Evaluating Defensive Catcher Performance in Major League Baseball: A Machine Learning-Based Scoring Model

A Thesis Submitted
In Partial Fulfillment of the Elon Master of Science in Business Analytics Program

By: James T. Tegler Jr.
May 2025

Thesis Mentor: Dr. Alex Traugutt

Approved by:
Alexander Traugutt
Name of Professor, Department (Thesis Mentor)

Abstract

This study was inspired by a question posed in a job posting by an MLB team's R&D department: "The R&D department has been asked to identify the best defensive catcher in baseball. What models would you build to answer that question, and how would you apply those models to decision-making?" In response, this research develops a machine learning-based model to evaluate defensive catcher performance. This is an area that historically has been difficult to quantify, compared to other positions. Using Statcast data from 2018 to 2024, principal component analysis (PCA) was applied to reduce dimensionality across defensive metrics (framing, blocking, pop time, and throwing). K-means clustering segmented catchers into statistical profiles, and logistic regression identified elite performers, defined as the top 25% in Defensive Runs Saved (DRS). The model generates two outputs: an Absolute Score (z-score of elite probability) and a binary elite label. Model 1, which excluded age and salary, achieved a 0.919 AUC, indicating strong predictive power. Model 2 then adds age and salary data to see if model performance increases. Results highlight inefficiencies in player compensation and reveal defensive assets. As MLB teams increasingly invest in AI, machine learning, and analytics to improve roster construction and maximize return on investment, this model offers a scalable and interpretable tool. By aligning performance with contract value, teams can make smarter decisions that boost revenue and roster efficiency, ultimately improving their chances of winning more games.

1. Introduction

Of the four major sports leagues in America (NFL, NBA, MLB, and NHL), the sport most heavily driven by analytics has been Major League Baseball. Baseball is unique in that it lends itself to statistical analysis more than any other sport due to the wealth and discrete nature of the data (Mizels et al., 2022). Each event in baseball is a separate and countable event. From the pitch to the swing, determining if it is a hit or not, a fielding attempt, and finally, if the runner is out or safe. For every action, there is a clear beginning and end, which is different from both basketball and soccer. In football, there are plays, but within those plays, the actions can still be highly chaotic and interdependent. With baseball, every pitch, at-bat, or play can be recorded as a distinct and clean datapoint. This makes baseball an ideal environment for analytics, with the application of advanced statistical techniques, we can derive valuable insights for decision making and player evaluation.

The goal of analytics is to turn data into actionable insights for decision makers. This is especially useful for player evaluation, where differences in player performance can be explained by a variety of factors (Alamar 2023). To create those actionable insights and figure out the most efficient ways to succeed in baseball, Sabermetrics were created. Sabermetrics began with the Society of American Baseball Research's Statistical Analysis committee, a group co-founded by researchers Dick Cramer, Bill James, and Pete Palmer in 1974 (Kelly 2019). Bill James is credited with creating the name, taking it from the acronym SABR. He is considered the paternal figure of the movement (Kelly 2019). Sabermetrics asks questions about how baseball is played and then answers the questions through empirical research. The answers must be backed up with quantifiable evidence. Questions can range from a simple data query to abstract questions, much like this paper attempts to answer. The argument against Sabermetrics is that it collides with baseball traditional strategies, which are built on feel and tradition.
Due to this, it was ignored by many big-league decision makers during the 1980s and early '90s (Kelly 2019). It wasn't until the Moneyball Oakland Athletics, who favored metrics like on-base percentage and slugging percentage and used that to build a competitive team during the 2002 season, that organizations began to use advanced statistical analysis. Sabermetrics has transformed baseball from traditional subjective evaluation methods to objective and data-driven analysis. By focusing on advanced metrics, teams can uncover inefficiencies in player valuation and gain a competitive edge.

Analytics in baseball took a turn during the 2003 MLB season, when General Manager Billy Beane and the Oakland Athletics revolutionized MLB and analytics through exploiting inefficiencies in the baseball labor market. The ability to get on base was severely underrated (Wolfe et al., 2007). *Moneyball* is the story of how analytics changed baseball player evaluation and impacted how we analyze players today, as machine learning and AI algorithms continue to become more sophisticated. For over a century, baseball talent evaluation relied heavily on scouting. Traditional player evaluation involved the five-tool approach: hitting for average, hitting power, fielding, foot speed, and arm strength (Wolfe et. al, 2007).

Pitchers were assessed on arm strength, types of pitches thrown, and control. *Moneyball* increased the tension between the use of subjective vs. objective data (Wolfe et al., 2007). What Beane illustrated was that sabermetrics provides teams with an advantage through statistical analysis over traditional scouting attributes. Through statistical analysis, Beane replaced traditional batting statistics like batting average and RBIs with on-base percentage (OBP) and slugging percentage (SLG). OBP measures how often a batter reaches base per plate appearance, incorporating hits, walks, and hit-by-pitches, capturing a hitter's ability to avoid making outs. SLG, on the other hand, accounts for the total number of bases a player earns per at-bat, assigning greater weight to extra-base hits, such as doubles, triples, and home runs. Unlike batting average, which treats all hits equally and excludes walks, both OBP and SLG offer a more comprehensive approach to evaluating offensive production.

For pitchers, Beane discarded stats like wins and ERA in favor of Defensive-Independent Pitching Statistics (DIPS), focusing on outcomes pitchers could control, such as walks, strikeouts, and home runs (Wolfe et al., 2007). Beane's approach demonstrated that patience at the plate was a highly valuable tactic. Ultimately, the use of sabermetrics challenged baseball's deeply ingrained reliance on subjective scouting, showing that objective and data-driven analysis could uncover undervalued skills and provide a sustainable competitive advantage for small-market teams.

Of all the defensive positions in baseball, catcher is the most physically and mentally demanding. No other position requires a player to be as involved on every pitch of every inning. Catchers spend most of the game squatting behind the plate, bearing the weight of heavy protective equipment, and are tasked with responsibilities other than receiving pitches. They call the game, manage pitcher-batter matchups, understand each batter's strengths and weaknesses, and maintain strong relationships with their pitchers to effectively adapt game plans mid-game.

Physically, the demand on a catcher's knees, legs, and overall stamina is unmatched. Agility, arm strength, and concentration are key physical skills required for a catcher's success. This helps to handle foul tips, wild pitches, and quick throws to the bases on attempted steals (Baseball Biographies, n.d.). Mentally, catchers act as the on-field managers, controlling the flow of the game through pitch selection and strategic adjustments. The relationship between pitcher and catcher is also crucial to team success, and effective communication can be a difference maker during the game.

This combination of mental acuity, physical endurance, and leadership makes the catcher position arguably the hardest in baseball. Their influence on the game is more subtle and not as easily measured as other positions on the field. However, catchers have historically been undervalued compared to other players, particularly when it comes to compensation.

Recent technological advancements like Statcast have revolutionized how catcher defense is measured. Statcast is a publicly available tool for data collection of baseball players and uses state-of-theart tracking technology for the collection and analysis of baseball data. Statcast was installed in all thirty MLB ballparks in 2015 (MLB). One of the most important developments is the Defensive Runs Saved (DRS) metric, which quantifies how many runs a player prevents compared to the league average at their position. DRS is particularly powerful because it breaks catcher defense down into distinct components such as pitch framing, blocking, controlling the running game, and normal fielding (DeForest 2022).

Among these, pitch framing stands out as the most important differentiator between elite and average catchers. Since a catcher is tested on every pitch, small differences in framing ability can add up to major impacts throughout a season (DeForest 2022).

Ultimate Zone Rating (UZR) is another widely used defensive metric, designed to measure a fielder's defensive contributions in runs saved compared to the average player at their position (Baumer & Zimbalist, 2014). While UZR is robust for most fielders, it is less commonly used for catchers because it doesn't adequately capture specialized skills like framing and blocking (DeForest 2022). Therefore, for evaluating catcher defense, DRS offers a more comprehensive framework but does not completely answer the question.

In this paper, I expand on previous research by building a scoring model based specifically on defensive catcher performance. While DRS provides a strong starting point, there remains a gap in comprehensive and reliable evaluation models tailored to the unique demands of the position. Traditional scouting methods have often struggled to objectively rank catchers based solely on defensive contributions. A catcher's impact is spread across framing pitches, blocking errant throws, and managing the pitching staff, which are difficult to quantify using traditional metrics.

The introduction of Statcast has been a major step forward, providing access to granular defensive data such as pop time, framing, catching, and throwing. However, organizations still face challenges in creating a unified ranking system that accurately captures defensive value.

My approach aims to complement existing frameworks by applying unsupervised learning techniques, specifically principal component analysis (PCA), to reduce the dimensionality of complex data and uncover latent patterns in catcher performance profiles. In doing so, I can create clusters of similar catchers based on their defensive skill sets, providing a clear picture of how catchers compare defensively across the league. Ultimately, the goal of this research is to provide a data-driven, repeatable, and actionable method for ranking catcher performance, addressing a critical and still-evolving area within baseball analytics.

