Leveraging Machine Learning to Analyze Player Injury Return Dates and Predict Tommy John Surgery in Pitchers

THIERNO DIALLO
JENNIFER MCNEW

ABIGAIL PARSLEY
KAYLI SMITH

JOSH STILL
SYLVIA TURNER

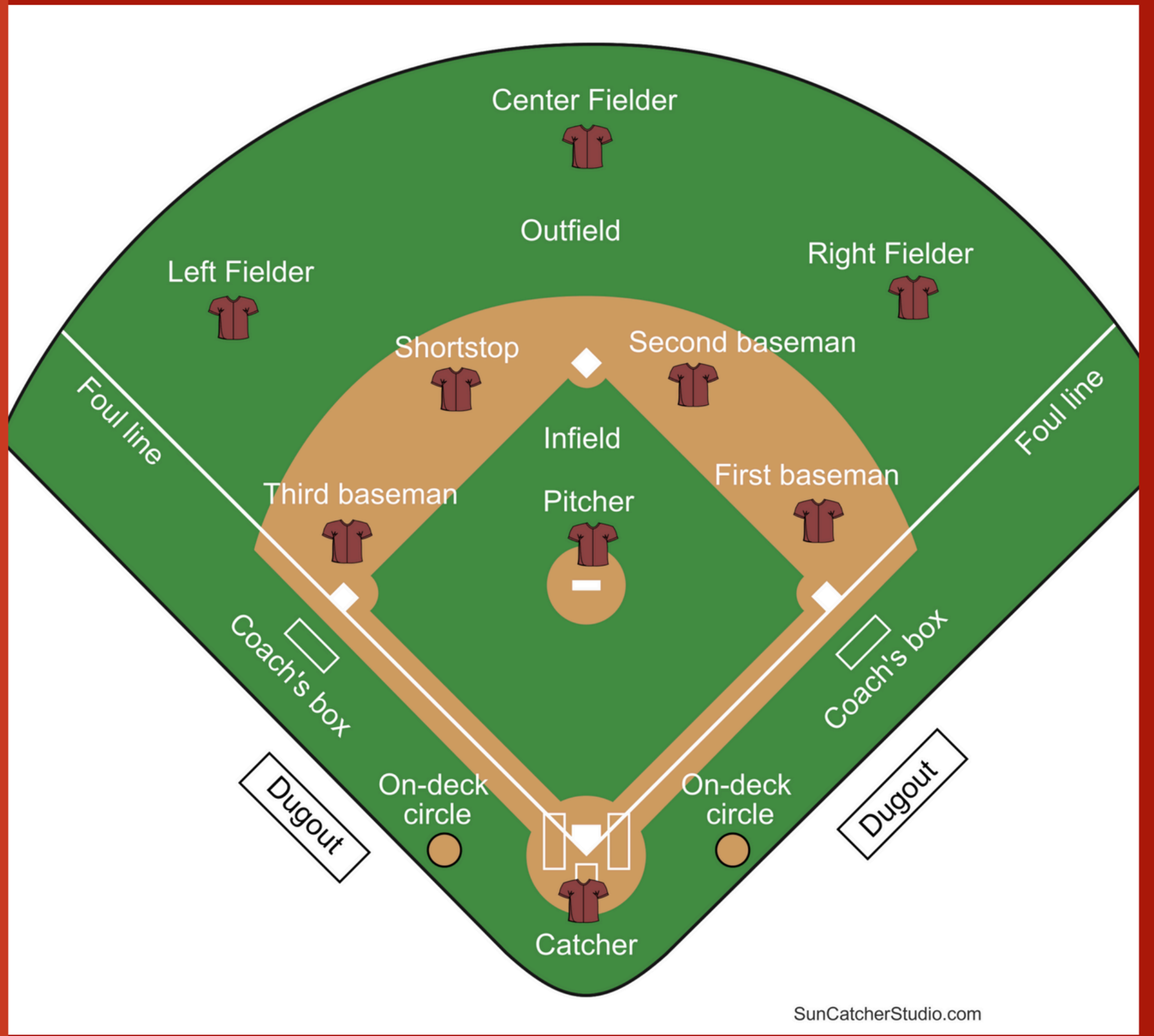
LISA YBARRA
JACK YEAGER

Project Overview and Objectives

Our focus is on two main areas: evaluating the accuracy of injury return dates and predicting Tommy John surgeries from pitch counts in baseball

