CREATING A PREDICTIVE PRICING MODEL FOR NATIONAL FOOTBALL LEAGUE TRADING CARDS

By

Susan Amanda Lucas

May, 2025

Director of Thesis: Nic Herndon, PhD

Major Department: Computer Science

ABSTRACT

Collecting trading cards is a decades-old hobby and fascination. Since statisticians were able to start assigning pricing values to cards, based on rarity of the card and the popularity of the player, the once casual hobby has transformed into a serious one, with some collectors focusing on financial prospecting. Currently, this "prospecting" and literature on the National Football League trading cards focuses on previous purchase prices and a subjective inference of player performance or popularity to "predict" whether a card could be a positive or negative investment. This project works to better understand correlates of average card values from a more comprehensive view of the industry: player's physical attributes, card condition, NFL performance metrics, and draft performance metrics are considered. This paper shows strong positive relationships between passing touchdowns, passing yards, passing completion, passing attempts, passing interceptions, weighted value for drafting team, approximate weighted career value, draft pick, sacks, and Pro Bowl selection ($R^2 > 0.60$ for all features) with average selling price of a trading card. Overall, ensemble learning produced the best model optimization when selecting the best model from Gradient Boost, XGBoost, Random Forest, and KNN outcomes. Stacking models led to increased explanation in variance and to reduced errors for all positions except Safety. MAE remained optimized for the Defensive End, Defensive Tackle, Linebacker, Offensive Lineman, Quarterback, and Running Back positions with XGBoost models.

CREATING A PREDICTIVE PRICING MODEL FOR NATIONAL FOOTBALL LEAGUE TRADING CARDS

A Thesis

Presented to The Faculty of the Department of Computer Science

East Carolina University

In Partial Fulfillment of the Requirements for the Degree

Master of Science in Data Science

By

Susan Amanda Lucas

May, 2025

Director of Thesis: Nic Herndon, PhD

Thesis Committee Members:

Venkat Gudivada, PhD

David Hart, PhD

Kebin Peng, PhD

ACKNOWLEDGEMENTS

I would like to thank Dr. Nic Herndon for serving as the Director of my Thesis Committee, for all of his support, and for all of his assistance in completing this project. I would also like to thank Dr. Venkat Gudivada, Dr. David Hart, and Dr. Kebin Peng for serving as Thesis Committee Members. Finally, I thank my family for their relentless support and tireless motivation.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

Collecting National Football League (NFL) trading cards is a decades-old hobby that has increasingly shifted from the casual hobby collector to an adult-oriented style of prospect collecting, often aimed at increasing financial investment. The "value" of these cards took on a quantifiable number once statisticians, such as Beckett, worked to create algorithms meant to assign values to cards capturing their rarity and market demand Beckett [1984]. NFL trading cards have evolved from the "premium with purchase" commodity, first offered with tobacco products and gum, into a dynamic market of collecting, trading, and prospect selling. Increased demand from card consumers helped drive a hobby towards investment status with the introduction of variations, autographed cards, and several special-edition releases. In fact, several selling platforms, such as eBay, offer "vault" options to their consumers where traders buy and sell the card digitally while the card remains physically vaulted by brokers to protect the condition of the card as it "changes hands." While demand ultimately drives the price a consumer is willing to pay, many factors drive this sense of demand. Very little research has attempted to examine potential predictors of card values. It is the goal of this study to move towards a predictive model of NFL trading card values.

Prior research helps us understand how the current market for NFL trading cards developed: "In 1888, nestled within packs of Old Judge and Gypsy Queen cigarettes were the very first American football cards. As some of the oldest football cards, they featured sepia-toned images of college players in reserved, portrait-like poses" GiantSportsCards [2024]. Primm et al. [2008] suggest that when Camel cigarettes appealed to the quality of their tobacco blend over the premiums or coupons (including football cards), they gained tremendous market share, challenging the hobby of collecting and trading football cards included in Duke tobacco products. The authors write: "The hobby, however, had a revival in the middle of the twentieth century, this time as a part of the chewing gum industry. Sports cards became part of the battle between rival companies in their quest for customers with the Topps Chewing Gum Company (Topps) and the Bowman Gum Company (Bowman) signing the biggest players in baseball and football." This re-emergence helped fuel a large spark in the trading card industry.

The NFL collectible market has changed tremendously since the emergence of the first football cards and many factors have contributed to the increasing value of football cards over recent years. Johnson [2023] suggests that the increased player and sport popularity contributed to huge increases in values. Jackson [2023] adds that cards are higher quality than they used to be, and values generally increase over time, but are likely sensitive to economic conditions that influence general spending. Graded cards, rookie cards (RC), age of card, number of cards printed (often serialized), popularity of the card set, and demand increases value (Johnson; Jackson). Primm et al. [2010] write on the topic of what drives the value of some trading cards over others for a collector: "What factors affect a card's value in the minds of collectors? Some are related to the cards themselves, such as card condition, the number of specimens available, and the card's age, while others are related to the athlete pictured: the player's level of skill or status and achievement. Additionally, there are other, more 'intangible' factors, including a player's off-the-field behavior." The authors find that a card's availability, age, performance of a player, Hall of Fame status, Quarterback position (followed by Offensive Linemen, Running Backs, and Wide Receivers), and Pro Bowl participation had significant relationships with card values.

While many factors contributing to card value have been discussed, there remains a gap in predictive, data-driven modeling of NFL trading card prices. This thesis proposes to address that gap by developing machine learning models that predict rookie NFL card prices based on a unique combination of player performance metrics and card attributes. This study aims to move beyond subjective valuation methods in order to find a scalable, interpretable framework for forecasting trading card prices.

1.1 Background: Predicting Card Values

In searching how to predict NFL card values among popular social media sites, it is common to encounter numerous reports on how individuals price their cards and predict future values – mostly this is based around player performance in response to their current statistics and prior purchase prices. Many videos and blogs weigh in on how to invest in trading cards – which cards to keep and which to let go. This is a very popular topic that does not have a good scientific approach to determining and predicting pricing. Thousands of blogs, websites, and videos are dedicated to the pursuit of understanding how cards will increase, or decrease, in value. Much time is dedicated to predicting which cards will be long-term investments.

Aldrich and McKilips [2022] used simple linear regression to try to find the best player statistic to predict the price of a card: they found that basketball cards are the highest priced cards – average top basketball card priced at $57K. Football follows as the second highest priced category of sports cards (average price of top card over $43K), then baseball ($34K), and hockey ($18K). The authors found that player awards, such as Most Valuable Player (MVP), are often associated with higher card values – 3.9 Football MVP's is the average among the ten highest priced football cards. Rookie cards tend to be more valuable with an average of $37K for RC versus $22K for non-RC in football. Many studies and social media posts have argued that player cards for players with increasing numbers of team wins are also associated with higher card values. DisplayZone [Zone, 2023] attributes higher values to the 2000 Tom Brady rookie cards: numerous Super Bowl wins and collector recognition of "significance of owning a piece of history with a Tom Brady card." Aldrich and McKilips find that the following will drive increased values: rookie card status, number of championships won, and MVP/award totals by player. They found that 2021 saw 14 different sports cards sold at auction for over a million dollars. Aldrich and McKilips also found that: across the top 20 highest priced cards included in their study, the players on these cards held an average of five and a half MVP's and that 72% of the top 200 cards were RCs. Further, the number of championships won and MVPs held were the two most important predictors of price – players on the top 200 cards have won an average of three championships and four MVPs.

DisplayZone similarly found that more modern examples of higher value rookie cards followed close trends: the 2017 Patrick Mahomes rookie card outpaces many peer cards due to his success in the league and for the player's ability to captivate fans; Lamar Jackson's 2017 RC will also likely be highly valued due to exceptional performance, multiple MVP awards, and agility; and that new RC Quarterbacks are emerging with great potential and popularity: CJ Stroud and Anthony Richardson are "superstar Quarterbacks who could also see their trading cards gain value over time." DisplayZone and Johnson have made predictions about what could drive the value of a card, including player performance, card rarity, card autograph, serial numbers, card condition, and player popularity. Claiming that social media and online trading platforms have helped boost the market, they ultimately state that player performance will be the biggest driver in card value. DisplayZone claims that higher values for 1998 Peyton Manning's rookie card from 1998 stem, at least in part, from the player having an "illustrious career as one of the most accomplished Quarterbacks in NFL history. His induction into the Pro Football Hall of Fame further solidifies his legacy and increases interest in his early cards."

Many blogs, websites, and videos offer pricing strategies and information. One such example, nooffseason.com [Hickey, 2020], is a card network that is entirely dedicated to football card pricing, research, and selling tips. Like other sites, this site claims to use correlates of player performance statistics and web platform (eBay in this case) pricing values. However, player performance statistics seem to be pulled based on either excellent player performance within the week of consideration or from increases in sales price or cards offered for sale. This pricing model is not historical, is very sensitive to immediate changes in performance, and introduces selection bias when considering which cards to evaluate. Selection bias is inherently introduced when an author uses current game performances as it relies on game coverage pre-determined by networks. This can overlook player performance in a less popular time slot or playing for an underdog team not covered as popularly as other players on more popular teams. While logic informs us that popularity should matter for card values, we can only confirm this by including all player performance data over time. No pricing strategy reviewed among social media posts in this study included a methodology for incorporating correlates of player performance across multiple games, NFL Combine performance, popularity, team performance, or awards won. It is important to note that Combine performance will be referred to throughout this paper in reference to a national scouting event designed to test various athletic and mental skills of an incoming draft class. Any reference to Combine data will always be presented as capitalized.

Many of the sites included in this study review common features: card variations, player injury status (a very time-pertinent consideration), number of cards sold in a specified period, autograph and patch status, raw or graded status, number of bids on a card, player rank in role, and popularity (often measured as "star" performance in a game or season). Often, these card values are reviewed within a season year (here defined as the pre-season, regular season, and post-season playoff cycle) and are very sensitive to short time frames for player performance. For example, nooffseason.com looks at a player's performance in a week and then compares improved or degrading performance statistics to the value of a card after a game to values from within the same season. Language such as "Thursday night game's big star efficiently touched the ball only 12 times with week-leading passing yards" illustrates how bias selections occur around who is perceived as a standout player. This overlooks long-term gains and losses to card values, especially over a longer period.

