Momentum in the MLB: A Data-Driven Analysis (2020–2024)

By: Anthony Fernandez

This project set out to examine whether “momentum” exists in Major League Baseball by measuring its impact on a hitter’s performance. Specifically, the focus was to see if a batter is more likely to get a hit following a hit or sacrifice fly by the previous batter in the same inning. Using Statcast data from the 2020 through 2024 regular seasons, I built a pitch-by-pitch dataset for each year and derived a momentum variable for every plate appearance. My code flagged momentum events whenever the preceding batter in the inning recorded a single, double, triple, home run, or sacrifice fly. These events were chosen because they are the most direct indicators of offensive success and would reasonably be expected to create psychological or game-state “momentum” that might carry over to the next hitter.

To ensure accurate comparisons, I carefully filtered the data. I excluded plate appearances that were not official (such as walks, hit-by-pitches, or errors) and isolated only the outcomes that directly count toward batting average…hits and outs. This allowed for a clean binary hit/no-hit outcome. I also filtered out any leadoff hitters in each inning, as there was no prior batter to generate momentum. Without a preceding plate appearance, there’s no valid momentum state to evaluate. This step ensured that each data point reflected an environment where momentum could theoretically exist, improving the validity of the comparison. By focusing on within-inning sequences, I avoided potential biases from inning or game-level resets that might obscure short-term momentum patterns.

After building clean datasets for each season, I created summaries of batting averages in two scenarios: with momentum (the previous batter got a hit or sac fly) and without momentum. The differences were then aggregated at the league level and broken down by team. For each season, I performed a two-sample Welch t-test to statistically evaluate whether batting average differences between momentum and non-momentum plate appearances were significant. Welch’s t-test was used because it does not assume equal variance between groups and is more robust when comparing two sample means from datasets with different sizes, which applies well to real world sports data.

The results varied by season, but the presence of a “momentum effect” was supported more often than not. In 2020, the difference was negligible and statistically insignificant. The average batting average without momentum was .2468 versus .2478 with momentum ($p = 0.9027$). However, in 2021, the effect became more noticeable. Batters hit .243 on average without momentum and .255 with momentum, a statistically significant difference ($p = 0.018$). This pattern persisted in subsequent seasons. In 2022, the difference was again significant ($p = 1.28e-10$), with batters hitting .242 without momentum and .260 with

momentum. Similarly, 2023 and 2024 continued this trend. The 2024 t-test produced a p-value of 4.46e-08, confirming a highly significant difference in average performance depending on momentum state. The average difference in batting average between momentum and non-momentum situations often ranged from .01 to .02, which, while not enormous, is meaningful over the course of thousands of plate appearances.

Beyond league-wide analysis, I aggregated momentum performance at the team level for each season. I grouped each team's batting average with and without momentum and calculated the difference. This provided a more granular look at which teams may have excelled at capitalizing on momentum across seasons. For example, in 2020, the Tigers showed one of the largest momentum boosts, with a nearly .08 higher batting average following a momentum event. In contrast, a few teams like the Phillies and Red Sox saw marginal or even slightly negative effects. These year-by-year comparisons were stored in separate CSV files and later combined into a master dataset to facilitate a longitudinal team-level analysis from 2020 through 2024.

To explore patterns in this multi-year momentum performance, I performed a k-means clustering analysis using each team's five-year average momentum batting average difference. I used a five-cluster model to identify team archetypes. Teams were grouped based on how consistently they outperformed (or underperformed) with momentum. One cluster featured teams that consistently showed positive momentum effects, possibly suggesting either lineup construction or team mentality that fosters hitting success in bunches. Another cluster included teams that seemed unaffected by momentum; their batting averages were flat regardless of the previous outcome. A third group even appeared to perform worse in momentum situations, though this could be statistical noise or lineup-related quirks. The remaining clusters had smaller variations, representing teams that hovered around neutral or showed inconsistent patterns across years.

The accuracy and reliability of the models used throughout this project were solid. Each season included over 50,000 relevant plate appearances after cleaning and filtering, ensuring that the t-tests and summary statistics were backed by substantial data. Even with this large sample size, the results remained statistically significant in four of the five seasons, reinforcing the credibility of the findings. I also validated the data by comparing my calculated total number of hits against official MLB statistics for each year. In most cases, the values matched closely, with a minor difference explained by filters (e.g., removal of walks, leadoff hitters, and sacrifice bunts).