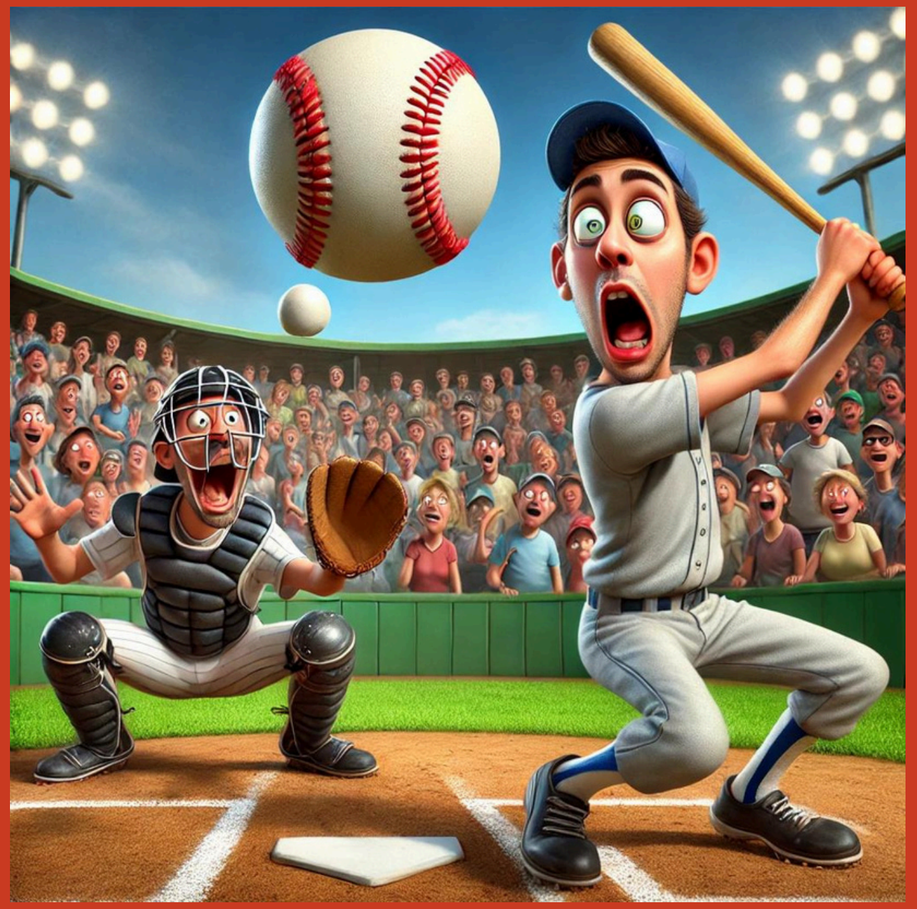
Baseball Basics

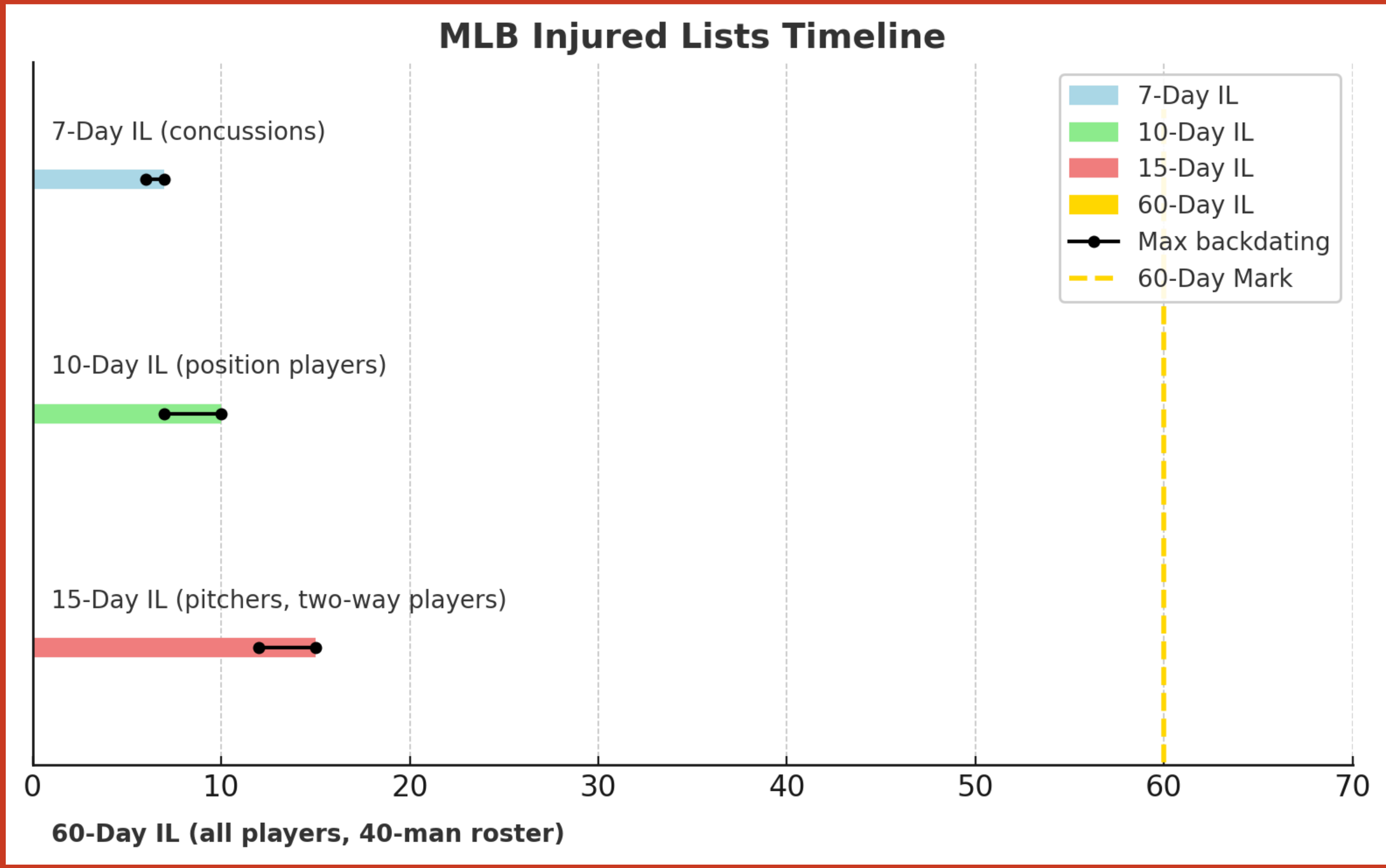
- Two teams alternate between offense and defense.
- Offense: Hits the ball and runs to four bases to score runs.
- Defense: Aims to get offensive players out and prevent runs.
- Game consists of nine innings.
- The team with the most runs at the end wins.
- Pitchers play a key role by throwing the ball to batters.
- Hitters try to score runs by successfully hitting the ball.