Johnson argues that market trends are influenced by seasonal fluctuations, with peak values occurring during the football season. Johnson recommends searching eBay for selling history to get a preferred “live” price over pricing guides. Port [2022] is among a strong support base claiming that Panini is the strongest NFL trading card brand: “Topps is probably the best-known brand in valuable baseball cards, but they lost their license to produce football cards a few years ago. This paved the way for Panini to become the largest and most popular brand in the sports card industry. As a result, Panini has been on a spending spree – buying up other sports brands left and right, like Donruss Playoff.” Barone [2023] argues that Panini’s Prizm is the best of the best and should be thought of as the best Panini cards around. Given the incredible number of available set options, this project will follow popular research and look exclusively at Panini Prizm cards. Numerous supporters argue that Prizm is the best Panini set, including Sports Cards Specialist [Clarke, 2025]. Pricing information for NFL trading cards in social media is largely driven by previous sales. Industry giants such as SportsCardsPros [SportsCardsPro, 2020], eBay [eBay, 1995], PSA magazine [PSA, 1991], Mavin [LLC, 2016], Beckett monthly magazine, and 130PT [Pearce, 2016] are heavily utilized for pricing information and for selling hubs. However, these prices are not tapping into the predictive potential of the market and this project seeks to explore how a value-predicting algorithm may or may not successfully predict NFL trading card values.

1.2 Purpose of Study

Prior work suggests that certain features should influence a card’s value. For card attributes, it has been suggested that card availability, card condition, card autographs, and card grading should play a role in card values. Prior research also indicates that several features related to players should also influence card values: rookie year, collegiate performance, draft performance, drank rank, awards won, Pro Bowl selection, popularity, games played, and Body Mass Index (BMI). Although prior work has not explored an intersection of these features, it is thought that doing so will contribute a comprehensive model of card pricing to the field of card and draft valuation which can help more accurately and efficiently predict player performance in a professional league and corresponding trading card pricing. This study presents a novel approach to predicting NFL trading card prices by combining structured card characteristics with player performance, draft data, and physical attributes across multiple seasons. The goal is to develop interpretable, position-specific machine learning models that forecast card prices more reliably than subjective, trend-based assessments common in the collectibles market. To achieve this goal, I will build a card pricing database from card pricing data points collected through a subscription to Card Hedge [2020]. This database will be merged with a statistical player database with data from Sports Reference [2000]. I will then work to develop Machine Learning algorithms that can offer enhanced predictive ability. I will examine results from Random Forest, KNN Regressor, Gradient Boost, and XGBoost models with various approaches to scaling, hyperparameter tuning, and handling of missing values. Ensemble learning will then be applied for potential model optimization. Finally, an exploration of the most influential machine learning features by position will help highlight which give us increased predictive ability. Results from this study will show that XGBoost will give the best pre-stacking results and that stacking models offer the best fit for our model overall. This research will show that key predictive features include weighted career value, draft value to the drafting team, Pro Bowl selection, autographing, grading, and transaction count, affirming both performance-based and collectible-specific factors as strong drivers of average selling price. In order to evaluate model performance, several measures will be examined. Model performance is assessed with Mean Absolute Error (MAE) to assess average prediction error, Root Mean Squared Error (RMSE) to evaluate prediction variance, and R^2 to quantify how well a model explains variance in card prices. This study has several broader implications for future work - performance features from college games can be used to help minimize draft class costs to recruiting teams – more efficient evaluation of skills can lead to better understanding of how certain players will contribute value to a drafting team. Data science can help facilitate a deeper understanding of how collegiate performance and Combine event features can build increased value for NFL teams. Feature analysis should also help with enhanced fantasy football drafting as player valuation is based on similar key features.

Chapter 2

Related Work

Currently, there is no existing public dataset that hosts historical pricing information or a comprehensive player feature dataset in an accessible environment. Pricing in practice relies heavily on the amount a card sells for over a historical period across different selling platforms. Another dataset will rely on combining many other player variables to attempt a large-scale player feature dataset. DisplayZone writes: "The value of a card is determined by a number of factors, including the player's performance, the rarity of the card, and whether or not the card is autographed." DisplayZone argues that player performance is likely the most important driver of price. While a bit of literature suggests similar approaches, locating research on how player performance is measured, how it serves as a predictive factor in statistical pricing methodologies, and statistical testing for whether it is a stronger pricing predictor than another factor has not been reviewed with scientific methodology. While the field is missing a comprehensive predictive pricing model, much research has been done to predict player performance and draft class decisions based on in-season player performance.

Lewis [2003] and Hakes and Sauer [2006] found a strong correlation between Major League Baseball (MLB) performance success among athletes with higher on-base percentages. The search for improved data analytics across sports has been an on-going effort. Hakes and Sauer, in their examination of player salaries to their on-base and slugging percentages, demonstrate that player's salary in correlation to their contribution to the team's winning percentage is undervalued by the MLB. The 2020 MIT SLOAN Sports Analytics Conference [Eager and Chahrouri, 2020] posits that the accurate valuation of a player has been relatively well-addressed in other sports, but "has largely been unsolved in football, due to unavailable data for positions like Offensive Line and substantial variations in the relative values of different positions." MIT Sloan also suggests that the number of distinct plays, incredible number of unique operators who run those plays uniquely, and the general noise of the sport make it more difficult to value player potential than in other sports.

MIT SLOAN, Yurko et al. [2019], and Citrone and Ventura [2017] have focused on a model of player valuation that uses normalized data for player performance to attempt a player valuation based on projected player wins in comparison to a replacement-level player in the same position. Barney et al. [2013] used the recognition factor of a player and games played and started as a measurement of player performance and found that Centers, Guards, and Kickers are undervalued while Cornerbacks tend to be overvalued.

2.1 A Case for College Performance as an Indicator of NFL Success

Prior research has been done around predicting player performance and the valuation of a player with respect to their college performance. The ability to accurately predict a player's pre-NFL performance has been largely under-researched. However, popular culture items such as Moneyball [Lewis, 2003] have brought attention to the use of applied analytics in the professional sports world. While academic studies have reviewed several potential predictors of NFL success, a large body of work explores how college performance can predict player success in the NFL. Much opinion work within the social media world suggests that performance plays a role in informing card values, so logic would suggest that prior collegiate performance could play a role in determining the NFL performance of players, which may, in turn, inform card values. Barbatsis et al. [2024] found that college completion percentage correlates strongly with NFL completion percentage: "This is because passing profiles (taking shots down the field versus check-downs) and decision-making remain relatively consistent throughout a player's career." When exploring college performance metrics as predictors of NFL Quarterback (QB) ratings, the authors found that player height and college completion percentage reliably predict QB rating. They found that while not statistically significant, touchdowns, interceptions, air yards per attempt, and yards per attempt show low p-values and contributed to a more successful model. They also found draft pedigree and average points per game to be strong indicators of NFL success. Interestingly, Barbatsis et al found Wide Receiver success to be much harder to predict than Quarterback success.

2.2 Combine Performance and Draft Valuation of Players

One approach to determining the predictive value of NFL players is to evaluate the potential productivity of potential incoming players during the draft season. Brower [2020] uses aggregate data from the NFL Combine and Pro Football Focus to attempt a measurement of player's first-year success in developing a player valuation model that highlights the possibility that performance on the 40-yard dash (a popular metric from draft Combines) may be over-utilized without merit. Similarly, Pitts and Evans [2019] found that NFL teams are undervaluing BMI, receiving yards, and vertical jump for Wide Receivers and early exit for Tight Ends in the Combine. They also find that draft position correlates with higher or lower productivity but suggest that teams should ultimately rely on traditional valuation methods and analytics as a supplement to best make draft decisions.

Berri and Simmons [2011] found a weak correlation between Quarterback NFL performance and draft day performance. The authors attribute this to the noise of the sport, which makes it hard to predict performance in what they term "the uncertain environment of professional sports." In a study of surplus value, the difference between the projected economic value and the compensation cost of a player, Massey and Thaler [2005] found that surplus value peaks in the second round of a draft. As seen in similar studies, the authors evaluated surplus value by comparing the draft cost of a rookie player to the cost of each veteran free agent existing in the league. Massey and Thaler suggest that over-evaluation by the team managers leads to an overcompensation of players drafted in the first round and a half of a draft class. Hendricks et al. [2003] argue that the hiring process of athletes out of the college system via the NFL draft is mostly about evaluating uncertainty around predicted player performance. They find that schools with greater resources to spend on training and coaches, what they term "the Bowl Subdivision," tend to be viewed as less risk-adverse by recruiters due to the level of visibility these players receive during their collegiate seasons. Hendricks et al find that players from less visible teams, who are picked later in the draft as "undervalued players," tend to have better careers than their peers from the Bowl Subdivision. Berri and Simmons and Hendricks et al attempt to evaluate the recruitment of all positions and explore variables such as years played in the NFL, Pro Bowl appearances, games played, and games started. While calling attention to the nature of football being a team sport and the likelihood that "noise" such as fellow- and opposing- team influences on performance, Berri and Simmons look at such variables for Quarterback health as height, BMI, and 40-yard dash. They also look at such Quarterback performance variables as completions, passing yards, touchdown passes thrown, interceptions thrown, Quarterback rating, net points, wins produced, and passing attempts. The authors found a strong relationship between draft position and number of plays a Quarterback runs. They write: "Our study of subsequent NFL performance – which was hampered by a lack of data – failed to find that the Combine factors had much of an impact on future performance. In essence, NFL decision-makers can be impressed by taller, smarter, and faster signal callers. But there is no evidence that the extra inches, better test scores, or faster 40-yard dash times make any difference in subsequent NFL performance. Indeed, Berri and Simmons find that draft pick is not a significant predictor of NFL performance."

Reynolds et al. [2015] found that draft pick value (based on first-round selections and overall number of incoming draft picks) is positively correlated with win percentage for a team. They also found that higher draft pick values shared a significantly positive relationship with making the playoffs three years after a draft class. This prior work points to building an inclusive dataset that will include player draft class information (health and Combine performance), multi-game performance metrics, team performance metrics, awards won, injuries, games played and started, championship wins, playoff participation, and position. Prior research has indicated that predicting performance may be position-specific and for the purpose of this study, player position will be examined in relation to developing a predictive pricing algorithm.

Pitts and Evans [2019] find that Wide Receiver performance on metrics such as BMI, receiving yards, and vertical jump is associated with significantly better draft position and greater expected NFL productivity. They find that Combine performance for Wide Receivers is not significantly associated with draft performance but is a strong predictor of number of NFL games played, suggesting that Wide Receivers on college teams with great offensive talent may be underestimated in predicted NFL productivity. Pitts and Evans [2018] found in prior research that earlier draft round pick positions for Running Backs, Tight Ends, and Wide Receivers better predict increased NFL productivity than draft pick for Quarterbacks – suggesting an over-evaluation of Quarterback Combine and college productivity as a predictor of NFL performance.

