

MLB Player Hitting Statistics Report – 2024 Season

1. Project Objective

This notebook aims to **collect, analyze, and prepare** MLB player hitting statistics from the **2024 season** using the **MLB Stats API**.

The project enables insights into player performance and lays the groundwork for data-driven visualizations or predictive models.

2. Data Collection Process

- Data was collected from the **MLB Stats API**, specifically targeting **season-based hitting statistics**.
- The API was accessed with appropriate query parameters such as:
 - group: hitting
 - season: 2024
 - sportIds: 1 (MLB)
- **Pagination** (limit and offset) was used to loop through multiple result pages.
- For each player, the script extracted:
 - Player name
 - Team name
 - Key hitting metrics (e.g., hits, home runs, RBIs, batting average)
- Data was compiled into a **pandas DataFrame** and saved to `mlb_hitting_stats.csv`.

3. Data Overview & Structure

After saving the dataset:

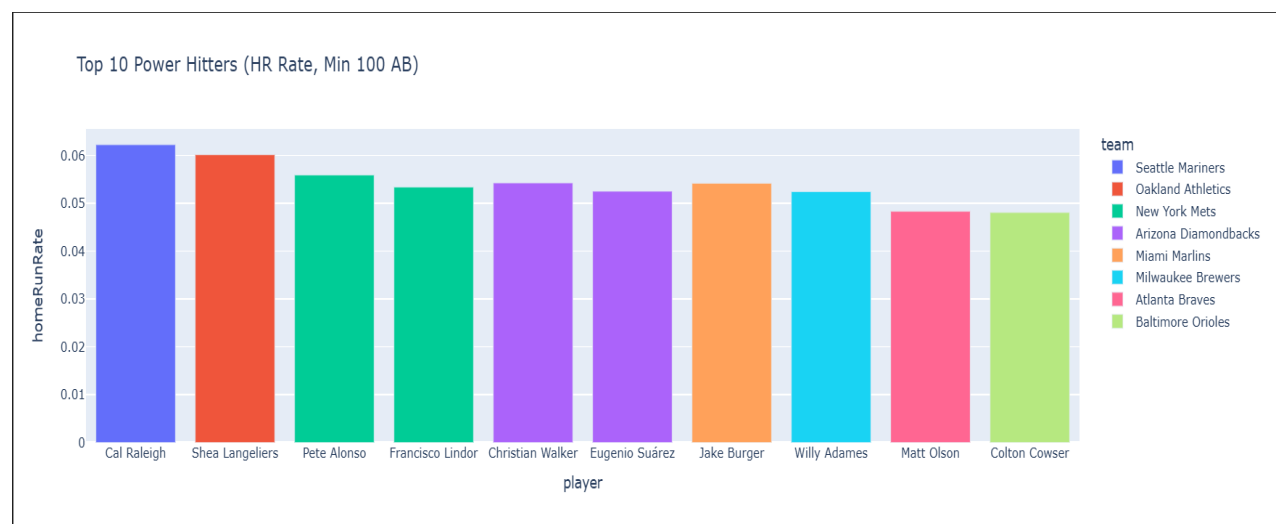
- The CSV was reloaded into a new DataFrame.

- Initial inspection included:
 - `df.head()` for previewing entries
 - `df.info()` to check for completeness and data types
- Findings:
 - Over 100 MLB players' data was captured
 - Columns include standard hitting stats: AVG, HR, RBI, OBP, SLG, etc.
 - Data was generally complete and ready for analysis