2. Literature Review

2.1 Introduction: The Rise of Analytics in Baseball and Business Efficiency

Baseball has evolved into one of the most analytically driven sports, thanks to its discrete, countable events and historical data (Koseler & Stephan, 2017; Mizels et al., 2022). Historically, baseball has been a sport that is rooted in tradition, often labeled "America's pastime" (National Baseball Hall of Fame, n.d.). However, it, like other sports and businesses today, is ever-changing and adapting to new technologies and innovations. Baseball analytics began in the 1920s with people developing models to predict outcomes and develop winning teams (Mizels et al., 2022). Then there was the rise of sabermetrics, which has led to baseball becoming one of the biggest data-driven sports. This growth of analytics was accelerated by new MLB ownership groups from tech, finance, and consulting backgrounds who prioritized ROI, Business Intelligence (BI) systems, and data-driven decisions (Hayduk, 2022). These owners have embraced data not only for on-field decisions but also for maximizing organizational efficiency and labor productivity. As analytics continue to expand, researchers have begun exploring the integration of machine learning (ML) models into the game's already data-rich environment, offering new methods for understanding performance patterns and player valuation. The current studies emphasize the growing use of machine learning techniques-such as PCA, clustering, and regression, to evaluate player performance across various sports. However, few have applied these methods to defensive catcher evaluation, leaving a critical gap that this research seeks to address through an integrated, data-driven scoring model grounded in DRS as a benchmark.

In a study done by Hayduk (2022), the author examines the impact of "tech-savvy" ownership on the financial performance of Major League Baseball teams. Data was collected on the valuation of each MLB team over fourteen years, to analyze the role business intelligence (BI) career experience has on an organization's success. The theory is that BI technologies can help firms optimize revenue and expenses if acquired and deployed proficiently (Hayduk, 2022). Operating margin was used as the dependent variable, calculated as operating earnings divided by revenue (Hayduk, 2022). The analysis found that executive influence on operating margin relates more to cost management than revenue maximization (Hayduk, 2022). The impact was positive, but marginal. Teams with BI-proficient ownership were better at managing payrolls and squeezing more value out of each dollar spent on players. This reflects a growing trend across the league. Newer owners from finance, tech, and consulting are bringing with them a data-driven mindset that contrasts with earlier generations who made decisions based on instinct and tradition. The plethora of data available is giving owners a new avenue to analyze players and make roster decisions. These small but impactful advantages reinforce the importance of developing tools that can help teams allocate payroll more effectively. My paper contributes to this line of research by applying machine learning to catcher defense, thus allowing owners to make informed decisions about roster and maximize value out of the position.

2.2 Challenges in Player Evaluation: Offensive vs. Defensive Metrics

When it comes to sabermetrics, offensive statistics are more standardized and widely accepted. It is much simpler to track offensive statistics such as batting average, home runs, runs batted in (RBIs), onbase percentage (OBP), and slugging percentage (SLG). However, this does not tell the whole story. Hitting is more complicated than just batting average. In Baumer & Zimbalist's book, *The Sabermetric Revolution: Assessing the Growth of Analytics,* they found that the correlation between batting average and runs scored is positively correlated at 0.82, and 67% of the variation in runs scored is explained by batting average. This is the same conclusion that was drawn in *Moneyball*: that on-base percentage and slugging percentage are even better at predicting runs. Their correlation was at 0.95 (Baumer & Zimbalist, 2014). The combination of on-base plus slugging is a simple metric, yet accurate in determining how effective a player is on offense, even more so than batting average (Baumer & Zimbalist, 2014). Since this book was published in 2014, Statcast was released by MLB. This innovation marked a turning point in data availability, enabling a new generation of performance metrics that continue to reshape how players are evaluated.

Pitching presents a more complicated evaluation challenge because it is inherently tied to team defense. Traditional pitching statistics such as earned run average (ERA), wins, and innings pitched (IP) have long been used to measure performance, but these metrics can be misleading. For instance, ERA is influenced by the quality of fielding behind the pitcher and the scorer's judgements of what constitutes an "earned" run. Pitching evaluation didn't change much until the theory of Defense Independent Pitching Statistics (DIPS). The introduction of DIPS has largely discredited ERA by demonstrating that the percentage of balls put in play against a particular pitcher that fall for hits is much more subject to chance than conventional wisdom allowed (Baumer & Zimbalist, 2014). This is fundamental in understanding the relationship between pitching and defense. It is largely out of the pitcher's control to prevent hits on balls hit in the field of play. Once the ball has been put in play, it doesn't matter all that much whether it was put in play by one pitcher or another (Baumer & Zimbalist, 2014). The difficulty of putting a ball in play against a particular pitcher varies dramatically and captures all the variation in pitcher skill. They should be judged by what happens when the ball is not put in play against them, the number of strikeouts, walks, and home runs given up that the pitcher records. As sabermetrics evolved, analysts developed defenseindependent metrics like Fielding Independent Pitching (FIP), which isolates outcomes that are solely under the pitcher's control, strikeouts, walks, hit batters, and home runs allowed, adjusted for league averages (Baumer & Zimbalist, 2014). FIP is a better predictor of future ERA than ERA itself because it reduces the noise introduced by defensive variability.

Further advancements, such as expected (xERA), powered by Statcast data, account for batted ball quality, exit velocity, and launch angle to provide a more accurate estimate of a pitcher's expected performance. These tools have enabled front offices to move beyond surface-level stats and gain a more precise understanding of pitcher effectiveness (Mizels et al., 2022). However, even with these developments, a pitcher's success still relies heavily on the defense behind them, reinforcing the need for better defensive metrics and a clearer understanding of the link between pitching and fielding performance.

While offensive statistics in sabermetrics are more standardized and widely accepted, defensive metrics, specifically fielding, have lagged in development and reliability. Historically, defense has been measured using errors and fielding percentage. An error is defined as failing to get an out on a routinely batted ball that is expected to result in an out (Mizels et al., 2022). The issue with this is that it requires the scorer's judgement, and it brings in human error. It also doesn't account for "bad defense", if a player gets a slow jump on a ball and doesn't make the play, technically it is not an error (Mizels et al., 2022). Fielding percentage is the common way to measure a player's defense, which is the ratio of total plays made (assists + put outs) to recorded total opportunities (assists + put outs + errors), which is a sensible way of combining these three statistics (Baumer & Zimbalist, 2014).

Attempts have been made to better measure a player's defensive performance. Revised Zone rating (RZR) measures how often a fielder converts batted balls in their zone into outs (Mizels et al., 2022). It is similar in theory to errors and fielding percentage. Defensive runs saved (DRS) and Ultimate Zone Rating (UZR) build on RZR by quantifying how many runs a player prevents while playing defense (Baumer & Zimbalist, 2014). For example, if a play is made only 40% of the time across the league, a successful attempt is weighted accordingly (Baumer & Zimbalist, 2014). However, these statistics are limited by the number of truly difficult batted balls that are hit to each fielder every year. It is out of the fielder's control where the batter puts the ball in play. These examples demonstrate the complexity of understanding defensive play. They have long been considered poorly captured with traditional sabermetrics (Baumer & Zimbalist, 2014).

Additionally, UZR faces methodological criticisms. It is a closed-source system, offering no transparency into its calculations (a black box), and lacks published confidence intervals or error margins. Year-to-year UZR scores are highly variable, with correlations similar to batting average, suggesting low consistency and poor predictive value (Baumer & Zimbalist, 2014). The metric also conflates range and pre-pitch positioning, a critical issue in the era of defensive shifts, where coaching strategies often play a major role. While newer tracking technologies like Statcast have introduced tools such as catch probability and reaction time, these advancements have primarily benefited outfield and infield evaluation, not catchers (Mizels et al., 2022). The wealth of data from Statcast has enhanced the fan experience and created jobs by forcing teams to hire statisticians to better understand the data and recommend how to act upon it. It is revolutionizing baseball strategy by finding new ways to win games and evaluate players.

This gap presents a significant opportunity. Despite increasing access to granular data on framing, blocking, pop time, and throwing metrics, catcher defense still lacks an integrated evaluation framework. My study seeks to address this by developing a machine learning based model that synthesizes these inputs and aligns them with Defensive Runs Saved, offering a more comprehensive and scalable method for quantifying catcher value.

2.3 Machine Learning Applications in Sports Analytics

In recent years, machine learning has become a cornerstone of modern sports analytics, enabling analysts to process high-dimensional data and uncover patterns that traditional statistics might miss. Analytics is a rapidly growing and evolving field that allows organizations to deploy many methods for different situations. The knowledge of advantages and disadvantages for each player can give added value in roster composition (Sarlis & Tjortjis, 2020). The issue with sports data is that it is irregular and sparse. A majority of players don't have long careers and don't remain in the same league for many years (Sarlis & Tjortjis, 2020). Baseball has numerous statistics that can be tracked, evident in the data collected for this study. There are many web pages and blog articles covering sports analytics, but there is a lack of credible and peer-reviewed research articles (Sarlis & Tjortjis, 2020). The challenge is the need for domain experts to combine scientific research experience with a professional sports career, which is a difficult task. To truly understand the intricacies of sports, one must play them at a high level. This allows the individual to critically analyze the sport and figure out how to maximize the value to be gleaned from analytics. Despite these challenges, sports organizations are placing a growing emphasis on analytics to improve roster construction, evaluate player performance, and support decision-making.