Chapter 3

Data

3.1 Data Overview

This project first sought to create a database for NFL trading card selling values to include Rookie player name, card number, whether the card was autographed, card variations, grading status (Raw, PSA-graded "9" and PSA-graded "10" cards), and historical data for dates sold and selling prices. It is important to note that the database created for this project was acquired through a subscription to Card Hedge [2020] and will not be publicly available due to the website use agreement. A card can be submitted to companies, such as PSA or Beckett, for grading by professional evaluators. These evaluators determine the condition of a card and encapsulate the card in a plastic card protector labeled with the card's rating and information. PSA is considered the standard in the trading card industry due to its large presence and brand consistency. Data was only readily available for raw cards and for cards graded by PSA as "9" or "10." Data for the creation of this dataset comes from Card Hedge, which is among the only sources of historical card values and certainly the most comprehensive source utilizing a wide web search for access to this information. At the time of this research, there is no comprehensive public dataset on trading card values over time and finding historical pricing data can prove daunting. Large-scale historical NFL trading card pricing information was only accessible through a paid subscription to Card Hedge, a pricing platform that allows users to search and track specific cards. Once a specific card iteration is selected, the site provides useful selling price points in the form of a line graph visualizing pricing over time. This project will conceptualize a card "iteration" as the sales information for a single card for one player with a specified card year, base/parallel/variation status, and graded status at a time. Such iterations allow control for physical attributes of a trading card that can influence pricing independent of player draft and league performance. Once a specific card iteration is specified on Card Hedge using a graphical user interface search on the website, a paying consumer can select to receive output card performance values for selling values over a month or longer. Users can even select an output format from multiple options. This project utilized a comma-separated values (CSV) output option for ease of compiling data. The creation of the dataset for this project required a good deal of manual input to bring data values for multiple players and card iterations into a consolidated dataset for analysis.

Values for card prices were collected into one dataset for this project by manually searching by Rookie player name, choosing variation of card, adjusting the time period for inclusion to include all available values, and selecting each pertinent grading per card. It was then necessary to download CSV files for each of these iterations. A Python script was written to then merge all CSV files together while creating a column for player name and whether the card was an autograph or not. This essentially was the beginning of the initial dataset and as all Rookie players within the respective year span were gathered, it was possible to then merge those together into a larger year file. When four years of data were collected, the files were finally merged together to create a final raw dataset with all pricing variables.

3.2 Set Selection

There are multiple manufacturers of NFL trading cards, and these cards are released under many sub-names such as Bowman, Donruss, Immaculate, Leaf, Panini, Topps, and Score. Each manufacturer also offers different card sets for overlapping player lists in a single year. For example, in 2023, Panini offered: Panini Contenders Optic Football (base set size of 207 cards), Panini One Football (base set size of 260 cards), Panini Select Football (set size of 500 cards), Panini National Treasures Football (set size of 184 cards), and Panini Flawless Football (set size of115 cards). This means that for the most popular players, a single player could have a rookie card from at least five different Panini base card sets in 2023. The sheer number of different card iterations for a player from different manufacturers grows exponentially when considering that each manufacturer issues different versions of card releases in a single year. Add to this that each of these card set releases can have over 60 potential variations per player card into the equation, we are approaching thousands of card price points for a player's rookie card value. This project will create a dataset only from the Panini Prizm NFL trading card line as this is a brand and line that has been available for many years and is arguably the most popular NFL card choice. Future research should evaluate how other trading card lines and brands perform relative to Panini Prizm, but for the sake of narrowing a potentially endless scope, this project will exclude them.

First, official card set lists were obtained from the official Panini website to identify which cards were offered each year for the Panini Prizm card set. These values included here were obtained through a semi-automated web scraping process on Card Hedge and included selling prices for Panini Prizm NFL rookie trading cards for the years 2020 to 2023. Cards released in earlier years (2020 versus 2023) have data points across the 2020-2023 included range and longitudinal review is possible. For cards released in 2023, the study can only review a year of trading history. Card Hedge houses data values for how much NFL trading cards were sold over time through eBay, PWCC (Fanatics.com), mySlabs, and ALT. Most data appear to come from eBay and Fanatics, but all available sales points were included.

3.3 Rookie Card and Variation Selection

Rookie cards were chosen as the focus, as prior research has demonstrated that they are typically the most valuable and most frequently collected – this would provide the most data points for analysis. Every player with a NFL rookie card has a base version of the card, which is the most common expression of any given NFL trading card. Variations from the base card are popular among collectors as they are rarer than the base version. The Silver variation is the most common variation, which is still much rarer than the base card, and should typically have higher selling prices than the base versions. Certain card iterations are made in lesser quantities. As a card variation becomes increasingly rare, fewer data points are available. This is inherently the case as there are fewer of these cards to sell. Originally, during the initial 2023 data collection process, values were collected for all variations of each player's card. However, variations were not consistently included in the trading card set for every year and rarer variations were only produced for a subset of players. For example, the 2023 card for Bijan Robinson included 62 variations. Beyond the number of variations, card pricing has been submitted for graded cards and for autographed cards. This project initially included all data points with the Panini Prizm Rookies of 2023, but found that only a few variations were popularly traded across enough players to be included for evaluation: Base, Silver, Disco, Lazer, and Red White Blue. Each of these variations represented at least 7.9% of the total data values from the 2023 data set.

Figure 3.1 shows a breakdown of variations per year for 2023. Variations were examined for their data points and most variations make up less than 1% of the data. This is not surprising as only more popular players – those thought to be popular or important players – are going to have rarer variations, autographs, and graded cards. Furthermore, many of the lesser-known players, who are less likely to have seen game time play, are not likely to have selling cards. Players who do not play, are not popular, or are not in popular positions are not going to be an essential part of a collection. Future research should examine the "rarer" variations that were excluded from this study.

3.4 Player Position Selection

Figure 3.2 highlights the number of unique players in each position for the selected card set. It shows that there are positions with extremely limited data points, which make the position a poor candidate for assessment in learning model selection. Kicker and Nose Tackle

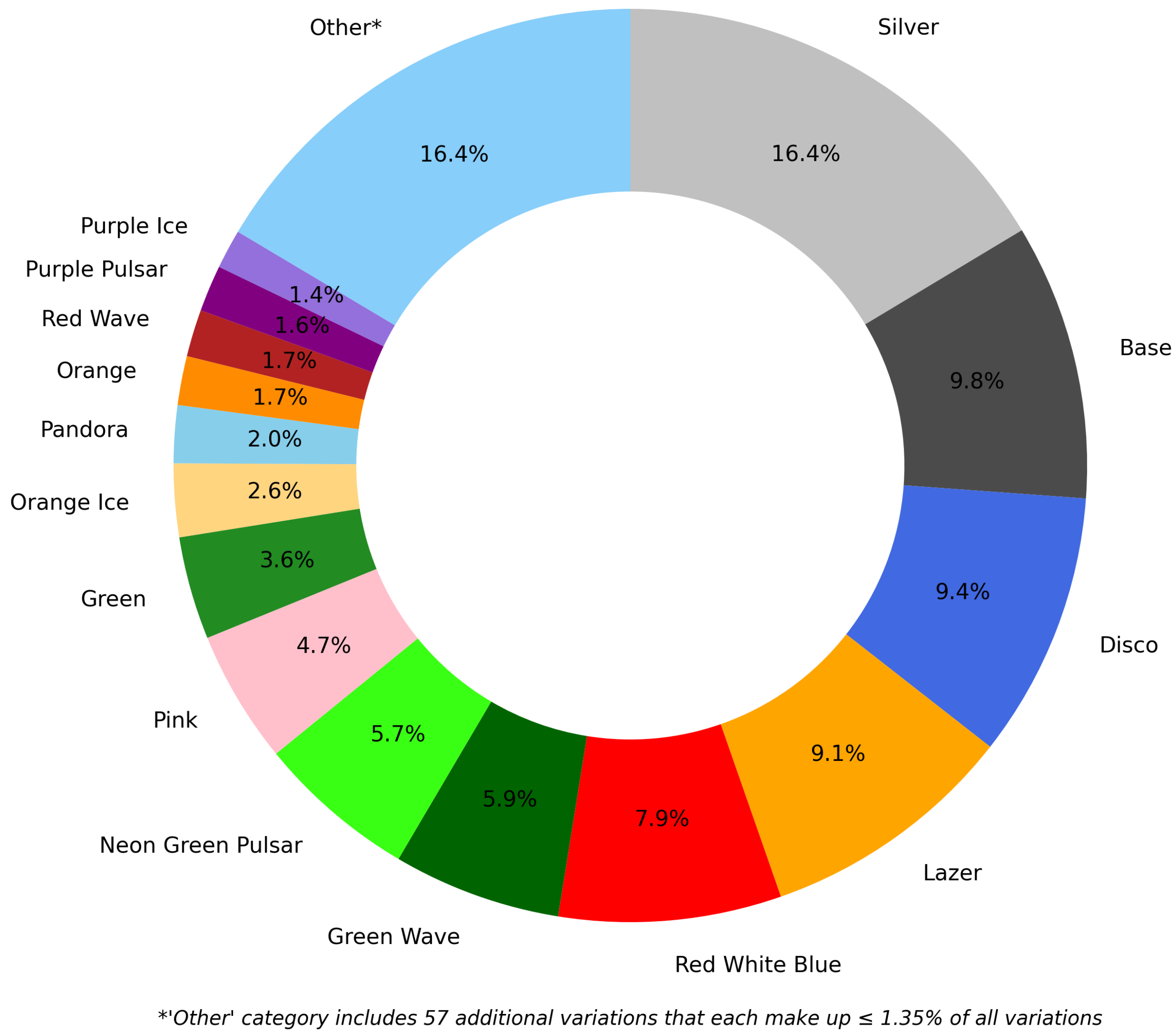

Figure 3.1: 2023 Initial Dataset Variation Percentages. This figure presents the percentage distribution of variation transactions for 2023 NFL Panini Prizm cards. The data informed the selection of the top five card variations—Silver, Base, Disco, Lazer, and Red White Blue—for inclusion in the dataset, as these variations accounted for the highest transaction volumes and represent the most popular types among collectors.