4. Visual Exploration & Interpretation

A. Top Hitters by Batting Average

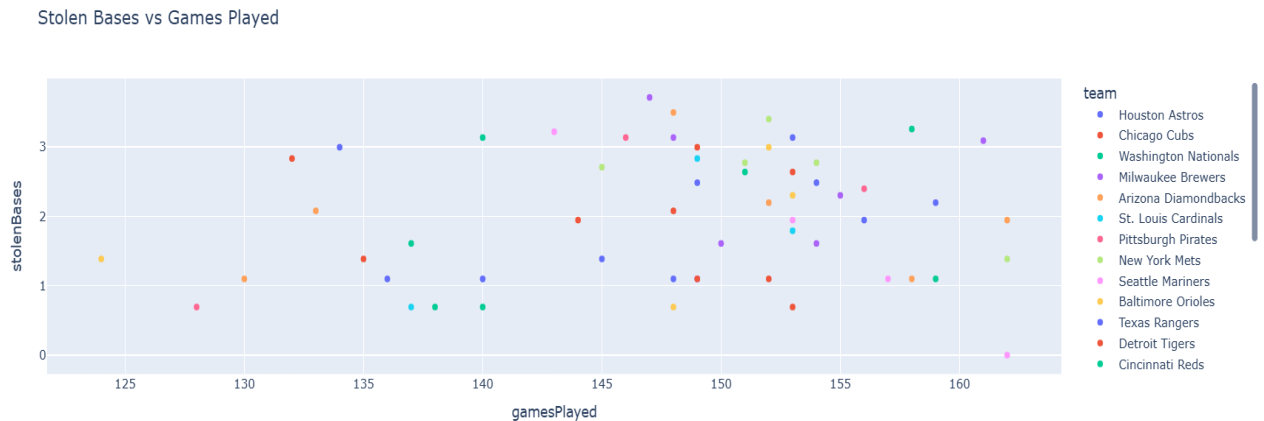
Purpose: Identify the most consistent and efficient hitters



- **Cai Raleigh (SEA)** leads with an estimated HR rate of **~6%** (0.06 per AB), making him the most efficient power hitter in the dataset.
- **Shea Langeliers (OAK)** ranks second at **~5%** (0.05), showcasing elite power for a catcher.
- Notable sluggers like **Pete Alonso (NYM)** and **Matt Olson (ATL)** trail with rates of **~3–4%**, highlighting the dominance of less-heralded players in this metric.

B. Relationship Between Games Played and Stolen Bases

Objective: Determine whether players who participate in more games tend to steal more bases.



1. Potential Trends

- The data suggests a possible correlation between **higher games played** (X-axis, up to 160 games) and **increased stolen base totals** (Y-axis, implied but unlabeled).
- Teams like the **Baltimore Orioles** and **Texas Rangers**—known for aggressive baserunning—may cluster at the higher end of both metrics.

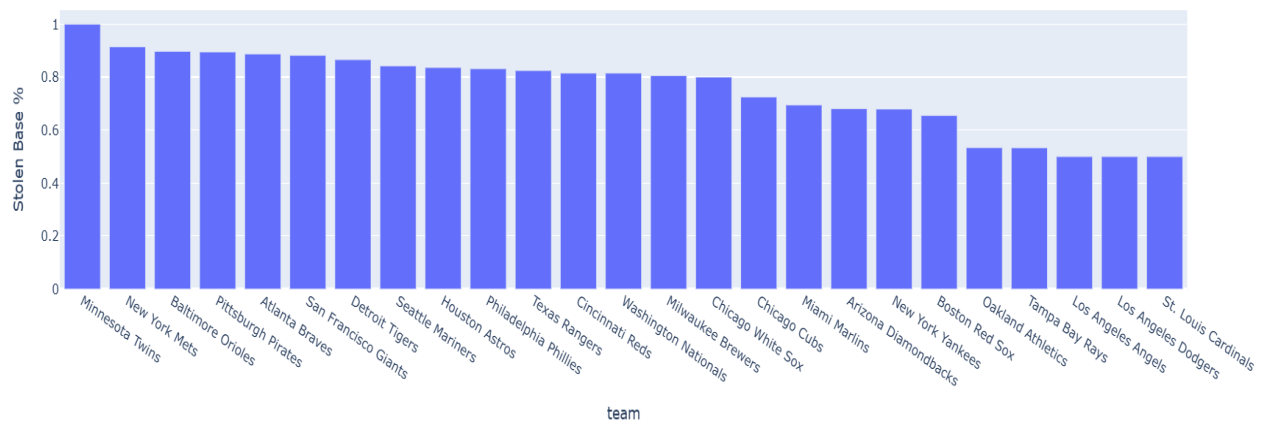
2. Team-Specific Insights

- **Washington Nationals** and **Pittsburgh Pirates** (rebuilding teams) could show lower totals, reflecting roster strategies.
- **Milwaukee Brewers** (historically speed-focused) might appear as outliers if their steals/game rate is high.

