## Introduction

- **Overview:** We conducted an exploratory data analysis on anonymized bat tracking data from multiple Phillies Minor League affiliate games. The data is in JSONL format and has been converted to CSV for analysis.
- **Objective:** To understand pitch types, their effectiveness, and identify missing data patterns.

## Data Description

- **Data Source:** JSONL files containing tracking data for each at-bat.
- **Key Fields:**
    - **Units:** Length, Velocity, Acceleration, Angle
    - **Summary Acts:** Pitch, Hit, Summary Score
    - **Events:** Detailed batted ball descriptions
    - **Samples:** Ball flight, Bat flight

## 1. Missing Values Analysis

The missing values analysis shows that several columns have a significant number of missing values. For example:

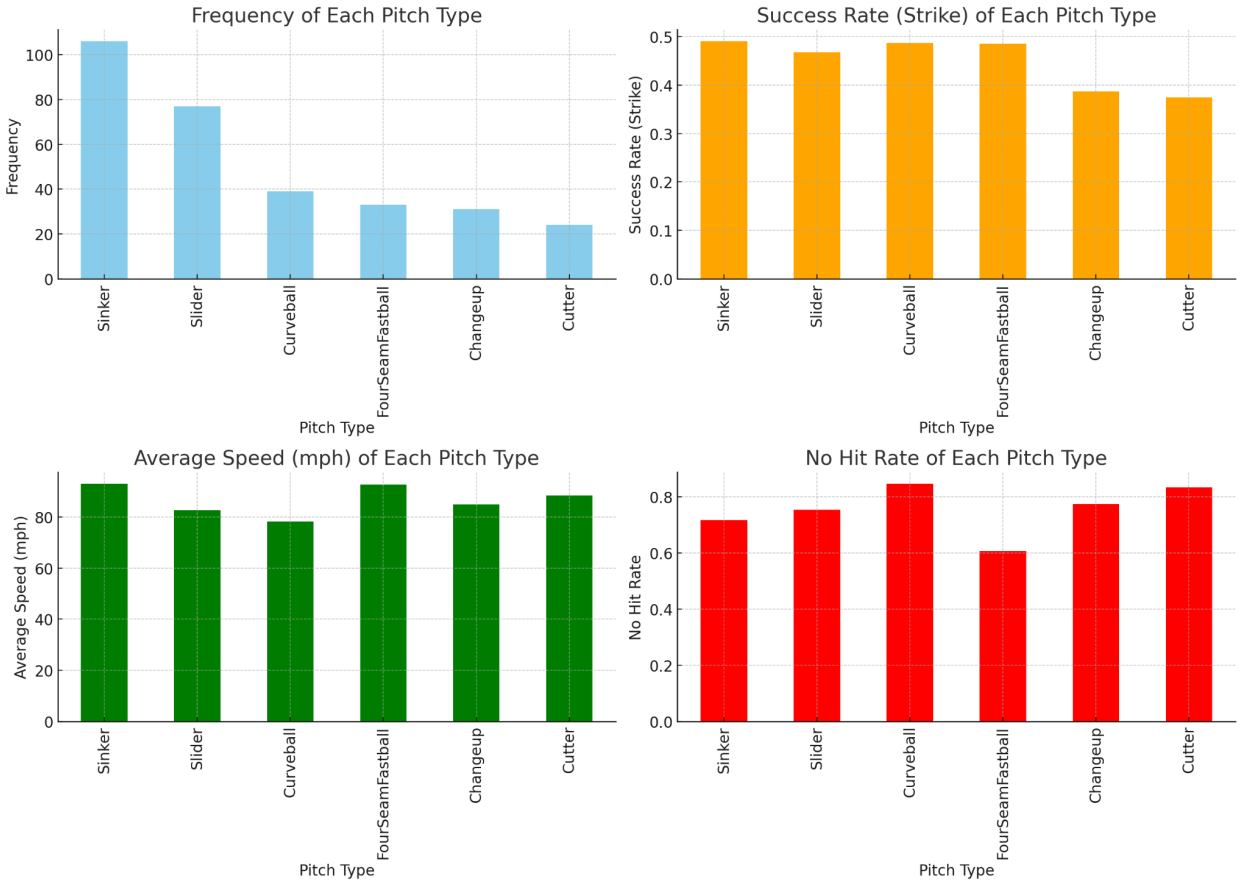
- `pitch_type` has 944 missing values.
- `pitch_action` has 326 missing values.
- Columns related to hit information (`hit_speed_mph`, `hit_speed_kph`, `hit_speed_mps`, `hit_spin_rpm`) have 928 missing values.