Sarlis & Tjortis (2020), in this paper, attempt to evaluate existing performance analytics used in European and NBA basketball to predict who the MVP and DPOY will be. The authors want to increase the understanding of important insights and minimize the possibility of uncertain current or future events (Sarlis & Tjortjis, 2020). The authors then evaluate multiple common algorithms that have been reported in literature, such as neural networks for classification and prediction, decision trees for predictive models, Bayesian networks, support vector machines for classification and regression analysis, linear and logistic regression, and unsupervised learning (Sarlis & Tjortjis, 2020). Unsupervised learning methods, particularly clustering algorithms, are emphasized as valuable tools for grouping players based on style or performance. These techniques are especially useful when traditional labels (e.g., "elite defender" or "franchise player") are unavailable or subjective. In clustering, for example, players are grouped by statistical similarities, allowing decision makers to identify archetypes and build more balanced rosters.

This approach directly supports the methodology used in this paper, which applies principal component analysis (PCA) to reduce dimensionality in catcher defensive data, followed by k-means clustering to identify player groupings. In their review of sports analytics, Sarlis and Tjortjis (2020) emphasize that modern sports data, especially in high-variance environments like basketball, is often irregular, sparse, and difficult to interpret using traditional statistics. They highlight how clustering allows analysts to group players with similar styles or performance profiles without relying on subjective labels, enabling more objective and data-driven evaluation. Similarly, this study leverages PCA and clustering to uncover hidden performance patterns among MLB catchers, offering a scalable way to differentiate defensive roles based on statistical profiles rather than reputation or surface metrics.

Sarlis and Tjortjis also construct predictive models to forecast award outcomes like MVP and Defensive Player of the Year, using performance data from current seasons. Their models incorporate supervised techniques such as linear and logistic regression to connect player groupings to real-world outcomes. This aligns with the current study's use of logistic regression to connect catcher cluster scores to DRS, a widely recognized benchmark for defensive value. Just as Sarlis and Tjortjis validate their clusters by comparing them to meaningful external targets, this study uses regression as a validation mechanism to ensure the proposed scoring model reflects true defensive impact.

Furthermore, their paper underscores the importance of dimensionality reduction when working with high-dimensional, unstructured sports data. They argue that methods like PCA are essential to simplify complex datasets and improve model interpretability. A conclusion further supporting this study's first step of reducing multicollinearity in catcher metrics before clustering (Sarlis & Tjortjis, 2020). Finally, their emphasis on creating aggregated performance indicators, to synthesize multiple stats into a single, actionable score, directly parallels the goals of this project. Just as their models sought to quantify overall player impact for award forecasting, this catcher model aims to condense defensive performance into a composite score that front offices can use to make more cost-effective roster decisions.

Beyond basketball, baseball has increasingly embraced machine learning to address its complex player evaluation challenges. Koseler and Stephan (2017) highlight how the sport's discrete, event-driven structure and extensive statistical history make it particularly well-suited for analytical techniques like clustering and dimensionality reduction. Their study emphasizes the importance of applying principal component analysis (PCA) to extract meaningful features from high-dimensional baseball data. PCA helps reduce multicollinearity and improve interpretability by transforming correlated variables such as framing, blocking, and throwing stats into a smaller set of components that still retain the majority of the variance. This step is critical in creating a reliable model that allows decision-makers to draw actionable insights from complex datasets. The authors argue that the ability to simplify such data is central to identifying undervalued players and optimizing roster construction in a highly competitive landscape (Koseler & Stephan, 2017).

Academic work by DeForest further demonstrates how clustering methods like k-means can be used to identify undervalued pitchers by grouping them according to similar performance characteristics. In this project, clustering algorithms were used to determine undervalued players and classify them based on pitch type and repertoire (DeForest 2022). This clustering approach aligns with the methodology of the current study, which groups catchers based on PCA-derived defensive traits. These player profiles help illustrate performance patterns otherwise hidden in traditional metrics. Regression techniques, including ordinary squares are frequently used across sports analytics literature to validate performance models by linking cluster-based scores outcomes such as MVP selection, player value, or salary (Koseler & Stephan, 2017). By applying OLS regression to associate catcher cluster scores with DRS, this paper follows an established framework that brings analytical rigor to catcher defense evaluation. The combination of PCA, clustering and regression reflects best practices in sports analytics and reinforces the model's utility for improving payroll efficiency and player valuation.

3. Research Framework

The research framework combines player demographic characteristics, advanced defensive metrics, and machine learning methods to evaluate MLB catcher performance. Figure 1 illustrates the research framework used in this study, which integrates Statcast metrics, demographic data, and machine learning techniques to model elite catcher performance. The framework involves dimensionality reduction using principal component analysis, followed by clustering with K-means to segment players into distinct tiers, and finally, regression models to quantify cost-performance alignment using DRS and contract data.

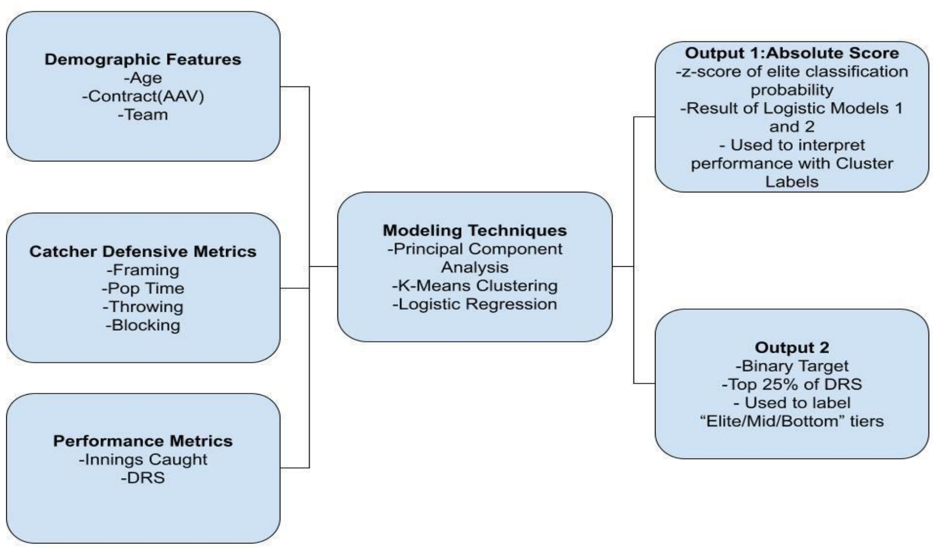


Figure 1. *Illustration of our research framework.*

3.1 Variables

Table 1 summarizes and defines some of the key variables in the dataset and the model that was created. These variables span both contextual player information (e.g., age and contract value) and Statcast-derived defensive metrics across four dimensions. Understanding the definitions of each variable is essential for ensuring analytical consistency.

Variable	Definition
Demographic Features	

Average Annual Value (AAV)	Average value of a player's contract for one year
Age	Age of the player during that season
Blocking	
Block Above Average	Number of Passed Balls + Wild Pitches saved compared to expectation of average catcher.
Blocks Above Average Per Game	Rate stat based on average catcher receiving 40 blocking chances per game.
Catcher Blocking Runs	Transition from blocks to a run value.
Framing	
Strike Rate	Strike rate, measured across specific zones (11-19) and aggregated as a cumulative metric
Catcher Framing Runs	Converts strikes to runs saved on a .125 run/strike basis, and includes park and pitcher adjustments
Pop Time	
Pop Time	Time from the moment the pitch hits the catcher's mitt to the moment the ball reaches the fielder's projected receiving point.
Exchange Time	How quickly the catcher releases the ball, in seconds.
Arm Strength	Average arm strength of catcher's throws to second, measured on average above players 90^{th} percentile performance.
Throwing	
Caught Stealing Percentage	Percentage of steal attempts that were caught stealing.
Catcher's Caught Stealing Above Average	Number of extra caught stealing compared to the expectation of the average catcher.
Catcher Stealing Runs	Translation of Catcher Stealing Above Average to a run value.

Table 1. *Variable definition and representation*

4. Experimental Setup

4.1 Data Collection

The data used in this research was collected from Baseball Savant, which leverages Statcast technology. Statcast captures high-speed, high-precision data on both player movements and ball trajectories, making it an essential tool for modern performance analysis (Casella 2015). This dataset spans seven MLB seasons (2018-2024) and includes all catchers who caught at least 400 pitches in a given season. In total, there are 405 players in the dataset. Defensive performance was broken down into four key categories: framing, throwing, blocking, and pop time. Within each category, multiple variables were collected to account for variation in skill and impact across players.

Preprocessing involved manually merging datasets and excluding players who did not appear in all four categories. Missing values were imputed using the median for each feature. Age was then collected from *Baseball Reference*, as well as contract value (AAV) for that season, collected from *Spotrac.com.* Defensive Runs Saved (DRS) and innings caught were collected from *Fielding Bible*. Table 2 presents descriptive statistics for the main variables used in this analysis, including contract value, age, and each of the four defensive components. With these variables, I was able to construct a comprehensive player profile used for further analysis, including PCA, clustering, and logistic regression modeling. The following subsection offers a deeper explanation of the four main components of catcher defense evaluated in this study: blocking, framing, pop time, and throwing.