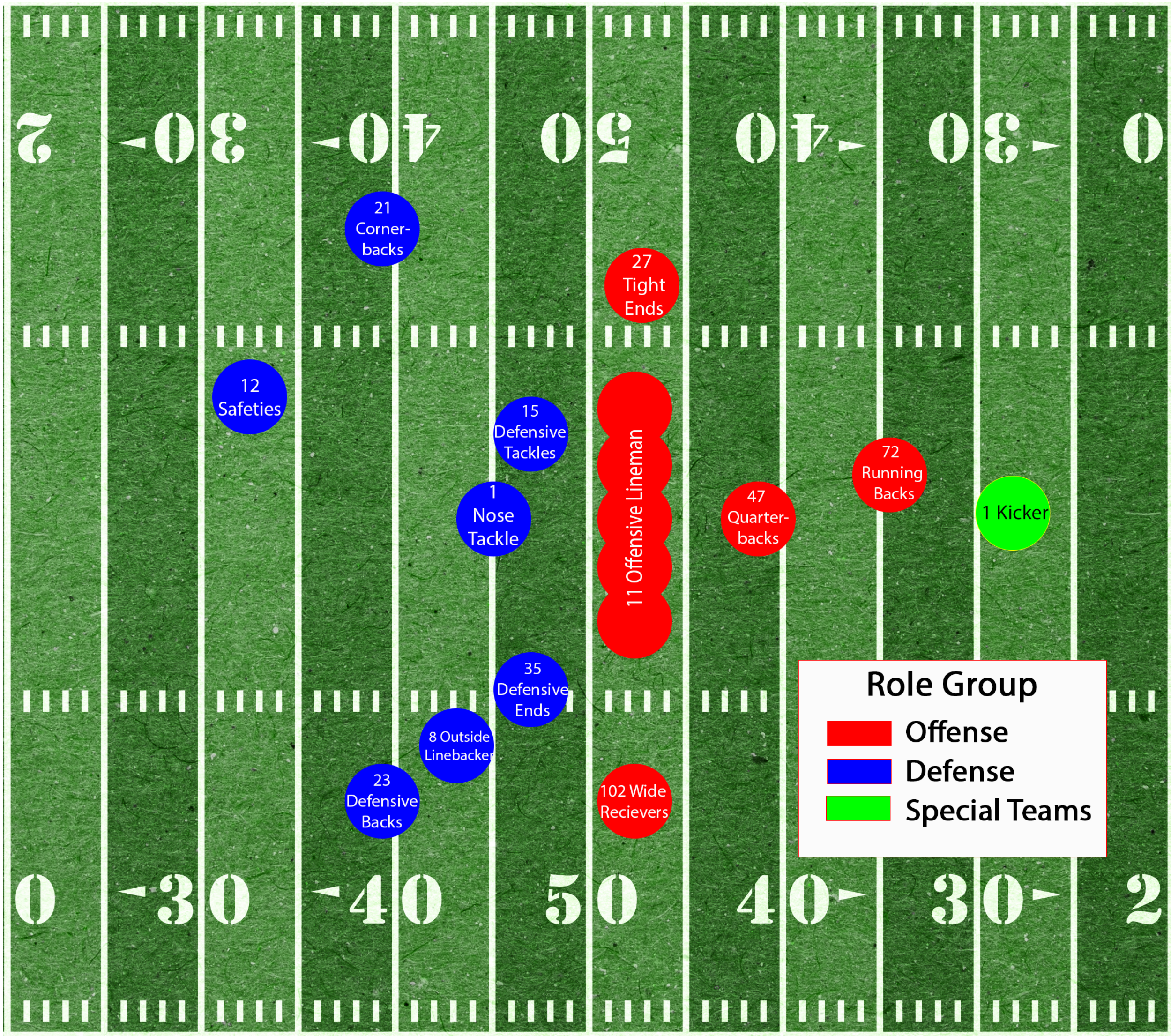

Figure 3.2: Unique Player Counts by Position. Offensive positions accounted for the majority of players (259), while defensive positions contributed 151 players. Only one player represented special teams (a kicker), which was insufficient for inclusion in the machine learning models due to the lack of data.

were only represented by one player each and were excluded from analysis for model parsimony. The dataset includes card values for the following positions: Cornerback, Defensive Back, Defensive End, Defensive Tackle, Linebacker, Offensive Lineman, Outside Linebacker, Quarterback, Running Back, Safety, Tight End, and Wide Receiver. NFL trading cards can be submitted to card grading companies, such as PSA and Beckett, for card grading and encapsulation. Cards are given scores up to 10, which would be considered a "mint" condition card in the best possible condition – the card has no flaws to the print, corners, finish, etc. Card Hedge includes data on "raw" cards that have not been submitted for review and on cards that have been graded as a "10" or "9," which would be the more valuable graded versions of a card. Lower grades are sold, but they are much less frequent and would only pertain to the most popular players. These values are not available and are not included as they could confound the impact of other features.

3.5 Performance, Draft, and Physical Attribute Selection

Prior research helped determine which features to include in the model. This research suggests that the draft process and Combine data produced from events held annually before college and high school players are drafted into the NFL provide key metric indicators for NFL success. Much research illustrates that NFL success is a major factor in driving card values. This study also utilizes performance data from the NFL to build as holistic a model as possible.

Player performance data comes from pro-football-reference.com, including Combine data, NFL performance, and draft data. All data points on performance were included for review in this study. For a full list of measures included, please refer to Figure 3.3. Player draft and Combine performance is included for the year in which a player's rookie card is released as the rookie card is released during the player's rookie season. Player performance data during the season and pre- and post-season is available for the rookie year and any subsequent years included within the scope of the study.

PERFORMANCE FEATURES USED IN DATA SET:

General:

- **Rnd** -- Round selected in draft
- **Pick** -- Overall Selection in Draft
- **Pos** -- Position
- **Age** -- Age as of September 1 of the draft year
- **To** – Last Year
- **AP1** -- First-team all-pro selections
- **PB** -- Pro Bowl Selections
- **St** -- Number of years as primary starter for his team at his position
- **Approx Val wAV -- Weighted Career Approximate Value --**
  The career AV is computed by summing 100 percent of the AV of his best season, 95 percent of the AV of his next-best season, 90 percent of the AV of his third-best season, and so on
- **DrAV** -- Weighted AV accumulated for team that drafted this player
- **G** -- Games played
- **College/Univ** – Most Recent College/University

Passing

- **Cmp** – Passes completed
- **Att** – Passes attempted
- **Yds** -- Yards Gained by Passing. For teams, sack yardage is deducted from this total
- **TD** – Passing Touchdowns
- **Int** -- Interceptions thrown

Rushing

- **Att** -- Rushing Attempts (sacks not included in NFL)
- **Yds** -- Rushing Yards Gained (sack yardage is not included by NFL)
- **TD** -- Rushing Touchdowns

Receiving

- **Rec** -- Receptions
- **Yds** -- Receiving Yards
- **TD** -- Receiving Touchdowns

Tackles

- **Solo** -- Tackles
- Before 1994: unofficial and inconsistently recorded from team to team. For amusement only. 1994-now: unofficial but consistently recorded.

Def Interceptions

- **Int** -- Passes intercepted on defense
- **Sk** -- Sacks (official since 1982, based on play-by-play, game film and other research since 1960)

Figure 3.3: NFL Performance Metrics Reference. This figure lists the performance-related features included in the dataset and provides a brief explanation of what each metric measures.

3.6 Cleaning and Missing Values

Once datasets were created for pricing values, Combine metrics, and performance measures, merging was necessary to create a final dataset, which introduced issues to address. One of the first issues identified was the use of different names for the same players – generally, this was due to the inclusion of nicknames or generational suffixes in different data sources. These had to be matched up to the correct players. To accomplish this task, Python was used for fuzzy matching with the `get close matches` function in class `difflib`.

Additionally, player Combine, draft, and performance data had to be linked to the correct year for each player. Performance data outside of the draft and Combine types were longitudinal and had to be matched up. Performance data was included per player for all years to create multi-year summaries for performance measures. This data was available on Sports Reference [2000] by draft class.

Combine performance data was also introduced into the data set from Sports Reference by annual Combine events. These had to be matched to each player. Missing data for Combine performance were then reviewed and cleaned. Additional data points for missing data were manually located for inclusion. See Table 3.1 for a missing data summary across performance features. Features such as height were transposed from feet and inches to just inches as a more usable format. Each position was evaluated for potential data points that would help drive predictive insight. Different positions were also evaluated for performance indicators by key features. See Table 3.2 for features included by position.

Categorical data expressed in categorical columns, such as college attended or draft team, had to be converted into numerical form. In anticipation of machine learning, one-hot encoding was necessary for the preprocessing of certain features such as college/university, draft team, and card variation.

Table 3.1: Missing Data Summary. While each position under review has different metrics of relevance, some will not meaningfully associate with certain metrics. For this reason, it will be important to utilize position-specific models when trying to predict card values.

Feature	**Missing Values**	**Percentage Missing**
Passes Intercepted on Defense	3232	83.82
Sacks	2988	77.49
Bench	2245	58.22
3Cone	2225	57.70
Shuttle	2151	55.78
Solo Tackles	1438	37.29
Broad Jump	1053	27.31
Vertical	939	24.35
Draft Pick	891	23.11
40yd	712	18.46
Weighted Value for Team That Drafted	210	5.45
Weighted Career Approx Value	155	4.02
Interceptions Thrown	149	3.86
Passing Touchdowns	149	3.86
Yards Gained By Passing	149	3.86
Passes Attempted	149	3.86
Passes Completed	149	3.86
Rushing Yards Gained (sacks not included)	130	3.37
Rushing Attempts (sacks not included)	130	3.37
Receiving Receptions	130	3.37
Receiving Yards	130	3.37
Receiving Touchdowns	130	3.37
Rushing Touchdowns	130	3.37
Round Drafted	99	2.57
Years Played	85	2.20
Years as Primary Starter	80	2.07
Pro Bowl Selections	80	2.07
1st Team All Pro Selections	80	2.07

Table 3.2: 2020–2023 NFL Performance Data by Position. This shows that performance features are not available for all positions. For example, "Passing Yards" is a performance metric that is aptly used as an indicator of Quarterback performance, but it is not an appropriate metric for Tight End performance. "Rushing Yards" is a good performance metric for Tight End performance and could also be a useful indicator for Quarterback performance.