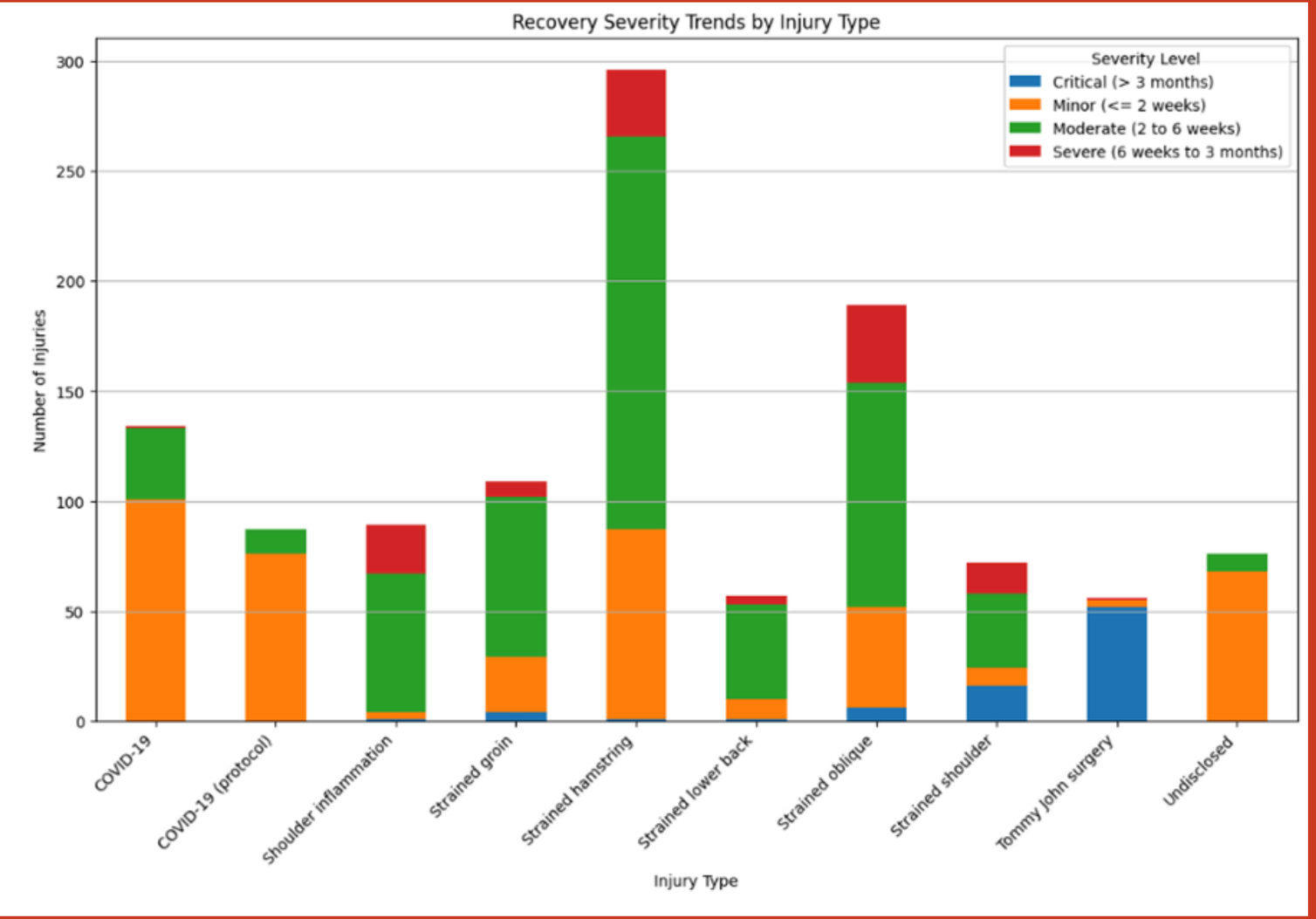


Key Facts

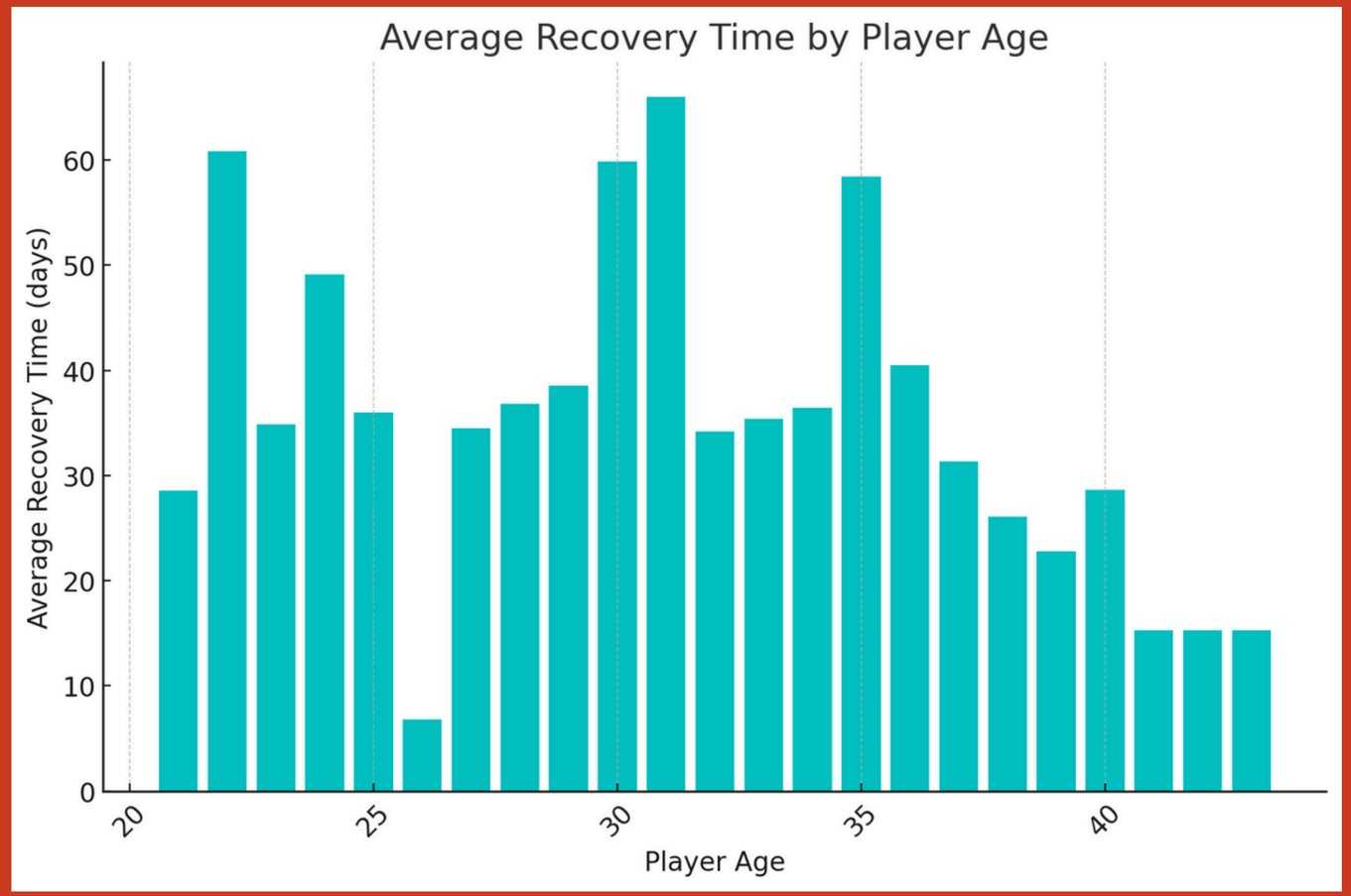
- Baseball: 9-inch leather-wrapped ball.
- Game Length: 9 innings; extra innings if tied.
- Teams: 9 players on defense, 30 MLB teams.
- Batting Average: .250 average, 25% hit rate.
- Strikeout Rate: 22% average.
- Home Runs: 1.2 per game (MLB 2023 average).