Variable	**Mean**	**SD**	**Max**	**Min**
Average Annual Value	$3,254,290.45	$4,616,730.08	$23,100,000	$190,000
Age	28.96	3.52	39	21
Blocks Above Average	0.428	7.38	24	-26
Blocks Above Average Per Game	0.000953	0.1156	0.37	-0.47
Catcher Blocking Runs	0.1494	1.84	6	-7
Strike Rate	47.52	2.827	53.8	40
Catcher Framing Runs	0.26	5.44	17	-18
Pop Time	1.93	0.0484	2.09	1.78
Exchange Time	0.677	0.394	0.822	0.565
Arm Strength	78.54	2.57	85.96	72.767
Caught Stealing Percentage	0.20	0.058	0.52	0
Catcher's Caught Stealing Above Average	0.1145	7.584	40.194	-27.65
Catcher Stealing Runs	0.074	4.93	26.126	-17.97

Table 2. *Descriptive Statistics for Catcher Variables*

4.2 Dataset Description

The first aspect of the dataset is catcher blocking, a metric designed to evaluate how skilled catchers are at preventing wild pitches or passed balls. One major league pitch takes less than half a second to go from pitcher to catcher (Baccellieri 2023). It is incredibly difficult to conceptualize that, and it happens in the blink of an eye. Statcast sets up twelve cameras positioned around each ballpark to collect data. There are pitch-tracking cameras that capture 300 frames per second which help to capture data from each pitch (Baccellieri 2023). For catcher blocking, each pitch is assigned a probability of being a passed ball or wild pitch based on several inputs, namely pitch speed, location, movement, catcher location, and handedness of the batter.

The next aspect in the dataset is catcher framing which involves the catcher receiving a pitch and moving the glove in a way to convince the umpire to call the pitch a strike (Statcast). Pitch framing is a defining skill for catchers, and a well-framed pitch can sway the umpire's call in favor of the pitcher. A pitch that may be on the edge of the strike zone can be maneuvered by the catcher's glove to make it seem like an easy strike. To qualify in this dataset, a catcher must receive six called pitches per team game. Strike rate represents the overall percentage of borderline pitches called strikes by that catcher, and framing runs is an estimate of how many runs a catcher has either saved or cost the team through pitch framing. Catcher framing runs are cumulative over the season, positive values are above average framers, and negative values indicate catchers potentially costing their teams runs by losing on borderline strike calls.

Pop time is a metric that measures the time it takes from the moment the pitch hits the catcher's glove to the moment the ball reaches the fielder's projected receiving point at the center of the base. It combines exchange time, how quickly the catcher releases the ball, and arm strength, in miles per hour. The average MLB pop time on steal attempts to second base is 2.0 seconds (Statcast). Pop time is a better representation of a catcher's ability to throw out baserunners than arm strength alone (MLB Advanced Media, n.d.). A catcher with only a great arm will not throw out many baserunners if the exchange time from ball to hand to release takes a long time. A short pop time is the best chance a catcher has of throwing out the runner. In this dataset, I have compiled exchange time, number of attempts, average pop times on throws to both second and third base, and times when the runner was caught and times when the runner was safe.

The final variable incorporated in the analysis is catcher throwing. Catcher throwing expresses the skill of the catcher at throwing out runners on steal attempts, looking at the specifics of the opportunities presented. Each steal attempt in the dataset is assigned a probability of success; while looking at several inputs at the time the pitch crosses the plate. These are: runner distance from second, runner speed, pitch location, and awareness of pitchouts or delayed steals. The most important variable in this dataset is caught stealing above average. This is the difference between actual caught stealing attempts and estimated caught stealing based on the attempts seen.

5. Results and Discussion

5.1 Problem Introduction

The inspiration for this research came from a question posed by the Philadelphia Phillies, where they asked, “The R&D department has been asked to identify the best defensive catcher in baseball. What models would you build to answer that question, and how would you apply those models to decision making?” To answer this question, I combined dimensionality reduction (PCA), unsupervised clustering, and supervised learning to identify and score catcher defensive performance using objective data.

The idea behind principal component analysis is simple: it reduces the dimensionality of a dataset while preserving as much of the variability as possible (Jolliffe & Cadima, 2016). PCA was applied to all standardized numerical defensive metrics, including blocking, framing, pop time, and throwing features. This reduced the dimensionality while accounting for high correlation among inputs. Since PCA is sensitive to scale differences between variables (e.g., runs saved vs seconds), all input features were standardized prior to analysis to ensure equal contribution to component formation (Jolliffe & Cadima, 2016). The final PCA transformation produced eighteen uncorrelated components that retained approximately 90% of the total variance in the defensive metrics.

To evaluate individual player performance, each catcher’s predicted probability of elite classification was transformed into a standardized z-score, referred to as the Absolute Score. This score was computed separately for both models. Model 1 used PCA inputs only, while Model 2 included age and salary enabling objective, side-by-side comparisons of defensive value.

5.2 K-Means Clustering

After the data was reduced with PCA, K-means clustering was applied to those components. K-means clustering allows us to take the unlabeled data points and group them based on commonalities among the data. For this research, four clusters were determined to be the optimal number, as determined by the elbow method, seen in Figure 2. A visual representation of the clustering results is provided in Appendix D. The resulting clusters were labeled “Top Tier, Middle Tier, Overpaid, Young & Underpaid”. These clusters were named based on average values for Defensive Runs Saved, age, and contract (AAV) within each group.

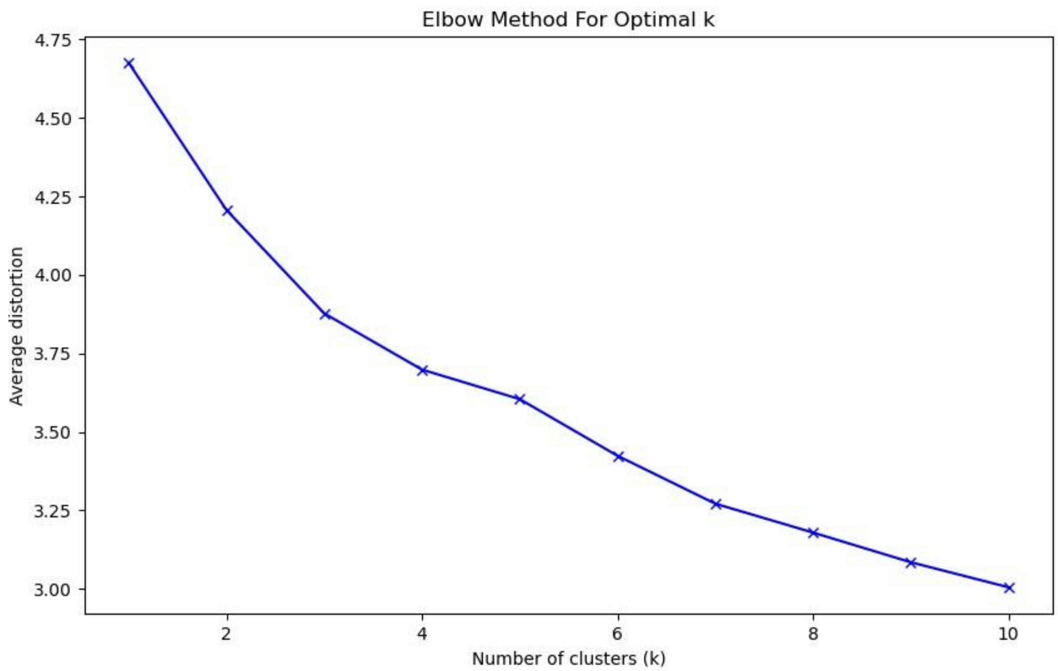


Figure 2. *Elbow Method Plot to Determine Optimal Number of Clusters(k=4)*

It is important to note that the cluster labels in this paper are the result of an unsupervised method based on statistical similarity. The performance tiers (Elite, Middle, Bottom) used later in the model are created through supervised learning and directly tied to elite defensive outcomes (defined as the top 25% of DRS scores across the dataset). The two segmentations serve different purposes and complement each other by combining player profile similarity with the outcome. This segmentation allows for more interpretable analysis of performance types and highlights potential market inefficiencies. The unsupervised approach provided an interpretable segmentation of defensive catcher types. Table 3 compares the cluster labels from unsupervised K-means with the performance tiers derived from logistic regression to clarify how these groupings differ in purpose and structure.

Group Type	**Source**	**Purpose**
Cluster Label	K-Means (Unsupervised)	Player segmentation
Performance Tier	Logistic Regression (Supervised)	Elite/Mid/Bottom classification

Table 3. *Distinction Between Cluster Labels and Performance Tiers*

5.3 Modeling Approach & Justification

Two logistic regression models were constructed to evaluate defensive catcher performance, using Defensive Runs Saved (DRS) as the binary target for elite classification. In this study, a catcher is classified as “elite” if they fall within the top 25% of DRS scores across the dataset. This threshold finds a balance between statistical robustness and practical relevance. It captures the standout defenders without being so narrow as to limit model training or so broad as to dilute the elite label. From a practical standpoint, this cutoff reflects how MLB teams potentially assess positional value. Placing an emphasis not just on being above average but identifying players who significantly impact run prevention and can justify contractual decisions.