## 2. Statistics on Missing Hits and Average Pitch Height

- **Number of Missing Hits**:
    - There are **928 missing hits** in the dataset.
- **Average Height of Pitches**:
    - The average height of pitches (based on initial position data) is approximately **5.52 units** (the exact units depend on the measurement system used in the dataset, which could be feet or meters).

- **Key Insight:** The significant number of missing hits indicates that a substantial portion of the dataset does not include hit information. This might be typical in games where many pitches do not result in hits.

## 3. Pitch Types and Their Effectiveness Analysis




Here is the analysis of different pitch types and their effectiveness:

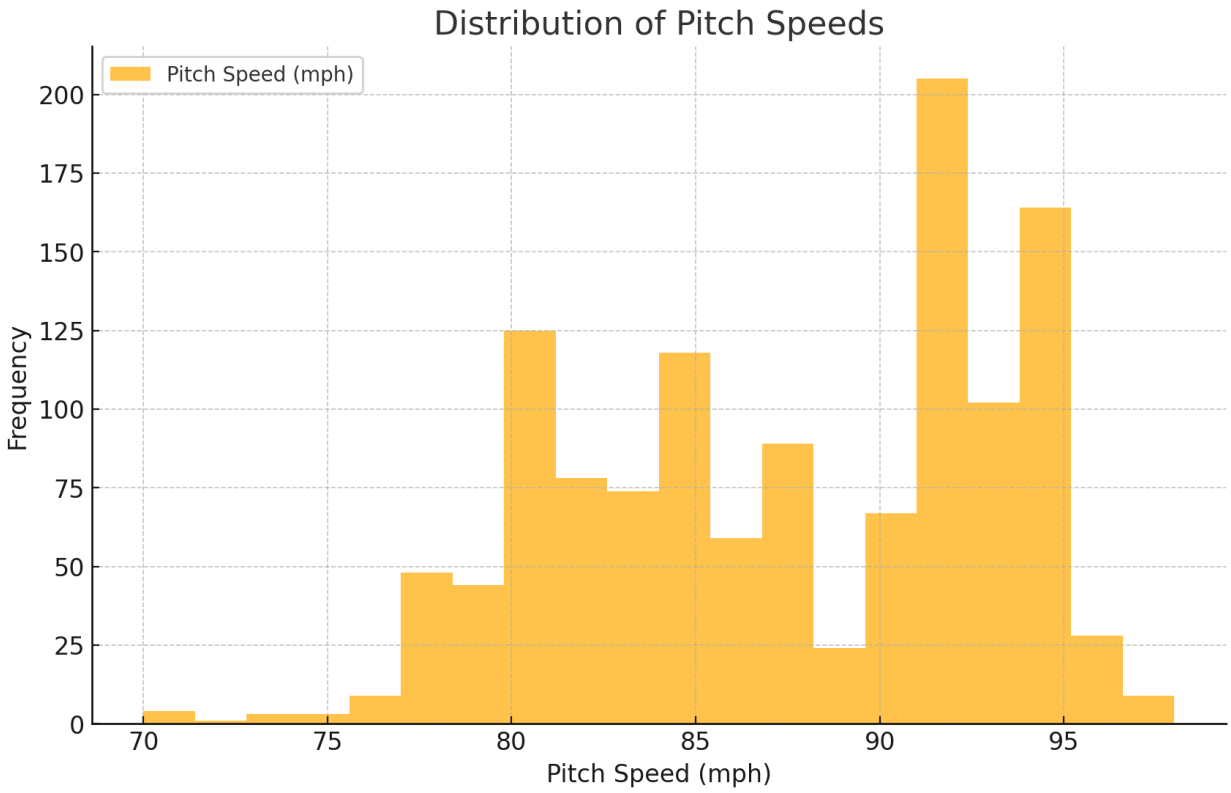
- **Frequency of Each Pitch Type**:
  - **Sinker**: 106 occurrences
  - **Slider**: 77 occurrences
  - **Curveball**: 39 occurrences
  - **FourSeamFastball**: 33 occurrences
  - **Changeup**: 31 occurrences
  - **Cutter**: 24 occurrences
- **Success Rate (Strike)**:
  - **Sinker**: 49.06%
  - **Slider**: 46.75%