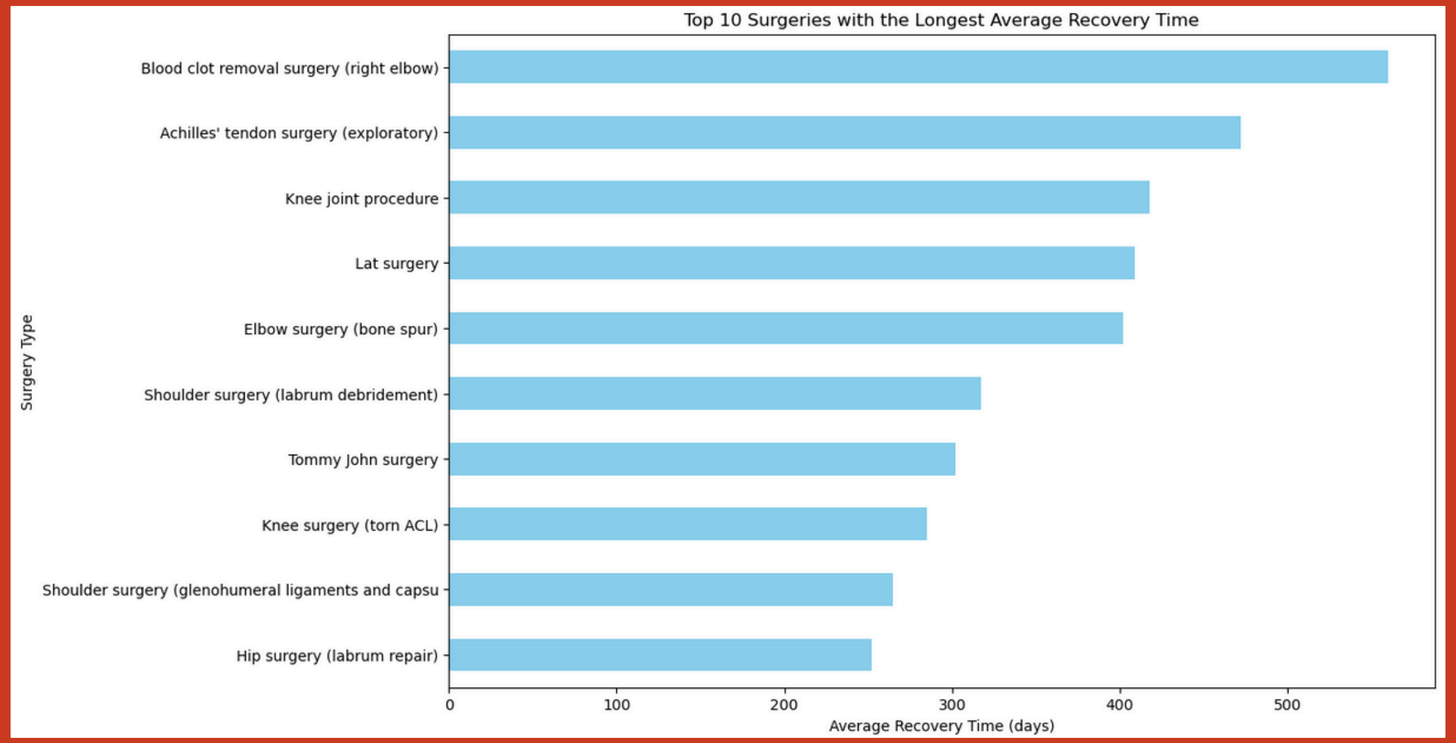

MLB Injured Lists Timeline
7-Day IL
10-Day IL
15-Day IL
60-Day IL
Max backdating
60-Day Mark
7-Day IL (concussions)
10-Day IL (position players)
15-Day IL (pitchers, two-way players)
0
10
20
30
40
50
60
70
60-Day IL (all players, 40-man roster)



Recovery Severity Trends by Injury Type



Recovery Time By Player Age



Return From Injury

- Target Variable: 'Days to Return' (days to return after treatment or surgery)
- Model Goal: Predict 'Days to Return' using dataset features
- Performance Evaluation: Based on accuracy and other metrics
- Model Accuracy: Achieved ~82% accuracy in predicting return time

Injury Breakdown

Strained Hamstring

The most frequently reported injury in baseball, occurring at a rate of 19.7% among players. It highlights the physical demand and risk of muscle injuries in the sport.

Elbow Inflamation

An injury that affects player performance, reported at a rate of 7%, showcasing the importance of arm health in baseball activities.