5.4 Model Results

Model 1 used only PCA components in logistic regression (framing, blocking, pop time, and throwing). This unsupervised foundation allowed the model to capture performance patterns without external bias. Model 2 added demographic and contract information, specifically age and average annual value (AAV), to assess whether these contextual factors improved prediction. As illustrated in Table 4, Model 1 performs slightly better across all evaluation metrics compared to Model 2, most notably in accuracy (0.886 vs. 0.815) and ROC AUC (0.919 vs. 0.873). The AUC of 0.919 for Model 1 indicates strong predictive power of the PCA-based features, validating the use of unsupervised clustering as a foundation for elite classification. Table 4 summarizes the performance of both logistic regression models, including accuracy, precision, recall, F-1 score, and AUC. These results suggest that raw defensive metrics, when condensed into PCA components, are the most reliable indicators of elite defensive performance. ROC curves comparing the performance of both models can be found in Appendix F. While age and contract value may influence team management decisions, they are not strongly correlated with on-field output.

Metric	**Model 1: PCA Only**	**Model 2: PCA + Age + Contract**
Accuracy	0.886	0.815
Precision (Elite)	0.842	0.739
Recall (Elite)	0.727	0.654
F1-Score (Elite)	0.780	0.694
ROC AUC	0.919	0.873

Table 4. *Performance Metrics for Logistic Regression Models (Elite = Top 25% of DRS)*

While Model 2 does not offer improved classification performance, it adds value through interpretability by connecting model predictors to market factors. This is especially useful in identifying mismatches between performance and compensation. For instance, players with high predicted probabilities of elite status but relatively low salaries may represent undervalued assets, while low elite classification probabilities and large contracts are players who are overpaid. Thus, although the second model does not outperform the first statistically, it offers additional insights into roster efficiency and financial decision-making.

5.5 Absolute Score and Probability Interpretation

To better contextualize individual player performance, each catcher was assigned an Absolute Score from Model 1, calculated by standardizing the elite classification probability from Logistic Model 1(PCA only). This was done using z-score standardization: (Elite Probability - μ) / σ, where μ is the mean elite probability across all players and σ is the standard deviation. This score quantifies how far above or below average each player's elite probability is, based solely on the model predictions. A higher score reflects a stronger likelihood of elite defensive value. The standardized score also allows for direct comparison of players, regardless of the original units (e.g., seconds vs. runs saved). Table 5 presents the top elite performers in the model.

For further analysis, a similar Absolute Score from Model 2 (PCA + age + contract) was also calculated, to evaluate how external factors like age and salary might influence perceived player value when on-field performance stays constant. Table 6 highlights both high and low performing catchers across both models, showing the distinction between elite and bottom-tier players.

5.5.1 Top Elite Probabilities (Model 1 vs. Model 2)

Player Name	Elite Probability Model 1	Elite Probability Model 2	Absolute Score (Model 1)	Absolute Score (Model 2)	Defensive Runs Saved	Age	Cluster Label	Performance Tier
Roberto Perez	0.999832	0.999472	2.095	2.116	31	30	Top Tier	Elite
Patrick Bailey	0.998837	0.999280	2.092	2.116	20	25	Top Tier	Elite
Adley Rutschman	0.992544	0.987237	2.073	2.079	18	24	Top Tier	Elite

Table 5. *Top Elite Classification Probabilities (Models 1 and 2)*

5.5.2 Catchers by Absolute Score

Player Name	Elite Probability Model 1	Elite Probability Model 2	Absolute Score (Model 1)	Absolute Score (Model 2)	Defensive Runs Saved	Age	Cluster Label	Performance Tier
Roberto Perez	0.999832	0.999472	2.095	2.116	31	30	Top Tier	Elite
Patrick Bailey	0.998837	0.999280	2.092	2.116	20	25	Top Tier	Elite
Jose Trevino	0.997681	0.997584	2.089	2.110	21	29	Top Tier	Elite

Patrick Bailey	0.997512	0.998269	2.088	2.113	13	24	Top Tier	Elite
Austin Hedges	0.997262	0.994549	2.087	2.101	20	26	Top Tier	Elite
Mitch Garver	0.000000000024	0.000012	-0.916	-0.923	-16	27	Young & Underpaid	Bottom
Willson Contreras	0.000000011	0.0523	-0.916	-0.764	3	26	Young & Underpaid	Bottom
Elias Diaz	0.000000012	0.0035	-0.916	-0.912	-2	27	Young & Underpaid	Bottom
Robinson Chirinos	0.000000017	0.0024	-0.916	-0.916	-9	34	Young & Underpaid	Bottom
MJ Melendez	0.000070	0.000076	-0.916	-0.923	-18	23	Overpaid	Bottom

Table 6. *Catchers ranked by Absolute Score from Model 1 and 2.*

5.5.3 Catchers by Performance Tier (Elite, Middle, Bottom)

Player Name	Elite Probability Model 1	Elite Probability Model 2	Absolute Score Model 1	Absolute Score Model 2	DRS	Age	Cluster Label	Performance Tier
Roberto Perez	0.999832	0.999472	2.095	2.116	31	30	Top Tier	Elite
Patrick Bailey	0.998837	0.999280	2.092	2.116	20	25	Top Tier	Elite
Yadier Molina	0.549	0.569	0.737	0.808	9	39	Middle Tier	Mid-Tier
Brian McCann	0.554	0.443	0.752	0.425	0	35	Middle Tier	Mid-Tier
Austin Romine	0.531	0.305	0.683	0.003	6	29	Middle Tier	Mid-Tier
Mitch Garver	0.000000000024	0.000012	-0.916	-0.923	-16	27	Young & Underpaid	Bottom
Willson Contreras	0.000000011	0.0523	-0.916	-0.764	3	26	Young & Underpaid	Bottom

MJ Melendez	0.000070	0.000076	-0.916	-0.923	-18	23	Overpaid	Bottom

Table 7. *Catchers Grouped by Performance Tier. Classification of catchers into Elite, Mid-Tier, and Bottom based on logistic regression outputs.*

Results from Table 7 show a clear distinction between elite catchers and the rest of the league. From this chart, one can see that the elite catchers typically have a 99% chance of being placed in the elite category, while the mid-tier catchers range from 30%-57%, and the bottom-tier catchers have a minuscule chance, close to 0%. This highlights the strength of the model in predicting elite catchers. The relationship between predicted probabilities and actual DRS scores is visualized in Appendix E. Notably, elite catchers such as Roberto Perez and Patrick Bailey consistently demonstrate both high elite probabilities and high standardized scores across both models, reinforcing their classification.

Middle-tier catchers, however, present a more nuanced picture. Players like Yadier Molina, Brian McCann, and Austin Romine have Absolute Scores from Model 1 between 0.68 and 0.75, only slightly above the mean. These values suggest that while these catchers are not consistently dominant defensively, they demonstrate some above-average skills and may contribute situational or veteran value to a roster. Yadier Molina is a potential Hall-of-Famer at the position, and the data from this model only captures the tail-end of his career. It would be interesting to see where he would've ranked if we had more years of available Statcast data and follow him during the prime of his career. The inclusion of players like Molina in the middle tier underscores one of the limitations of any current model: performance classifications are constrained by the window of available data. Still, the presence of the middle tier reflects meaningful defensive contributions and highlights the values of consistency and leadership, qualities not always captured solely through metrics. The middle tier serves as a reminder that defensive value can exist on a continuum, and that catchers who aren't in the elite threshold still offer important on-field contributions.

The logistic regression output includes a predicted probability of elite classification, those that belong in the top 25%. That is, how often the model predicts that specific player to be in the elite category. Based on the model, Roberto Perez performed the best, scoring a 0.999 for both elite probability models, one and two. He had an absolute score of +2.095, placing him firmly among the elite. In contrast, Mitch Garver has an elite classification probability of just 0.0025 and an Absolute Score from Model 1 of -0.91, making him one of the worst performers in the model. His salary for this data point was $5 million, indicating a large discrepancy between his contract and the defensive performance that the team was getting from him.

Other examples show how the model's insights can inform evaluation. Patrick Bailey and Jose Trevino, two highly regarded defensive catchers, appear near the elite classification threshold in both models with scores at 2.088 and 2.092, respectively. Interestingly, Patrick Bailey had an elite probability of 0.998 in the first model, but once factoring in age and contract, his elite classification probability decreased to 0.81. This could be due to his young age (24) and his low contract average annual value, where the model predicts younger players will not perform as well as those who are in their athletic prime and a couple of years older.

These outputs not only highlight individually elite players but also provide actionable insights for front office decision makers. By combining absolute score and elite classification probability, teams can evaluate both raw performance and their alignment with contract value. Players such as Roberto Perez and Jose Trevino stand out as defensive assets, while Mitch Garver's results flag a potential overpay. These insights can guide contract negotiations, trade evaluations, and roster construction by identifying undervalued performers and cost inefficiencies.