Performance Metric	CB	DB	DE	DT	K	LB	NT	OL	OLB	QB	RB	S	TE	WR
Passes Completed										X	2			X
Passes Attempted										X	X			X
Yards Gained By Passing										X	2			X
Passing Touchdowns										X				2
Interceptions Thrown										X	1			1
Rushing Attempts (sacks not included)										X	X	1	X	X
Rushing Yards Gained (sacks not included)										X	X	1	X	X
Rushing Touchdowns										X	X		X	X
Receiving Receptions						1		1		X	X		X	X
Receiving Yards						1		1		X	X		X	X
Receiving Touchdowns										X	X		X	X
Solo Tackles	X	X	X	X	X	X	X	2	X	X	X	X	X	X
Passes Intercepted on Defense	X	X	X	X		X	X		X	1		X		
Sacks	X	X	X	X		X	X		X			X		

**Red notes the number of unique players that have this metric*

Chapter 4

Methodology

4.1 Methods Overview

To examine how to best predict a card's average price or maximum sold price, many relationships were evaluated between card values and player performance, draft data, and physical attributes. In evaluating how to best operationalize a card's "value," the card value feature was examined for average price sold, maximum price sold, total sales dollars produced by a card, and number of recorded purchasing transactions. I found that for the most popular players, using the maximum sold price, number of transactions, and dollars from recorded transactions produced highly skewed distributions that could create extreme bias in the learning models. As shown in Figure 4.1, correlations are very strong ($r > 0.70$ for all features) for highest price of a sold card and 40yd, vertical, 3cone, and shuttle. More nuanced relationships are illustrated in Figure 4.2, which presents correlations between the average selling price of a card and various player attributes. Interestingly, the direction of some correlations differs from those observed with the highest price, suggesting that the two operationalizations of the dependent variable will demonstrate unique relationships with the same set of independent features. It is important to note that all values are rounded to the hundredths place for ease of reading and interpretation.

Table 4.1 shows strong, positive correlations ($r > 0.60$) between highest price sold for a card and passing touchdowns, passing yards, passing completion, passing attempts, passing interceptions, weighted career value, average value to drafting team, draft pick, draft round,

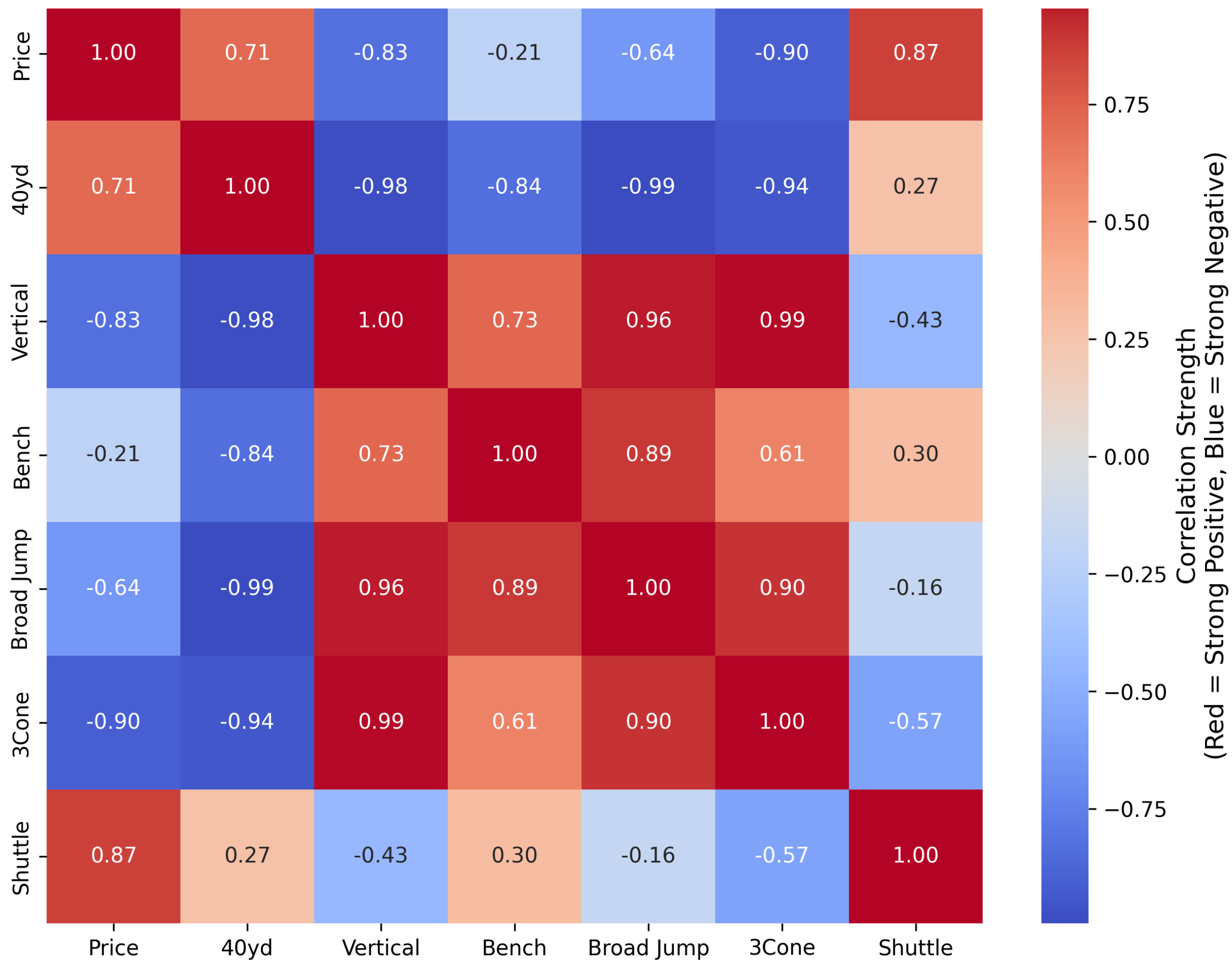

Figure 4.1: Features from the NFL Combine Event with Correlations with Highest Price. Combine correlations are very strong ($r > 0.70$) for highest price of a sold card and the 40-yd, vertical, 3cone, and shuttle.

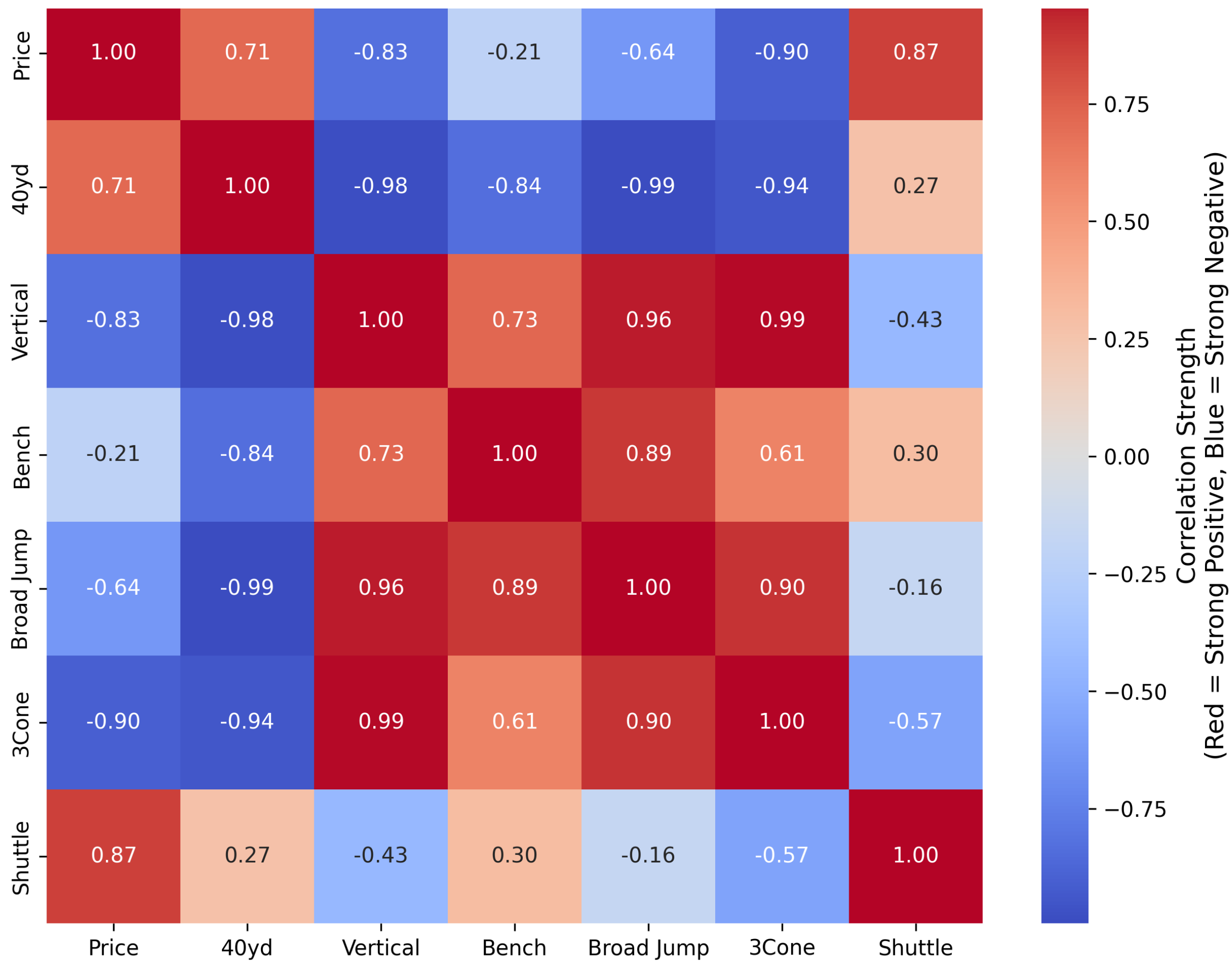

Figure 4.2: Features from the NFL Combine Event with Correlations Average Price. More nuanced relationships are shown in the above figures, where correlations between average selling price for a card and player attributes are shown. The direction of some relationships reversed from highest price to average price, indicating that each operationalization of the dependent variable will demonstrate unique relationships with the same set of independent features.

Table 4.1: Correlations Between Performance Metrics and Card Prices. Passing TD, Passing Yds, Passing Cmp, Passing Att, Passing Int, wAV, DrAv, Pick, Rnd, Sack, PB all demonstrate strong, positive relationships with average price sold $r > 0.60$). Interestingly, negative relationships of moderate strength ($r > -0.30$) suggest that increases in Tackles Solo, Receiving Rec, Receiving TD, and Receiving Yds yield lower average selling prices. The figure illustrates an almost absence of causality between Rushing Att and G ($r > 0.00$) and average price sold. For these reasons, the average price sold and maximum price will be evaluated for model fit.

Performance Metric Feature	**Avg Price Corr.**	**Max Price Corr.**
Passing TD	0.98	0.98
Passing Yds	0.95	0.95
Passing Cmp	0.91	0.92
Passing Att	0.89	0.90
Passing Int	0.87	0.88
wAV	0.71	0.71
DrAV	0.71	0.71
Pick	0.71	0.70
Rnd	0.68	0.66
Sack	0.67	0.49
PB	0.61	0.60
Passes Int	0.31	0.09
Rushing TD	0.27	0.26
St	0.10	0.09
Rushing Att	0.01	0.00
G	0.01	-0.02
Rushing Yds	-0.04	-0.04
Age	-0.06	-0.06
AP1	-0.12	-0.12
Receiving Yds	-0.31	-0.32
Receiving TD	-0.33	-0.33
Receiving Rec	-0.34	-0.35
Tackles Solo	-0.44	-0.44

and Pro Bowl selections. Receiving yards, receiving touchdowns, receiving receptions, and tackles solo demonstrate moderately strong, negative relationships ($r > -0.30$) with highest price sold per card. Table 4.1 also shows a similar absence between rushing attempts and games played ($r > 0.00$) and highest price sold.

