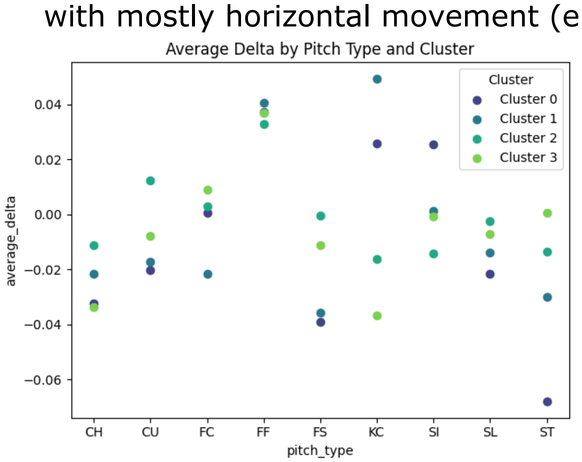


	Silhouette	Davies-Bouldin
Hitter Clusters	.067	3.824
Pitcher Clusters	.165	1.2



The clusters are not very compact or distinct. It may have been better to filter out the center points on the plot where AA and VBA are near the mean. This would have led to more distinct clusters and possibly different results. The pitching clusters were better but still had relatively poor metrics.

- Statistical Significance
  - To determine is hitter clusters outperformed against specific pitches, zones, or pitchers, I took the average xwOBAcon for each player and calculated the delta between the xwOBAcon for each play, and the player's season average. I then performed Anova to see if each cluster had a statistically significant difference.
  - Comparing clusters against specific pitch types, only Sweepers and Sinkers had a p value below .05 (.004 and .007 respectively). Cluster 0 performed much better against sinkers and worse against sweepers. This makes sense because cluster 0 has a steep upward swing with high VBA which would be better at hitting low pitches with a lot of vertical movement (e.g. sinkers) and struggle against pitches with mostly horizontal movement (e.g. sweepers).



Comparing batter cluster performance against pitch locations showed cluster 0 was much better against low pitches than high pitches while cluster 3 was the opposite. This was expected, but only 3 of the pitch locations showed significance.

f_statistic	p_value	zone
3.053447	0.027905	high
4.384042	0.004508	low
1.556753	0.201239	low&away
0.496019	0.685451	low&in
0.503579	0.679884	middle
3.108054	0.030735	up&away
1.143433	0.333038	up&in

When doing two-way Anova between hitter clusters and pitcher clusters, only one pair showed a statistically significant difference, but the difference was so small it was inconsequential. This is likely due to having poor clusters.

- Practical Implications
  - The findings of this study are not enough to inform strategic decisions for teams, such as optimizing lineups or bullpen usage, but it could potentially help at a player development level. Hitters should understand their swing shapes and what pitches they can best hit. This could help the batters make better swing decision by swinging at fewer pitches where they are more likely to underperform and swinging more at pitches where they can do more damage.
- Limitations
  - This research has many limitations which can be improved upon.
    - Clusters were not very distinct. Focusing more on edge-cases and filtering out players with average swings could lead to better results.
    - Data used was from the 2023 season. Looking at multiple years would increase the sample size and confirm if the results stay the same year to year.
    - Only balls in play were considered. A lot of a batter’s value comes from walks and strikeouts which were not included.
    - Swings are dynamic. More research can be done to see if VBA and AA have much variability within each game.

References

Blast connect: What should attack angle be?. Blast Connect | What should Attack Angle be? (n.d.). https://blastconnect.com/training_center/item/209

Chamberlain, A. (2019, July 9). *Pitch Type xwOBA on contact (xwobacon)*. RotoGraphs Fantasy Baseball. https://fantasy.fangraphs.com/pitch-type-xwoba-on-contact-xwobacon/

Jldbc. (n.d.). *Pybaseball /pybaseball at master · JLDBC/Pybaseball*. GitHub. https://github.com/jldbc/pybaseball/tree/master/pybaseball

Jones, B. (2022, December 1). *Hitting biomechanics*. Driveline Baseball. https://www.drivelinebaseball.com/2022/12/hitting-biomechanics-barrel-direction/

Richards, D., & Richards, D. (2019, April 19). *Going Deep: The Real Value of Statcast Data Part I*. Pitcher List. https://pitcherlist.com/going-deep-the-real-value-of-statcast-data-part-i/#:~:text=Fourth%2C%20though%20not%20reflected%20in,(approximately%2018%20games%20played).

Statcast search CSV documentation. baseballsavant.com. (n.d.). https://baseballsavant.mlb.com/csv-docs#estimated_woba_using_speedangle

SwingGraphs Member Zone – VBA Data. Swinggraphs. (n.d.). https://swinggraphs.com/vbadata/

What is attack angle? - baseball connect. Baseball Connect - Educating and connecting people through the game of baseball. (2024, September 19). https://baseball-connect.com/learn/what-is-attack-angle/