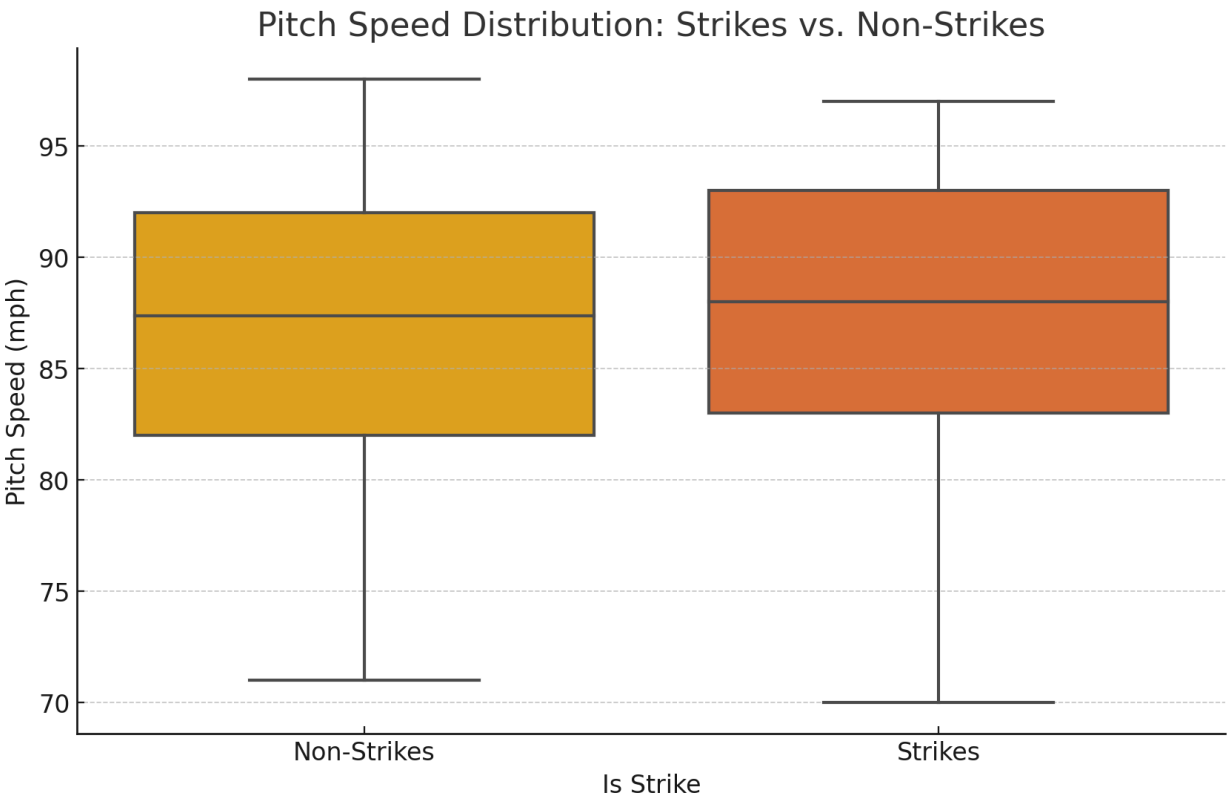
    - **Curveball**: 48.72%
    - **FourSeamFastball**: 48.48%
    - **Changeup**: 38.71%
    - **Cutter**: 37.50%
- **Average Speed (mph)**:
    - **Sinker**: 92.95 mph
    - **Slider**: 82.53 mph
    - **Curveball**: 78.26 mph
    - **FourSeamFastball**: 92.67 mph
    - **Changeup**: 84.81 mph
    - **Cutter**: 88.33 mph
- **Effectiveness in Preventing Hits (No Hit Rate)**:
    - **Sinker**: 71.70%
    - **Slider**: 75.32%
    - **Curveball**: 84.62%
    - **FourSeamFastball**: 60.61%
    - **Changeup**: 77.42%
    - **Cutter**: 83.33%

- **Key Insights:**
    - **Curveball** and **Cutter** have the highest no-hit rates (84.62% and 83.33% respectively), indicating their effectiveness in preventing hits.
    - **Sinker** and **FourSeamFastball** have high average speeds (around 93 mph), but their effectiveness in preventing hits is relatively lower compared to Curveball and Cutter.
    - **Changeup** has the lowest success rate for strikes (38.71%) but a reasonably high no-hit rate (77.42%).

## 4. Pitch Speed vs Strike and No-Strike Analysis

The visualizations indicate the distribution of pitch speeds for strikes and non-strikes and highlight the average speeds. As observed, non-strike pitches have an average speed of approximately 86.59 mph while strike pitches have an average speed of approximately 87.68 mph. As a result, there is no strong correlation between pitch speed and strike or no strike.


Distribution of Pitch Speeds
Pitch Speed (mph)
Frequency
0
25
50
75
100
125
150
175
200
70
75
80
85
90
95
Pitch Speed (mph)


Pitch Speed Distribution: Strikes vs. Non-Strikes
Pitch Speed (mph)
70
75
80
85
90
95
Non-Strikes
Strikes
Is Strike