The results from exploring correlates of both versions of the dependent variable fall in-line with prior research. This study will utilize machine learning to identify potential predictors of card values. Relationships between card values and many performance indicators were explored.

4.2 Review of Output Target by Position

For model development, a selection process occurred through multiple stages. First, a baseline Random Forest regressor helped examine the performance of the highest price sold target and the average price sold target by position. Using zero or the median to replace missing values was evaluated to determine whether one or the other would optimize model performance – these values are necessary to run a Random Forest regressor. As discussed in the Data review, player positions allow different feature analysis as not all performance metrics are available for each position.

Table 4.2 shows the results of this comparison. Consistent evaluations were used throughout all model iterations. These are Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R^2 Score. These metrics yielded a reliable framework for comparing the effects of changes to preprocessing and model parameters. MAE provides the average of the absolute difference between the actual and predicted values, measuring the average of the residuals. RMSE takes the square root of the mean squared error to facilitate easier interpretation of the units. It gives the standard deviation of residuals (the average difference between the original and predicted values) in the same unit as the dependent variable, a dollar amount to represent a selling price, and indicates how well a model fits the dataset. R^2 is the proportion of variance in the dependent variable and helps with understanding variability in the dependent variable. Ideally, we want to see lower MAE and RMSE values, indicating higher accuracy in the regression model, and a higher R^2 value.

Table 4.2 shows substantially better MAE values for all positions when using average price versus maximum price as the target. RSME and R^2 values were better in all but two positions (DT and LB) and RMSE was very similar. RMSE is a better option than R^2 for comparing accuracy of our models, but both help quantify fit to the dataset. Of interest,

Table 4.2: Baseline Random Forest Regressor Results. The models using Average Price outperformed those using Maximum Price as the target variable in 136 out of 144 performance metrics. In the eight cases where Maximum Price performed better, the largest performance difference was 0.95.

Position	MAE	RMSE	R^2
Cornerback	6.82	12.33	0.64
Defensive Back	3.37	5.82	0.88
Defensive End	3.84	7.90	0.76
Defensive Tackle	5.75	14.56	0.47
Linebacker	9.20	40.27	0.61
Offensive Line	4.09	5.70	0.97
Outside Linebacker	6.16	11.60	0.47
Quarterback	44.11	114.33	0.80
Running Back	7.21	23.91	0.70
Safety	6.69	10.75	0.42
Tight End	7.13	15.69	0.84
Wide Receiver	9.70	19.39	0.78

(a) Average Price (0's fill missing values)

Position	MAE	RMSE	R^2
Cornerback	6.86	12.24	0.64
Defensive Back	3.41	5.89	0.88
Defensive End	3.85	7.84	0.76
Defensive Tackle	5.88	14.63	0.47
Linebacker	8.99	39.76	0.62
Offensive Line	4.41	6.19	0.96
Outside Linebacker	6.18	11.51	0.48
Quarterback	43.29	113.54	0.80
Running Back	7.13	22.87	0.72
Safety	6.43	10.63	0.43
Tight End	7.17	15.22	0.85
Wide Receiver	10.02	20.29	0.75

(b) Average Price (Median fill missing values)

Position	MAE	RMSE	R^2
Cornerback	12.37	18.62	0.53
Defensive Back	10.47	14.85	0.63
Defensive End	10.47	14.85	0.63
Defensive Tackle	6.64	13.61	0.55
Linebacker	14.30	39.70	0.75
Offensive Line	8.90	11.15	0.88
Outside Linebacker	10.20	17.02	0.33
Quarterback	156.24	376.87	0.58
Running Back	21.32	52.15	0.67
Safety	9.07	11.93	0.37
Tight End	23.73	45.35	0.49
Wide Receiver	41.52	161.94	-3.26

(c) Max Price (0's fill missing values)

Position	MAE	RMSE	R^2
Cornerback	12.31	18.43	0.54
Defensive Back	10.33	14.69	0.63
Defensive End	18.87	40.14	0.67
Defensive Tackle	6.66	13.77	0.54
Linebacker	14.22	39.44	0.75
Offensive Line	8.97	11.22	0.88
Outside Linebacker	10.16	17.02	0.33
Quarterback	157.12	377.34	0.58
Running Back	20.68	48.12	0.72
Safety	8.91	11.88	0.38
Tight End	23.48	44.54	0.51
Wide Receiver	40.48	162.53	-3.30

(d) Max Price (Median fill missing values)

*Shading highlights the best-performing results when comparing models.

error increases significantly for the Quarterback position, which may reflect the tremendous variation in what cards sell for in this position. In reviewing the performance of maximum price sold and average price sold by position as our predictive measure, I found that the average price sold was a much stabler measure for our future models.

4.3 Evaluating Machine Learning Models: Random Forest Regressor

Applying an iterative and structured process, the development of the Random Forest Regressor (RFR) progressively refined both the features set and model configuration. The modeling pipeline was implemented in eight steps – each step focused on modifying a single workflow component to isolate its effect and maintain interpretability. The process began with a baseline model trained using `scikit-learn RandomForestRegressor` with default parameters (n estimators = 100, random state=42) and average price as the target variable. Missing values in the feature set were filled using the median value for each respective feature. The "player" feature (each player's name) was dropped from the dataset before running all models.

Although tree-based models like Random Forest typically do not require scaling, feature scaling using `StandardScaler` normalized numerical inputs and was explored for potential impact. As we will discuss in Results, scaling did not have much of an impact in the results and only yielded at most a 0.01 difference in most models. Models were also run to compare filling missing values with zero or the median value. Hyperparameter tuning for model optimization was explored for exploring tree depth (`max depth`), the minimum number of samples required to split a node (`min samples split`), and the total number of estimators (n `estimators`). These were tuned with `GridSearchCV` to efficiently assess multiple combinations. By introducing these changes step-by-step, it was possible to isolate each step for performance evaluation, providing transparency, comparability, and control over model complexity. The results for all model combinations were collected for all positions for later evaluation.

4.4 Evaluating Machine Learning Models: XGBoost Regressor

Also following an iterative process, the development of the XGBoost model refined both the feature set and model configuration for predictive accuracy. The modeling pipeline was implemented through stages, with each stage focused on specific aspects of preprocessing, model training, and performance optimization. This process started with a baseline model using the `XGBRegressor` from the `XGBoost Python library` initialized with default hyperparameters (n estimators $= 50$, `max depth = 3, learning rate=0.1, random state=42`). The target variable again was average price. Missing values in the feature set were handled natively by XGBoost. This is ideal as it can manage missing values during training without explicit imputation. Again, the "player" feature was dropped as it was not used in machine learning.

As XGBoost is robust to unscaled data, feature scaling was not applied. However, the impact of scaling was considered and found to be unnecessary for this tree-based method. Hyperparameter tuning was then introduced to optimize key model parameters, including the number of n estimators, the depth of each tree, the step size at each iteration (learning rate), the fraction of samples to use for each tree (subsample), and the fraction of features to use for each tree (colsample bytree). To evaluate different hyperparameter combinations, first we used `RandomizedSearchCV` with a parameter grid of n `estimators: [50, 75, 100], max depth: [3, 4, 5], learning rate:[0.05, 0.1, 0.15], subsample: [0.8, 1.0], colsample bytree: [0.8, 1.0]` and `Search Setup` was `RandomizedSearchCV (`n `iter=30, cv=5, scoring=`R^2) for the second model.

Interestingly, as attempts were made to enhance the model, such as early stopping and hyperparameter tuning, repeated compatibility issues arose. For example, newer versions of `XGboost (v2.0+)` introduced changes to the `XGBRegressor.fit()` method, resulting in many training arguments, such as early stopping rounds, eval metric, and callbacks, that were unsupported or unstable in combination with `scikit-learning tools`. These issues made further progress impossible when using the standard `XGBRegressor API`. In an attempt to resolve this, the model was transitioned to the native `xgboost.train()` interface, offering lower-level access to the booster. It is also fully compatible with modern versions of XGBoost. This led to the third model. Data was converted into an optimized DMatrix format. Then, `xgb.train()` was used to specify hyperparameters, track validation performance using evals, and apply early stopping. Early stopping was utilized with early stopping rounds=10 to stop training if the model performance on the validation set did not improve after ten rounds to reduce overfitting and computation time. The results for all model combinations were collected for all positions for later evaluation.

4.5 Evaluating Machine Learning Models: Gradient Boosting Regressor

The Gradient Boosting Regressor was selected for splitting the data as it performs well with building an ensemble of weak learners (decision trees) that work together for more accurate predictions. We again created a pipeline of models that would evaluate subtle differences to find various results and to therefore find which model was the best. The model was first initialized with the following default parameters: the number of boosting stages (trees) used in the model (n estimators = 100), the contribution of each tree to the final prediction (`learning rate=0.1`), and the maximum depth of individual trees to prevent the model from overfitting to the training data (`max depth=3`).

Scaling was evaluated, but results varied little with or without scaling. Like Random Forest and XGBoost models, this is a tree-based model. These models are generally not affected by the scale of input features as decision trees split data based on feature thresholds rather than calculating distances or dot products between features. Filling missing values with median or zero was also looked at - this gave us four models which we then used to incorporate `GridSearchCV` for hyper-tuning.

`GridSearchCV` was added to search over a grid of hyperparameters. This tested different values for the following key hyperparameters: n estimators, learning rate, max depth, min samples split, and subsamples. The data was divided into five subsets with `GridSearchCV` using five-fold cross-validation, and the model was trained and evaluated five times. Each time the model was trained, a different subset was used for testing and the remaining subset was used for training. R^2 was used to evaluate the best set of hyperparameters. The results for all model combinations were collected for all positions for later evaluation.

4.6 Evaluating Machine Learning Models: KNN Regressor

The KNN Regressor predicts the target based on the average of the "K" closest neighbors and was initialized with n estimators = 3. The KNN algorithm is well-suited for scenarios where nearby data points are likely to share similar target values and works by calculating the distance between data points in the feature space. Scaling was not performed on the initial training model so that it could serve as a baseline for performance assessment. Two different imputations were tested for handling missing values – filling with zero and filling with the median. After evaluating initial model performance, results demonstrated that KNN models are very sensitive to the scale of the data and `StandardScaler` was applied to address potential issues with such sensitivity. Scaling helps ensure that all features contribute equally to the KNN distance metric so that no one feature with large differences in range could overly influence the calculation.