Moreover, the strong agreement between high absolute scores and elite probabilities reinforces the integrity of the model itself. By standardizing inputs and applying PCA, the model reduces noise from unit discrepancies and inter-variable correlation, enabling the logistic regression to make more robust classifications. The drop in Bailey's elite probability in Model Two illustrates how contextual features like age and salary adjust perceived value, a critical component when evaluating efficiency. This is a prime example of how the model can identify elite talent before they reach their athletic prime and command major contracts. The San Francisco Giants, in theory, should benefit from years of high-level defensive production from Patrick Bailey at a fraction of the cost.

5.6 Salary Implications

Table 8 below breaks down the average annual contract value, number of innings caught, Absolute Score from Models 1 and 2, average age, and average elite classification probability for each cluster. This model not only distinguishes elite catchers from the rest of the league but also reveals meaningful inefficiencies in how teams compensate catchers across performance tiers.

Cluster Label	**Mean AAV ($M)**	**Innings Caught**	**Absolute Score (Model 1)**	**Absolute Score (Model 2)**	**Average Age**	**Average Elite Classification Probability**
Middle Tier	2.28	383.4	-0.18	-0.22	29.2	0.24
Overpaid	3.62	590.7	-0.48	-0.45	29.2	0.15
Top Tier	4.35	861.4	1.09	1.11	28.2	0.67
Young & Underpaid	1.02	805.1	-0.92	-0.88	28.5	0.00

Table 8. *Cluster-Level Averages for Salary, Playing Time, Defensive Performance, and Elite Probability (Model 1)*

The Top Tier elite cluster defines what elite looks like in Major League Baseball. These catchers average the highest salaries ($4.35 million), play the most innings (861.4), and have the highest Absolute Score from Model 1(1.09). The probability that they are placed in the elite category, according to Model 1, is also the highest at 0.67, meaning these players are consistently recognized by the model as top defensive performers. This group makes up 25% of the recorded data; around 100 catchers over the past seven seasons are grouped into the elite tier, and 64 distinct catchers within this category. This indicates that while elite defensive performance is rare, it is not limited to a handful of names, instead, there is a rotating pool of top-tier defenders that emerge across seasons. This reinforces the importance of year-to-year evaluation of players and highlights how the model can adapt to changing player performance, rather than relying on the name value of a player alone.

The Overpaid cluster tells the opposite story. Despite averaging $3.62 million per year, these catchers log fewer innings (590.7) and have a negative absolute score (-0.48). Their average elite classification probability is the lowest among the four groups, indicating that teams are likely overpaying based on name value, reputation, or perhaps a single standout defensive season that was not sustained. Another possibility is that these players are being compensated for offensive production, with teams opting to tolerate below-average defense in exchange for a high-impact bat at the catcher position. This tradeoff illustrates how roster decisions can skew toward offensive metrics, even at critically defensive positions such as catcher.

The Middle Tier group likely represents many backup catchers and lower-tier starters. On average, they log significantly fewer innings (383.4) than the Top Tier catchers. This group is paid a modest $2.28 million, have a near-neutral absolute score from Model 1(-0.18), and an elite classification probability of 0.24. While slightly higher than the Overpaid group, this is still relatively low. It's possible that this tier includes a few undervalued or underutilized players, for instance, a capable catcher stuck behind an established starter who hasn’t had a full opportunity to showcase their defensive skills. In general, however, these players appear to be serviceable but not game-changing, and their compensation reflects their role as depth options rather than key contributors.

The most interesting group to point out is the Young & Underpaid cluster. Despite only averaging $1.02 million in salary, these players caught on average 805.1 innings. This is similar to the Top Tier of catchers. However, the absolute score from Model 1 is negative, at -0.92, and the average elite classification probability is 0.00, suggesting they still need development at the position. For analysis, it is important to note that this group only contains four players, one of them being 34-year-old catcher, Robinson Chirinos. This shows that unsupervised clustering grouped the players based on statistical similarity rather than age alone. A key limitation of the clustering model, cluster labels don’t always align perfectly with real-world expectations. With the small sample size, if a team can identify one or two players poised to break out, there could be a massive return on investment for that player.

Taken together, this analysis supports two key insights from the model. First, elite defensive catchers are rare, and the model successfully identifies them using objective Statcast metrics. Second, organizations aren't perfect, and teams still overpay players who they feel are deserving of that contract, even though their statistical output may not be worth the salary. Appendix G illustrates the surprising relationship between performance tier and compensation. By directly comparing pay with performance, MLB front offices can improve their decision making, whether it be optimizing roster payroll or identifying cost-saving opportunities at each position, especially at a crucial position such as catcher.

6. Contributions & Implications

This study makes both theoretical and methodological contributions to the literature and provides front offices with a practical approach for evaluating catchers defensively in the MLB, and other leagues.

6.1 Theoretical Contributions

This research expands on the existing literature by providing a framework with which catchers, and eventually other positions, can be evaluated using machine learning. The model developed in this study uses PCA, clustering, and logistic regression to understand better what makes a catcher elite defensively and how that performance aligns with their contract value. While sabermetrics has made huge strides in offensive evaluation, defensive analysis has lacked consistency and lagged in terms of development and depth of statistics available.

Catcher is a unique and complex position. Unlike other positions in baseball, they are involved in every pitch through framing, blocking, throwing, and managing pitchers, all of which are difficult to quantify. Traditional fielding metrics, as previously mentioned, fail to capture the nuances of the catcher position, since baseballs aren't typically hit at catchers, unless it is a bunt. Even newer metrics, such as DRS or framing runs are helpful additions, but they don't paint the full picture. From this complexity, few models have attempted to comprehensively evaluate catcher defense in a way that combines performance data with player context. This is why teams are looking for analysts to evaluate catchers using machine learning models. This area is still evolving, and this research offers a starting point that others can build upon. As new tracking technologies and defensive metrics develop, this framework can be revised and expanded to provide further insights into the position.

As technology continues to advance, more teams are adopting business intelligence to guide decision-making. This shift is creating new opportunities for organizations to use analytics in their approach to roster construction. This framework not only helps quantify performance at a crucial defensive position but also sets the stage for broader use in different leagues and positions, making it a valuable addition to the growing field of sports analytics.

6.2 Practical Contributions

From a practical and managerial perspective, this paper and model can serve as a valuable tool for MLB front offices in optimizing contracts and rosters. The model successfully clustered players based on defensive performance and then layered in age and contract data to assess whether players were living up to the value of their deals. These results present actionable insights and are especially useful in helping teams make informed decisions about how to allocate playing time, evaluate return on investment, and identify players who are potentially over- or under-performing relative to their contract value

MLB teams could use this tool internally to audit their current roster and past contracts. It can help flag players who are exceeding expectations and those who may no longer justify their cost. It can also support long-term decision-making by identifying early signs of decline in performance and helping teams determine when it is time to either extend, trade, or release a player. Teams are actively seeking analysts to build models such as this and find new ways to evaluate catchers using advanced defensive metrics and machine learning.

This is especially valuable in today's game, where small-market teams must extract maximum value from every roster spot. The way the MLB is set up, there is no salary cap and significant payroll disparities between teams. This means that identifying undervalued players is crucial for small-market organizations for them to remain competitive with big-market teams.

6.3 Methodological Contributions

From a methodological perspective, this study contributes to machine learning and analytics research by combining principal component analysis, k-means clustering, and logistic regression to build a comprehensive defensive evaluation model for every qualifying catcher over the past seven seasons. While previous research has typically focused on offensive statistics or pitcher performance, defensive analysis, especially for catchers, has been limited. Most existing defensive studies focus on fielders, as it's easier to analyze their opportunities (e.g., balls in play), where catchers' contributions are more complex and less visible with traditional stats.

7. Conclusions & Future Research

Catcher defense remains one of the most difficult areas to quantify, often overlooked by traditional metrics. The position is incredibly nuanced, and some statistics, such as Defensive Runs Saved and Ultimate Zone Rating aren't perfect. They offer some insights into the quality of the catcher's defense, but don't capture the full scope of the catcher's responsibilities. Evaluating framing, blocking, throwing, and pop time independently is essential to understanding their overall defensive value.

To solve this, the study introduced a unified, machine-learning-based model using Statcast data to measure catcher performance more holistically. Beyond statistical clustering and elite classification, the inclusion of age and salary variables adds important context. These features help assess roster efficiency and return on investment, both of which are critical for MLB front offices.

Results showed that the strongest predictors of elite defensive performance came from actual game metrics, not related to age or salary. This shows the importance of focusing on what players do on the field, and not necessarily correlating age with a decline in performance. However, adding contract and age features helps reveal which players were overperforming or underperforming relative to expectations set by the team or contract. That insight can help teams optimize payroll, build the most competitive roster, and plan for the future.