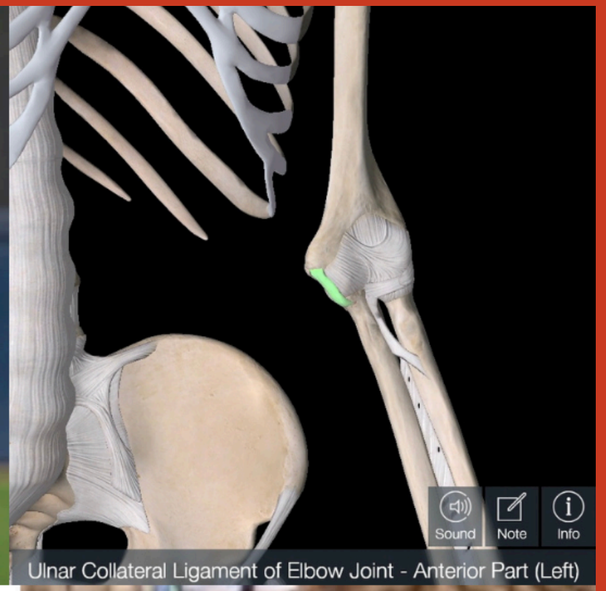
Covid-19

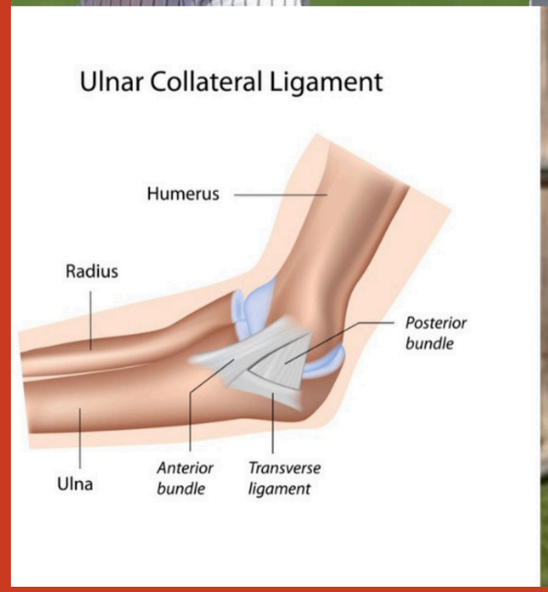
Surprisingly noted as the least common injury, though it significantly impacted seasons and player availability, reflecting the pandemic's influence on sports.

Injury Statisics

Understanding the distribution and frequency of various injuries helps in developing better training and recovery protocols for players.

Need Pie chart Kayli





Understanding Injuries Leading to Tommy Jone Surgery

Tommy John Surgery
Surgery is required when a player suffers a tear or severe damage to the ulnar collateral ligament (UCL) in the elbow.