One visual highlight of the project was a bar plot that showed each team's difference in batting average with and without momentum for each season. Bars were colored green for positive differences and red for negative ones. This made it intuitive for the viewer to see which teams benefitted from momentum and

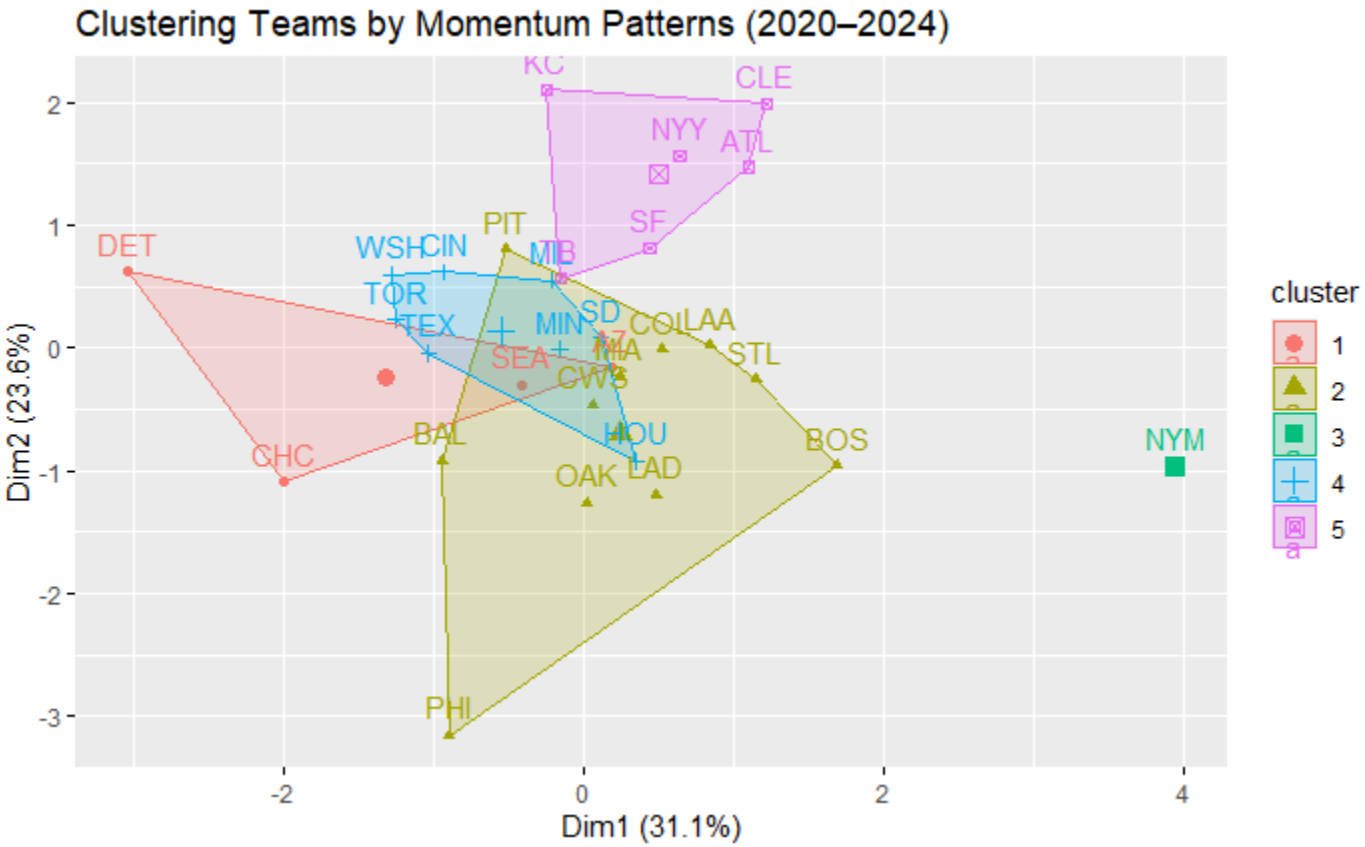
which did not. In 2020 teams like Detroit, Seattle, and Arizona stood out with strong positive momentum effects, while a few others like Red Sox showed slightly worse outcomes in momentum situations. Although this did not guarantee playoff success or even making the playoffs, it was still valuable to see which teams really road the coattails of momentum swinging situations during the regular season. To make the chart more accessible to casual fans, I reworded the axis title from “BA Difference” to “Difference in Batting Average When Previous Batter Got a Hit or Sac Fly,” which improved clarity.

Looking forward, there are several potential improvements to this analysis. One area of future research could involve incorporating pitcher identity and leverage situation to better control for confounding variables. Another possibility is to examine if certain hitters are more momentum-sensitive than others. Perhaps power hitters versus contact hitters, or young players versus veterans. Finally, visualizing momentum trends over time could add storytelling depth to the R-Markdown reports, especially if paired with game outcomes or win probabilities.