This study is not without limitations, and a plethora of opportunities for future research. First, adding offensive metrics into the analysis could further segment players into all-around hitters, power hitters, and ability to get on base. One can also compare elite offense and elite defense, and if teams are willing to take weaker defense for a catcher who is an elite hitter and compensate them properly. Also, the scope of the data is limited. Since Statcast was only introduced during the 2015 season, there are only ten seasons of complete ball-tracking data. As time passes and more seasons are logged, future research could track a catcher's entire career arc using advanced metrics. This would allow researchers to study aging curves, performance peaks, and decline phases in greater detail, and potentially forecast when a player is most likely to break out or begin to regress defensively. Lastly, more advanced modeling techniques, such as ensemble methods or neural networks could be explored to test whether they improve predictive accuracy over logistic regression while maintaining that same level of interpretability. These models could also be adapted to incorporate time-series data to observe how defensive metrics change across seasons, or within the same season.

Defensive Runs Saved as a benchmark is not without its limitations as well. DRS captures a player's total defensive value. It does provide an estimate of how well one player compares to another at the same position, on average. While DRS accounts for play difficulty and is run-weighted, it relies on fielding "buckets" to categorize batted balls, which fails to capture how far a fielder must move to make a play. This is an increasingly relevant shortcoming in the era of strategic defensive positioning. DRS is relative to positional average, so one needs to factor in that some positions are harder to play than others. For catchers in particular, DRS does not fully capture critical responsibilities such as framing, or game calling. Pitch framing influences strike zone control and run prevention but is largely excluded from DRS and WAR metrics (Petriello, 2019). DRS also does not work well for small sample sizes, the data is quite variable (Slowinski 2010), so one needs a large sample size to accurately analyze DRS. While DRS offers a useful aggregate view, its shortcomings highlight the needs for complementary models, particularly when evaluating a nuanced position such as catcher.

Analytics and machine learning are still evolving, with new tools and sources of data emerging each season. As technology advances, future models will be able to capture even more nuance in player performance, offering deeper and more precise insights that continue to push the boundaries of how teams evaluate talent. In the future, there may be a definitive model to evaluate catchers defensively. In the meantime, this model adds to the conversation around how we evaluate catcher defense and provides a strong starting point for MLB front offices to build from. It directly answers the question posed by the Philadelphia Phillies' R&D department: "*The R&D department has been asked to identify the best defensive catcher in baseball. What models would you build to answer that question, and how would you apply those models to decision-making?"*

References

Alamar, B. C. (2013). Introduction to sports analytics. In *Sports analytics: A guide for coaches, managers, and other decision makers* (pp. 1–23). Columbia University Press. http://www.jstor.org/stable/10.7312/alam16292.5

Baccellieri, E. (2023, June 2). How MLB pitch tracking works: Behind baseball's complex system. *Sports Illustrated*. https://www.si.com/mlb/2023/06/02/mlb-pitch-tracking-statcast-explained

Baseball Biographies. (n.d.). *How Hard is it to be a Catcher? A Comprehensive guide to the challenges and skills required*. How Hard Is It to Be a Catcher? A Comprehensive Guide to the Challenges and Skills Required – Baseball Biographies.

Baseball Reference. (n.d.). *Catcher stats and player profiles*. https://www.baseball-reference.com/

Baumer, B., & Zimbalist, A. (2014). An overview of current sabermetric thought I: Offense. In

The sabermetric revolution: Assessing the growth of analytics in baseball (pp. 38–56). University of

Pennsylvania Press. http://www.jstor.org/stable/j.ctt13x1njg.6

Baumer, B., & Zimbalist, A. (2014). An overview of current sabermetric thought II: Defense, WAR, and strategy. In *The sabermetric revolution: Assessing the growth of analytics in baseball* (pp. 57–84).

University of Pennsylvania Press. http://www.jstor.org/stable/j.ctt13x1njg.7 Casella,

P. (2015, April 21). Statcast: Baseball will never be the same. *MLB.com*.

https://www.mlb.com/news/statcast-primer-baseball-will-never-be-the-same/c-119234412

DeForest, W. (2022). *Modeling the performance-based compensation of MLB catchers* (CMC Senior Theses, No. 3155). Claremont McKenna College. https://scholarship.claremont.edu/cmc_theses/3155

Dewan, J. (n.d.). *The Fielding Bible*. Sports Info Solutions. https://fieldingbible.com/

Hayduk, T., III. (2022). Are "tech-savvy" owners better for business? Evidence from Major League Baseball. *Journal of Sport Management*, *36*(6), 559–574. https://doi.org/10.1123/jsm.2021-0252

Jolliffe, I. T., & Cadima, J. (2016). Principal component analysis: A review and recent developments. *Philosophical Transactions: Mathematical, Physical and Engineering Sciences*, *374*(2065), 1–16. http://www.jstor.org/stable/24760364

Kelly, L. (2019). *Sabermetrics in baseball: A casual fan's guide.* Major League Baseball. Sabermetrics in Baseball: A Casual Fans Guide

Koseler, K., & Stephan, M. (2017). Machine learning applications in baseball: A systematic literature review. *Applied Artificial Intelligence*, *31*(9–10), 745–763.

https://doi.org/10.1080/08839514.2018.1442991

Major League Baseball. (n.d.). *Catcher blocking leaderboard*. Baseball Savant.

https://baseballsavant.mlb.com/leaderboard/catcher-blocking?players=693307-2025-116&selected_idx=0 Major League Baseball. (n.d.). *Catcher framing leaderboard*. Baseball Savant.

https://baseballsavant.mlb.com/catcher_framing?year=2025&team=&min=q&type=catcher&sort=4,1

Major League Baseball. (n.d.). *Catcher throwing leaderboard*. Baseball Savant.

https://baseballsavant.mlb.com/leaderboard/catcher-throwing

Major League Baseball. (n.d.). *On-base percentage (OBP)*. https://www.mlb.com/glossary/standardstats/on-base-percentage

Major League Baseball. (n.d.). *Pop time leaderboard*. Baseball Savant.

https://baseballsavant.mlb.com/leaderboard/poptime

Major League Baseball. (n.d.). *Slugging percentage (SLG)*. https://www.mlb.com/glossary/standardstats/slugging-percentage

Marchi, M. (2011, June 10). Evaluating catchers: Quantifying the framing pitches skill. *The Hardball Times*. https://tht.fangraphs.com/evaluating-catchers-quantifying-the-framing-pitches-skill/

Mizels, J., Erickson, B., & Chalmers, P. (2022). Current state of data and analytics research in baseball.

Current Reviews in Musculoskeletal Medicine, *15*(4), 283–290. https://doi.org/10.1007/s12178-02209763-6

Morris, S. P. (2014). Deception in sports. *Journal of the Philosophy of Sport*, *41*(2), 177–191. https://doi.org/10.1080/00948705.2013.785419

National Baseball Hall of Fame and Museum. (n.d.). *Baseball history, American history and you*. https://baseballhall.org/discover-more/stories/baseball-history/baseball-history-american-history-and-you

Petriello, M. (2019, February 6). A baseball analysis primer: Fielding and Defensive Runs Saved. *Battery Power*. https://www.batterypower.com/2019/2/6/18202678/baseball-analysis-primer-fielding-defensiveruns-saved-sabermetrics

Sarlis, V., & Tjortjis, C. (2020). Sports analytics—Evaluation of basketball players and team performance. *Information Systems*, *93*, 101562. https://doi.org/10.1016/j.is.2020.101562

Slowinski, S. (2010, August 25). What do we know about catcher defense? *FanGraphs*. https://library.fangraphs.com/what-do-we-know-about-catcher-defense/

Spotrac. (n.d.). *MLB player contracts and salaries*. Retrieved May 10th 2025, from https://www.spotrac.com/mlb/

Statcast. (n.d.). *Statcast-Baseball Savant.* Statcast Leaderboard | baseballsavant.com.

Wolfe, R., Babiak, K., Cameron, K., Quinn, R. E., Smart, D. L., Terborg, J. R., & Wright, P. M. (2007). Moneyball: A business perspective. *International Journal of Sport Finance*, *2*(4), 249–262.

Appendices

Appendix A. Final Catcher Dataset

The full dataset used in the study. This includes catcher-level defensive metrics, demographic information, and the output from the PCA model, including explained variance, cluster, cluster labels, elite probability from Models 1 and 2, Absolute Score, and Performance Tier.

Final Catcher Dataset

Appendix B. Modeling Notebook (PCA, Clustering, Logistic Regression)

The full Jupyter Notebook used for PCA dimensionality reduction, K-means clustering, and logistic regression, as well as some descriptive visuals, is included here.