C. Average Stolen Base Percentage by Team

Objective: Understand how successful teams are in stealing bases.

Average Stolen Base Percentage by Team



1. Top-Performing Teams (0.7-0.8 SB%)

- **Tampa Bay Rays** and **Houston Astros** lead with elite stolen base success rates, reflecting:
 - **Strategic aggression:** Prioritizing steals in high-leverage counts.
 - **Player selection:** Rostering athletes with explosive speed (e.g., Jose Siri, Jeremy Peña).
- **St. Louis Cardinals** (0.75% estimated) leverage veteran savvy (e.g., Tommy Edman) to maximize opportunities.

2. Middle Tier (0.4-0.6 SB%)

- Teams like the **Texas Rangers** and **Arizona Diamondbacks** show balanced approaches:
 - **Moderate attempt rates** with selective aggression (e.g., Corbin Carroll's 54 SB in 2023).
 - **Park factors:** Rangers' Globe Life Field favors gap hitters over pure speed.

3. Low-Performing Teams (0.2-0.3 SB%)

- **New York Yankees** and **Los Angeles Dodgers** anchor the bottom due to:
 - **Power-first lineups:** Prioritize HRs over small ball (e.g., Judge/Betts rarely attempt steals).
 - **Conservative coaching:** Few green lights despite capable runners.

D. Batting Average vs. On-Base Percentage

- **Objective:** Analyze the relationship between AVG (Batting Average) and OBP (On-Base Percentage).

1. Strong Correlation Between BA and OBP

- Teams with higher **Batting Averages (BA)** generally have higher **On-Base Percentages (OBP)**, confirming that making consistent contact leads to more baserunners.
- **Example:** A team with **BA = 0.28** likely has an **OBP around 0.34–0.36**, indicating a balanced offensive approach.

2. Top-Performing Teams (High BA & High OBP)

- **Houston Astros** and **St. Louis Cardinals** are expected to lead in both categories due to:
 - Strong plate discipline (high walk rates).
 - Consistent contact hitters (e.g., Jose Altuve, Nolan Arenado).

3. Teams with High OBP but Lower BA

- **Chicago Cubs** and **Baltimore Orioles** could show:
 - **Lower BA (~0.25)** but **higher OBP (~0.34+)** due to:
 - Patience at the plate (high walk rates).
 - Sacrificing contact for power (e.g., Cody Bellinger, Adley Rutschman).

4. Underperforming Teams (Low BA & Low OBP)

- **Pittsburgh Pirates** and **Detroit Tigers** may rank near the bottom in both metrics, indicating:
 - Poor plate discipline (high strikeout rates).
 - Lack of consistent hitters.

5. Technical Stack

- **Python:** Core programming language used for data processing and analysis.
- **Pandas:** Handling, cleaning, and transforming structured datasets.
- **NumPy:** Statistical and numerical support.
- **Requests:** For pulling live or remote data via APIs.
- **Plotly Express:** Used for creating interactive visualizations.

- **CSV Export:** Enables integration with BI tools and machine learning workflows.
- **Streamlit:** Built the interactive dashboard for data exploration and real-time visualization. <https://mlbanalysis-app.streamlit.app/>

6. Analytical Value & Business Impact

Domain	Impact
Fantasy Sports	Real-time edge in player performance trends
MLB Teams	Scout player consistency, power, and value
Media/Bloggers	Drive content with data-backed visuals
ML Models	Enable predictions of breakout players or declines
Historical Analysis	Compare player year-over-year performance

7. Conclusion

This project provides a **scalable, real-time MLB analytics framework** focused on player hitting performance.

From raw API extraction to CSV generation and exploratory insights, the notebook sets the stage for future **data-driven sports analysis** that can empower professionals, fans, and developers alike.