In conclusion, my project supports the idea that momentum, at least in a short-term intra-inning sense, has a measurable impact on batting performance in the MLB. Although the size of the effect varies by year and team, the pattern is statistically significant in most seasons between 2020 to 2024. The combination of rigorous data cleaning, thoughtful filtering (including the removal of leadoff hitters), and robust statistical testing provides confidence in the results. Momentum may not be a game-changer in every instance, but across five years of MLB data, it seems to give hitters a slight edge and some teams may be better at riding that wave than others.

All code used for this analysis was written in R-Studio. Code is available upon request.

Appendix



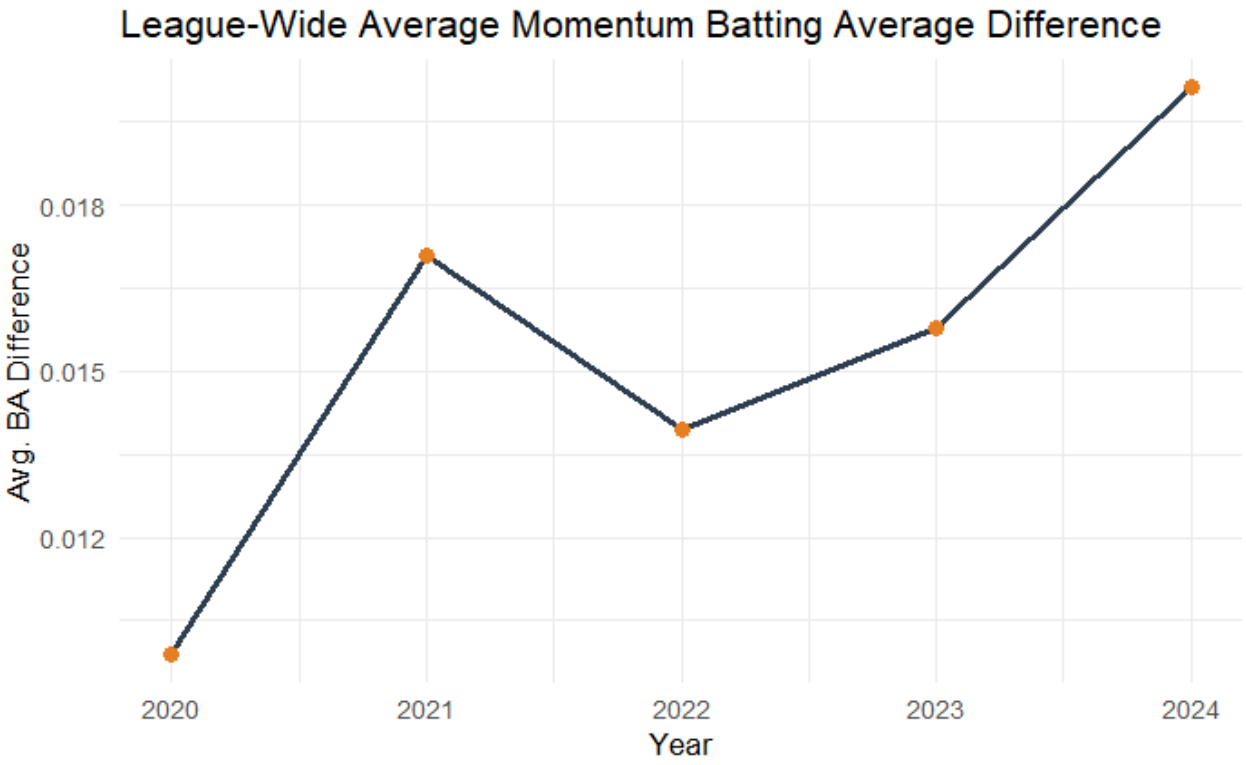
Cluster 1 – Upward Momentum Growers: Slightly negative in 2020, improving every year through 2024.