For model optimization, both `GridSearchCV` and `RandomizedSearchCV` were used to search over a range of hyperparameters: the number of neighbors to consider (tested in a range from 3 to 20 (n neighbors)), distance metrics used for calculating nearest neighbors (including Euclidean, Manhattan, and Minkowski), the weight function used for prediction (either uniform where all neighbors are equally weighted or distance where closer neighbors greater influence). `GridSearchCV` searched all combinations of the hyperparameters, performing a five-fold cross-validation to evaluate model performance across different subsets of the data with negative mean squared error used as the scoring metric. `RandomizedSearchCV` was employed as a faster alternative to `GridSearchCV`, randomly sampling from the hyperparameter space to test a subset of the combinations. This was a more efficient approach that allowed for a larger range of parameter values to be tested in less time. The results of all model combinations were collected for all positions for later evaluation.

4.7 Evaluating Machine Learning Models: ElasticNet and SVM Regressors

It should be noted that I also tried both ElasticNet and Support Vector Machines regressors. These yielded poor results from baselines and therefore were not further evaluated. They struggled with the complex, non-linear patterns in our data. On the other hand, tree-based models (like Random Forest, Gradient Boosting, XGBoost) are non-linear and can therefore better capture trends, outliers and local patterns. Also, KNN regressor is also non-linear and therefore worked when similar items should have similar target values, like two similar players having similar future prices.

Chapter 5

Results and Discussion

5.1 Results of Machine Learning Models

Performance results for the machine learning models are first evaluated by position. This is important as each position has unique player performance and draft features that interact collectively in unique ways. Different models therefore performed with varying efficacy for different positions. The number of model iterations presented in these results corresponds to the number of needed iterations. For the KNN Regressor, Random Forest, and Gradient Boosting models, eight iterations were needed to explore all hypertuning parameter options and to best optimize each model. XGBoost models were optimized with hypertuning in only three iterations, and only two iterations were needed for optimizing Stacking models. It is important to note that these differences will be reflected in subsequent figures capturing Machine Learning Model performance by position.

5.1.1 Cornerback

As seen in Table 5.1, Random Forest modeling using missing values to fill the mean, scaling, and `GridSearchCV` produced the highest R^2 value (0.70) and the lowest RMSE (11.29) for the Cornerback position. MAE (6.56) was the second highest for this position in this model iteration, but it still offers the best performance overall. R^2 (0.08) is lowest with the KNN Regressor models. Error is also the highest with the KNN models (MAE 12.78; RMSE 19.65).

Table 5.1: Five Machine Learning Model Results for the Cornerback Position. Each table presents a different learning model, with columns representing configuration variations. An "X" indicates inclusion of a specific preprocessing step or method. Yellow shading highlights the configuration that achieved the best performance.

KNN Regressor Models		1	2	3	4	5	6	7	8
Cornerback - Average	Missing Values Median	X	X	X	X				
	Missing Values Fill 0					X	X	X	X
	Scaling/Standardize Features		X	X	X		X	X	X
	No Scaling	X				X			
	GridSearchCV			X				X	
	RandomizedSearchCV				X				X
	MAE	10.23	12.78	12.18	11.74	10.23	12.78	12.41	12.20
	RMSE	14.27	19.65	17.62	17.15	14.27	19.65	17.71	17.51
	R^2	0.52	0.08	0.26	0.30	0.52	0.08	0.26	0.27

Random Forest Models		1	2	3	4	5	6	7	8
Cornerback - Average	Missing Values Median	X	X			X		X	
	Missing Values Fill 0			X	X		X		X
	Scaling/Standardize Features		X		X			X	X
	No Scaling	X		X		X	X		
	GridSearchCV					X	X	X	X
	MAE	6.86	6.88	6.82	6.85	6.46	6.45	6.56	6.48
	RMSE	12.24	12.26	12.33	12.35	11.88	11.83	11.29	11.85
	R^2	0.64	0.64	0.64	0.64	0.66	0.67	0.70	0.67

Gradient Boosting Models		1	2	3	4	5	6	7	8
Cornerback - Average	Missing Values Median	X	X			X		X	
	Missing Values Fill 0			X	X		X		X
	Scaling/Standardize Features		X		X			X	X
	No Scaling	X		X		X	X		
	GridSearchCV					X	X	X	X
	MAE	6.48	6.48	6.44	6.44	6.62	6.26	6.48	6.26
	RMSE	12.07	12.07	11.80	11.80	11.73	11.54	11.64	11.54
	R^2	0.65	0.65	0.67	0.67	0.67	0.68	0.68	0.68

XGBoost Models		1	2	3
Cornerback - Average	Simple XGBoost Model (no-tuning)	X		
	XGBoost with Hypertuning RandomizedSearchCV		X	
	XGBoost with Hypertuning Manual Random Search using xgb.train() and Dmatrix.			X
	MAE	6.65	6.81	6.32
	RMSE	11.69	11.84	11.45
	R^2	0.68	0.67	0.69

Stacking Models		1	2
Cornerback - Average	Simple Averaging Ensemble	X	
	Stacked Ensemble		X
	MAE	7.09	6.29
	RMSE	11.90	9.94
	R^2	0.66	0.77

5.1.2 Defensive Back

For the Defensive Back position, as shown in Table 5.2, R^2 is highest (0.89) and RMSE is lowest (5.57) with the XGBoost model with hypertuning manual random search using `xgb.train()` and `DMatrix`. Again, we find that the corresponding MAE (3.59) is not the lowest value. It is lowest with the Random Forest model that replaces missing values with the median average card price, utilizes scaling, and applies `GridSearchCV` (MAE 3.31). R^2 (0.29) is again the lowest and error the highest (MAE 11.93; RMSE 14.30) with the KNN Regressor models.

5.1.3 Defensive End

The XGBoost model with hypertuning manual random search using `xgb.train()` and `DMatrix` gives us the best performance for the Defensive End position (R^2 0.78; RMSE 7.61; MAE 3.55). Table 5.3 shows that each of our performance values for this position is strongest with this model. Much like with the best performance in R^2 and RMSE metrics for the Defensive Back position, we find that the XGBoost is best optimized with the hypertuning manual random search using `xgb.train()` and `DMatrix` over using simple XGBoost with no tuning and the XGBoost with hypertuning `RandomizedSearchCV`. R^2 values are again lowest and error highest for the Defensive end position when using the KNN Regressor models (R^2 0.14; RMSE 15.02; MAE 9.52).

5.1.4 Defensive Tackle

XGBoost with hypertuning manual random search using `xgb.train()` and `DMatrix` gives the best performance for the Defensive Tackle position. All performance metrics are highest with this model iteration (R^2 0.59; RMSE 12.80; MAE 5.10). Error is again the highest in the KNN Regressor models (RMSE 18.57 and MAE 10.07). Table 5.4 reflects these findings.

Table 5.2: Five Machine Learning Model Results for the Defensive Back Position. Each table presents a different learning model, with columns representing configuration variations. An "X" indicates inclusion of a specific preprocessing step or method. Yellow shading highlights the configuration that achieved the best performance.

KNN Regressor Models		1	2	3	4	5	6	7	8
Defensive Back - Average	Missing Values Median	X	X	X	X				
	Missing Values Fill 0					X	X	X	X
	Scaling/Standardize Features		X	X	X		X	X	X
	No Scaling	X				X			
	GridSearchCV			X				X	
	RandomizedSearchCV				X				X
	MAE	8.00	7.27	10.13	11.93	8.39	7.37	9.92	9.29
	RMSE	11.19	10.60	13.82	14.30	11.35	10.76	13.52	12.98
	R^2	0.56	0.61	0.33	0.29	0.55	0.60	0.36	0.41

Random Forest Models		1	2	3	4	5	6	7	8
Defensive Back - Average	Missing Values Median	X	X			X		X	
	Missing Values Fill 0			X	X		X		X
	Scaling/Standardize Features		X		X			X	X
	No Scaling	X		X		X	X		
	GridSearchCV					X	X	X	X
	MAE	3.41	3.37	3.38	3.34	3.36	3.37	3.30	3.34
	RMSE	5.89	5.80	5.82	5.75	5.76	5.81	5.70	5.76
	R^2	0.88	0.88	0.88	0.88	0.88	0.88	0.89	0.88

Gradient Boosting Models		1	2	3	4	5	6	7	8
Defensive Back - Average	Missing Values Median	X	X			X		X	
	Missing Values Fill 0			X	X		X		X
	Scaling/Standardize Features		X		X			X	X
	No Scaling	X		X		X	X		
	GridSearchCV					X	X	X	X
	MAE	3.97	3.96	3.79	3.79	3.93	4.31	4.35	4.25
	RMSE	6.72	6.67	6.35	6.35	6.33	6.78	6.83	6.83
	R^2	0.84	0.84	0.86	0.86	0.86	0.84	0.84	0.84

XGBoost Models		1	2	3
Defensive Back - Average	Simple XGBoost Model (no-tuning)	X		
	XGBoost with Hypertuning RandomizedSearchCV		X	
	XGBoost with Hypertuning Manual Random Search using xgb.train() and Dmatrix.			X
	MAE	3.51	3.31	3.59
	RMSE	6.10	5.77	5.57
	R^2	0.87	0.88	0.89

Stacking Models		1	2
Defensive Back - Average	Simple Averaging Ensemble	X	
	Stacked Ensemble		X
	MAE	4.49	3.13
	RMSE	6.77	4.98
	R^2	0.84	0.91

Table 5.3: Five Machine Learning Model Results for the Defensive End Position. Each table presents a different learning model, with columns representing configuration variations. An "X" indicates inclusion of a specific preprocessing step or method. Yellow shading highlights the configuration that achieved the best performance.