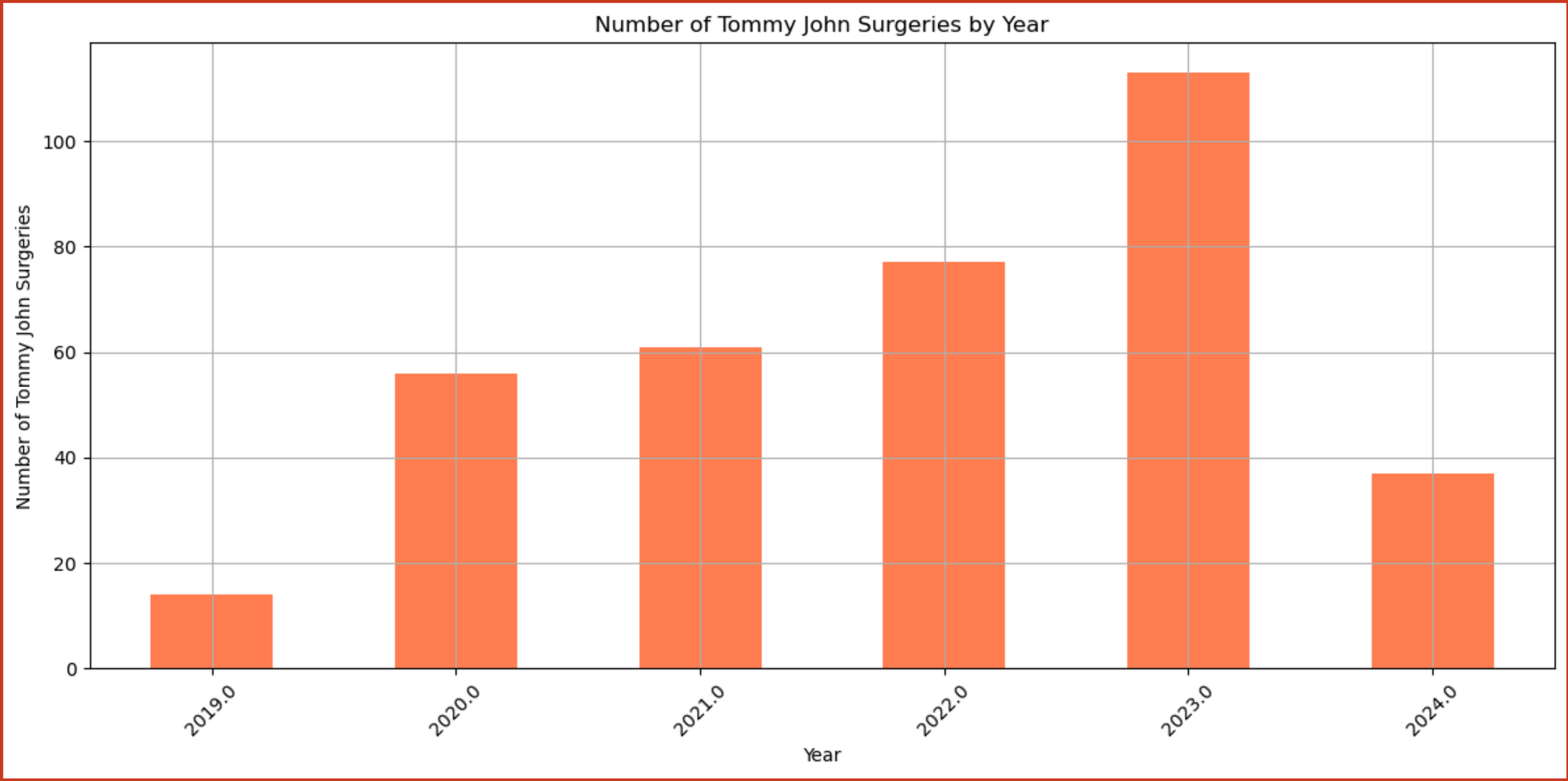
Causes of UCL Injuries
Repetitive Stress: Overuse from throwing, especially for pitchers, can cause the UCL to weaken over time.
High Velocity: The increased speed and intensity of pitches can put extra strain on the elbow ligament.
Poor Mechanics: Incorrect pitching techniques can increase the risk of UCL tears.

Tommy John Surgery

- Dataset: Analyzed data on pitches and Tommy John (TJ) surgeries across players.
- Target Variable: Tommy John Surgery.
- Model: Trained a Random Forest Classifier to predict TJ surgeries based on pitch count.
- Performance: Achieved 60.06% accuracy on the test set.

- Model Optimization: Enhanced model by fine-tuning parameters, focusing on 'n_estimators.'
- Accuracy Improvement: Optimization led to a notable accuracy increase on the test set.
- Final Results: Optimized model accuracy improved to 70% from the initial 60.06%.
- GridSearchCV Details: Explored 108 hyperparameter combinations with 5-fold cross-validation, totaling 540 fits.

Tommy John Surgery



Limitations of Our Baseball Injury Analysis

Add more text

Incomplete or Inaccurate Data:
Injury reports may be inconsistent, incomplete, or lack detail, which could affect the accuracy of predictions.

2. Overabundance of Data:
The sheer volume of historical player and injury data can be overwhelming, making it difficult to isolate relevant factors and identify key trends.

3. Bias in Data Collection:
Injury data could be biased toward higher-profile players, potentially skewing results.

4. External Factors Not Captured:
Factors such as player fatigue, environmental conditions, and pre MLB medical histories may not be fully represented in the dataset.

5. Complexity of Injury Causes:
Injuries like Tommy John surgery result from a combination of factors, making it challenging to create accurate predictive models.

6. Predictive Model Generalization:
Models may overfit to the current data, making it difficult to generalize predictions to new players or seasons.