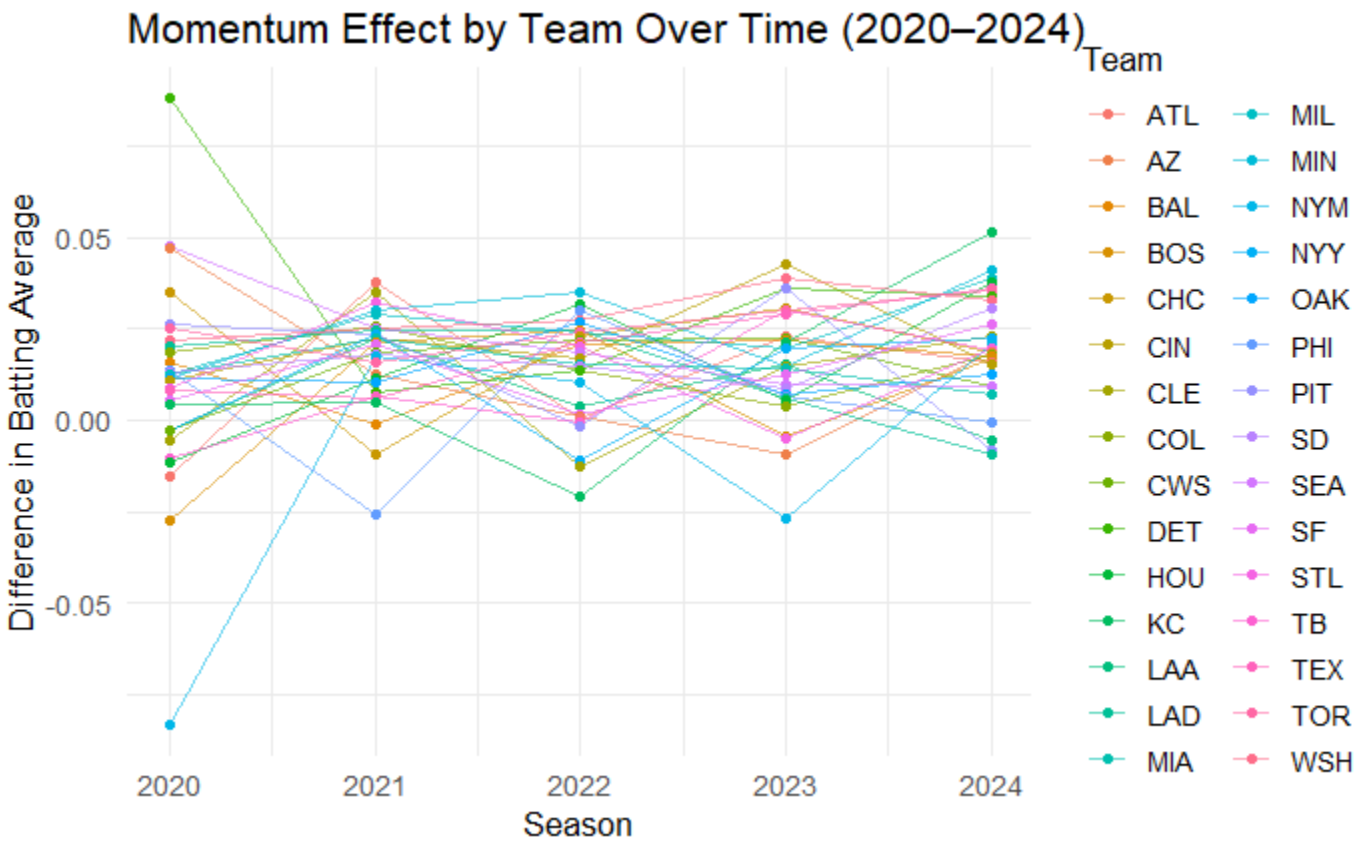
Cluster 2 – Steady, Modest Momentum Teams: Mildly positive and consistent.

Cluster 3 – Inconsistent or Rebounding Momentum Teams: Big dip in 2020, mixed recovery.

Cluster 4 – Elite Momentum Maximizers: Strongest and steadily rising BA differences.

Cluster 5 – Front-Loaded Momentum Teams: Huge 2020 momentum boost, flatter since.


League-Wide Average Momentum Batting Average Difference
Avg. BA Difference
0.018
0.015
0.012
2020
2021
2022
2023
2024
Year


Momentum Effect by Team Over Time (2020–2024)
Difference in Batting Average
0.05
0.00
-0.05
2020
2021
2022
2023
2024
Season
Team
ATL
AZ
BAL
BOS
CHC
CIN
CLE
COL
CWS
DET
HOU
KC
LAA
LAD
MIA
MIL
MIN
NYM
NYY
OAK
PHI
PIT
SD
SEA
SF
STL
TB
TEX
TOR
WSH

Sources

- Baseball Savant. *Statcast Search Tool*. Major League Baseball. https://baseballsavant.mlb.com/statcast_search
    - Used to collect pitch-by-pitch and plate appearance data for MLB regular seasons from 2020 to 2024.
- Major League Baseball. *Official Statistics and Leaderboards*. https://www.mlb.com/stats/
    - Referenced to verify total hits and batting averages for each team and season.
- Petti, B. (2024). *baseballr: R Package for Baseball Data Analysis*. GitHub repository. https://github.com/BillPetti/baseballr