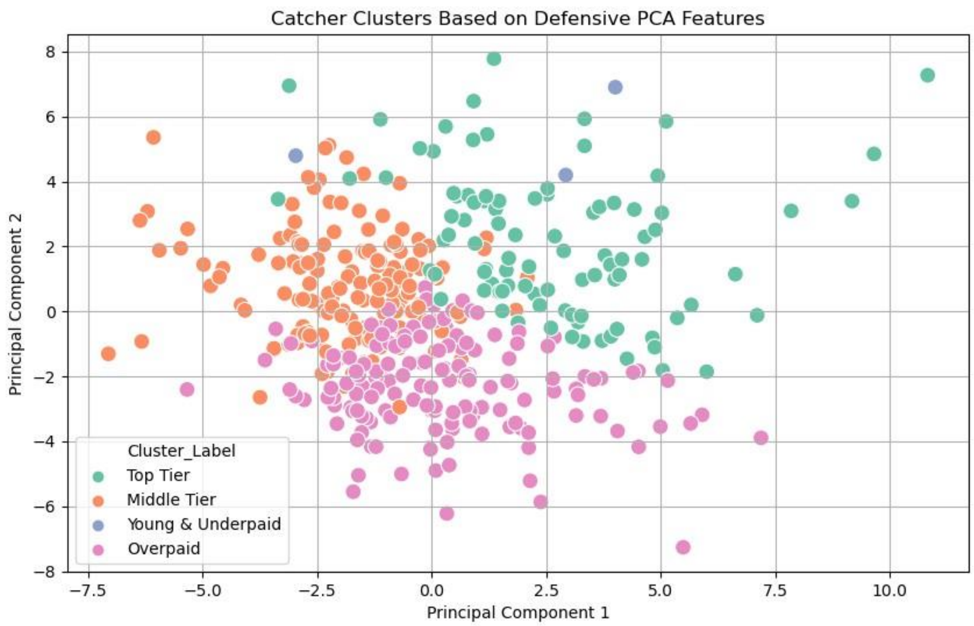
Clustering, PCA & Logistic Regression Notebook

Appendix C. Initial Model Setup & Elbow Method Notebook

This notebook shows the initial model setup. This includes a correlation matrix and histograms of key descriptive variables, the Elbow Method for determining the optimal number of clusters for k-means, and some initial result graphs before finalizing the model.

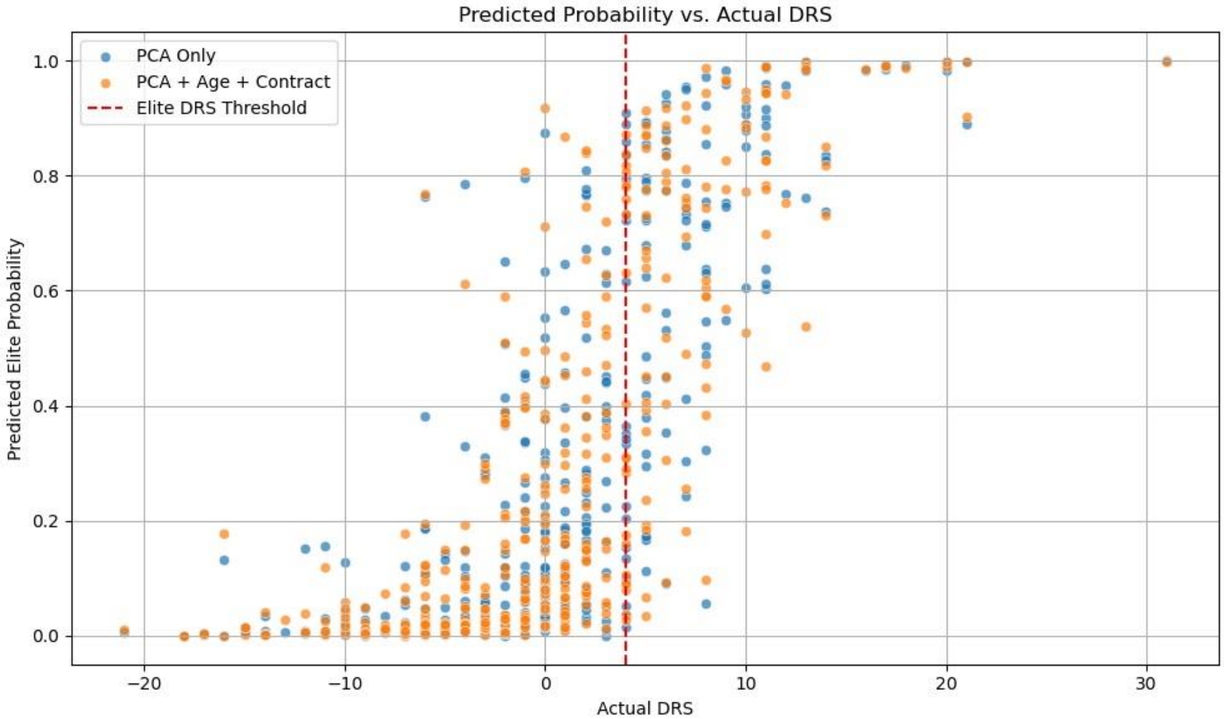
Preliminary Notebook **Appendix**

D.



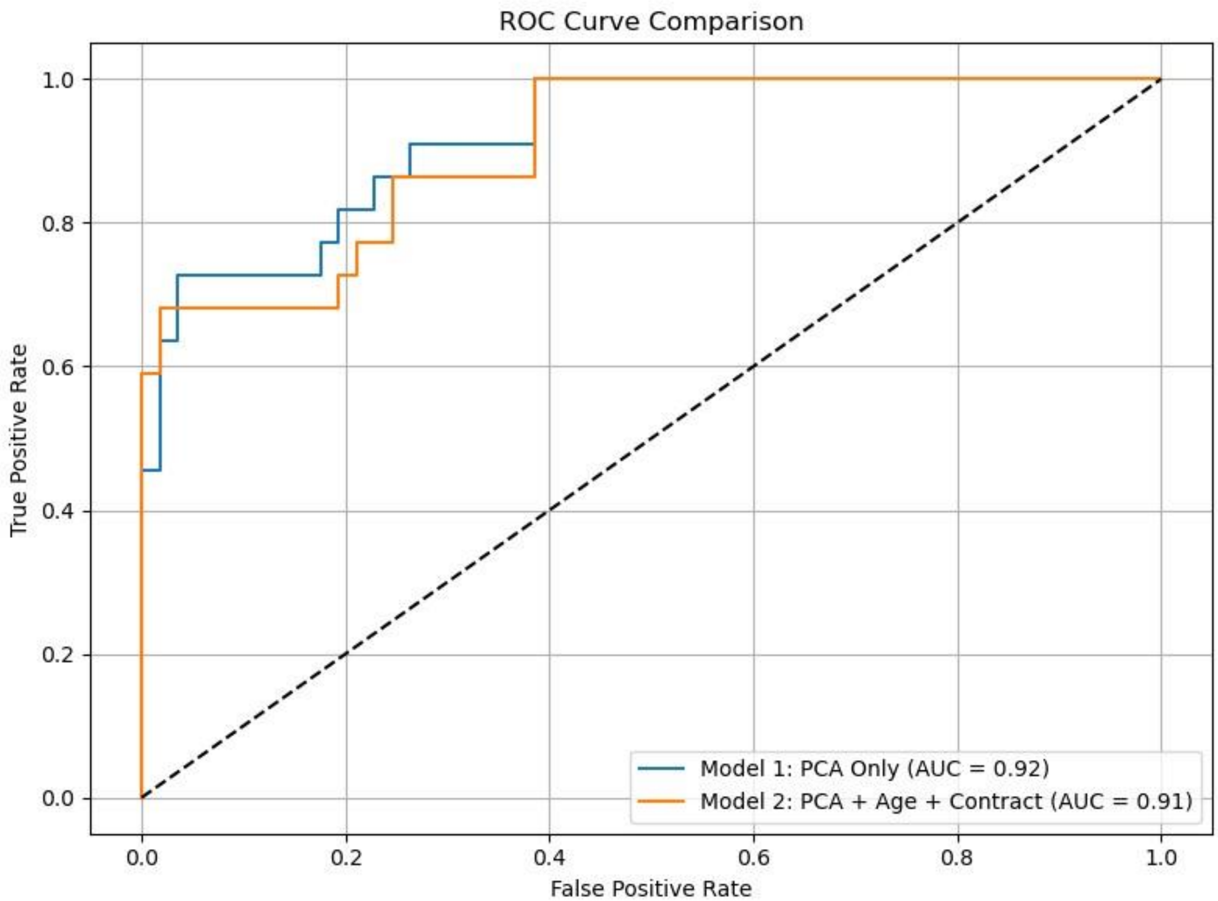
A visual depicting the K-means clustering results in PCA space. Each catcher is plotted based on their score with the top two principal components and grouped by Cluster Label. This figure highlights distinct defensive profiles across the league, with most catchers clustered close together, indicating the difficulty in analyzing defense.

Appendix E.



This figure highlights the model results of predicted vs actual for Defensive Runs Scored from both logistic regression models. The vertical red line represents the elite DRS threshold (top 25%), used to define elite status in the model.

Appendix F.



This figure is the Receiver Operating Characteristic (ROC) curves comparing Model 1 (PCA only) and Model 2 (PCA + age + contract) for elite classification. The AUC values indicate that Model 1 achieves slightly better classification performance than Model 2.

Appendix G.

This bar chart displays the average contract value (AAV) for catchers grouped by performance tier (Elite, Mid-Tier, Bottom) based on logistic regression classification. Surprisingly, the mid-tier defenders earn more on average than elite defenders, an interesting observation when it comes to defensive value at the catcher position for MLB front offices.