KNN Regressor Models		1	2	3	4	5	6	7	8
Defensive End - Average	Missing Values Median	X	X	X	X				
	Missing Values Fill 0					X	X	X	X
	Scaling/Standardize Features		X	X	X		X	X	X
	No Scaling	X				X			
	GridSearchCV			X				X	
	RandomizedSearchCV				X				X
	MAE	9.52	9.06	8.64	8.73	8.97	8.96	8.67	8.86
	RMSE	15.02	14.00	13.45	13.42	14.60	13.93	13.42	14.00
	R^2	0.14	0.25	0.31	0.31	0.18	0.26	0.31	0.25

Random Forest Models		1	2	3	4	5	6	7	8
Defensive End - Average	Missing Values Median	X	X			X		X	
	Missing Values Fill 0			X	X		X		X
	Scaling/Standardize Features		X		X			X	X
	No Scaling	X		X		X	X		
	GridSearchCV					X	X	X	X
	MAE	3.85	3.85	3.84	3.83	3.85	3.84	3.92	3.83
	RMSE	7.84	7.84	7.90	7.90	7.84	7.90	7.96	7.90
	R^2	0.76	0.76	0.76	0.76	0.76	0.76	0.76	0.76

Gradient Boosting Models		1	2	3	4	5	6	7	8
Defensive End - Average	Missing Values Median	X	X			X		X	
	Missing Values Fill 0			X	X		X		X
	Scaling/Standardize Features		X		X			X	X
	No Scaling	X		X		X	X		
	GridSearchCV					X	X	X	X
	MAE	4.42	4.42	4.36	4.36	4.17	4.07	4.16	4.16
	RMSE	10.41	10.41	10.27	10.27	9.14	9.04	9.12	9.14
	R^2	0.58	0.58	0.60	0.60	0.68	0.69	0.68	0.68

XGBoost Models		1	2	3
Defensive End - Average	Simple XGBoost Model (no-tuning)	X		
	XGBoost with Hypertuning RandomizedSearchCV		X	
	XGBoost with Hypertuning Manual Random Search using xgb.train() and Dmatrix.			X
	MAE	4.59	4.45	3.55
	RMSE	10.78	9.37	7.61
	R^2	0.55	0.66	0.78

Stacking Models		1	2
Defensive End - Average	Simple Averaging Ensemble	X	
	Stacked Ensemble		X
	MAE	4.72	4.35
	RMSE	9.17	6.37
	R^2	0.68	0.84

Table 5.4: Five Machine Learning Model Results for the Defensive Tackle Position. Each table presents a different learning model, with columns representing configuration variations. An "X" indicates inclusion of a specific preprocessing step or method. Yellow shading highlights the configuration that achieved the best performance.

KNN Regressor Models		1	2	3	4	5	6	7	8
Defensive Tackle - Average	Missing Values Median	X	X	X	X				
	Missing Values Fill 0					X	X	X	X
	Scaling/Standardize Features		X	X	X		X	X	X
	No Scaling	X				X			
	GridSearchCV			X				X	
	RandomizedSearchCV				X				X
	MAE	9.52	6.65	8.79	8.14	8.47	6.65	7.85	10.07
	RMSE	17.69	13.87	16.80	15.55	16.79	13.87	15.75	18.57
	R^2	0.22	0.52	0.30	0.40	0.30	0.52	0.38	0.14

Random Forest Models		1	2	3	4	5	6	7	8
Defensive Tackle - Average	Missing Values Median	X	X			X		X	
	Missing Values Fill 0			X	X		X		X
	Scaling/Standardize Features		X		X			X	X
	No Scaling	X		X		X	X		
	GridSearchCV					X	X	X	X
	MAE	5.88	5.82	5.75	5.75	5.81	5.65	5.74	5.62
	RMSE	14.63	14.62	14.56	14.56	14.72	14.52	14.61	14.52
	R^2	0.47	0.47	0.47	0.47	0.46	0.48	0.47	0.48

Gradient Boosting Models		1	2	3	4	5	6	7	8
Defensive Tackle - Average	Missing Values Median	X	X			X		X	
	Missing Values Fill 0			X	X		X		X
	Scaling/Standardize Features		X		X			X	X
	No Scaling	X		X		X	X		
	GridSearchCV					X	X	X	X
	MAE	5.74	5.74	5.38	5.38	6.02	5.83	6.02	5.80
	RMSE	13.62	13.62	13.66	13.66	15.16	14.53	15.16	14.53
	R^2	0.54	0.54	0.54	0.54	0.43	0.48	0.43	0.48

XGBoost Models		1	2	3
Defensive Tackle - Average	Simple XGBoost Model (no-tuning)	X		
	XGBoost with Hypertuning RandomizedSearchCV		X	
	XGBoost with Hypertuning Manual Random Search using xgb.train() and Dmatrix.			X
	MAE	5.83	5.50	5.10
	RMSE	14.37	14.43	12.80
	R^2	0.49	0.48	0.59

Stacking Models		1	2
Defensive Tackle - Average	Simple Averaging Ensemble	X	
	Stacked Ensemble		X
	MAE	5.83	6.91
	RMSE	13.90	9.30
	R^2	0.52	0.79

5.1.5 Linebacker

As with the Defensive End and Defensive Back positions, Table 5.5 the XGBoost with hypertuning manual random search using `xgb.train()` and `DMatrix` gives the best model optimization for Linebackers (R^2 0.82; RMSE 27.71; MAE 6.90). Overall, error for this position is higher, however.

5.1.6 Offensive Line

As shown in Table 5.6, XGBoost with hypertuning `RandomizedSearchCV` tied with Random Forest models replacing missing values with zero and using scaling to give the best results for R^2 among the Offensive Linemen position (0.97). Error, however was minimized the most in the XGBoost model with `RandomizedSearchCV` (MAE 3.44; RMSE 5.42). Error was highest and R^2 was the lowest in the KNN Regressor models.

5.1.7 Outside Linebacker

Gradient Boosting models replacing missing values with median values for average price sold give the best model performance for the Outside Linebacker position, as shown in Table 5.7. We achieve the same performance metrics for the first and second iteration (R^2 0.68; RMSE 9.05; MAE 5.47). Interestingly, the first iteration does not apply scaling, and the second iteration does apply scaling, but we get the same results. This indicates that scaling may not impact optimization for this position as strongly as for others. KNN Regressor models again deliver the lowest R^2 values (-0.050) and highest error (RMSE 16.33; MAE 10.04).

5.1.8 Quarterback

For the Quarterback position, as shown in Table 5.8, we get the best R^2 (0.80) and lowest RMSE (112.92) with the Random Forest model that replaces missing average price sold value with the median and applying scaling. MAE is lowest with the XGBoost model with hypertuning manual random search using `xgb.train()` and `DMatrix` (MAE 39.84). Overall,

Table 5.5: Five Machine Learning Model Results for the Linebacker Position. Each table presents a different learning model, with columns representing configuration variations. An "X" indicates inclusion of a specific preprocessing step or method. Yellow shading highlights the configuration that achieved the best performance.

KNN Regressor Models		1	2	3	4	5	6	7	8
Linebacker - Average	Missing Values Median	X	X	X	X				
	Missing Values Fill 0					X	X	X	X
	Scaling/Standardize Features		X	X	X		X	X	X
	No Scaling	X				X			
	GridSearchCV			X				X	
	RandomizedSearchCV				X				X
	MAE	17.8	17.6	17.7	16.5	18.9	17.5	17.6	16.4
	RMSE	52.7	52.0	55.3	46.0	53.0	52.0	55.3	46.0
	R^2	0.3	0.4	0.3	0.5	0.3	0.4	0.3	0.5

Random Forest Models		1	2	3	4	5	6	7	8
Linebacker - Average	Missing Values Median	X	X			X		X	
	Missing Values Fill 0			X	X		X		X
	Scaling/Standardize Features		X		X			X	X
	No Scaling	X		X		X	X		
	GridSearchCV					X	X	X	X
	MAE	8.99	9.00	9.20	9.21	9.14	9.21	9.09	9.15
	RMSE	39.76	39.76	40.27	40.27	40.04	40.25	39.48	39.65
	R^2	0.62	0.62	0.61	0.61	0.62	0.61	0.63	0.62

Gradient Boosting Models		1	2	3	4	5	6	7	8
Linebacker - Average	Missing Values Median	X	X			X		X	
	Missing Values Fill 0			X	X		X		X
	Scaling/Standardize Features		X		X			X	X
	No Scaling	X		X		X	X		
	GridSearchCV					X	X	X	X
	MAE	8.32	8.32	8.31	8.31	9.43	8.17	9.49	8.17
	RMSE	31.35	31.35	32.35	32.35	31.88	28.94	31.97	28.94
	R^2	0.77	0.77	0.75	0.75	0.76	0.80	0.76	0.80

XGBoost Models		1	2	3
Linebacker - Average	Simple XGBoost Model (no-tuning)	X		
	XGBoost with Hypertuning RandomizedSearchCV		X	
	XGBoost with Hypertuning Manual Random Search using xgb.train() and Dmatrix.			X
	MAE	8.85	8.45	6.90
	RMSE	31.62	33.10	27.71
	R^2	0.76	0.74	0.82

Stacking Models		1	2
Linebacker - Average	Simple Averaging Ensemble	X	
	Stacked Ensemble		X
	MAE	9.61	7.03
	RMSE	35.28	9.67
	R^2	0.70	0.98

Table 5.6: Five Machine Learning Model Results for the Offensive Line Position. Each table presents a different learning model, with columns representing configuration variations. An "X" indicates inclusion of a specific preprocessing step or method. Yellow shading highlights the configuration that achieved the best performance.

KNN Regressor Models		1	2	3	4	5	6	7	8
Offensive Line - Average	Missing Values Median	X	X	X	X				
	Missing Values Fill 0					X	X	X	X
	Scaling/Standardize Features		X	X	X		X	X	X
	No Scaling	X				X			
	GridSearchCV			X				X	
	RandomizedSearchCV				X				X
	MAE	16.94	16.43	17.85	18.45	15.86	16.43	18.09	17.95
	RMSE	25.07	21.93	28.20	29.22	24.22	21.93	28.21	28.39
	R^2	0.40	0.54	0.24	0.19	0.44	0.54	0.24	0.23

Random Forest Models		1	2	3	4	5	6	7	8
Offensive Line - Average	Missing Values Median	X	X			X		X	
	Missing Values Fill 0			X	X		X		X
	Scaling/Standardize Features		X		X			X	X
	No Scaling	X		X		X	X		
	GridSearchCV					X	X	X	X
	MAE	4.41	4.25	4.09	4.08	4.17	4.10	4.29	4.27
	RMSE	6.19	6.01	5.70	5.69	6.03	5.92	6.24	6.13
	R^2	0.96	0.97	0.97	0.97	0.97	0.97	0.96	0.96

Gradient Boosting Models		1	2	3	4	5	6	7	8
Offensive Line - Average	Missing Values Median	X	X			X		X	
	Missing Values Fill 0			X	X		X		X
	Scaling/Standardize Features		X		X			X	X
	No Scaling	X		X		X	X		
	GridSearchCV					X	X	X	X
	MAE	5.75	5.67	4.82	4.75	6.17	7.71	6.08	7.32
	RMSE	10.47	10.37	9.57	9.46	8.96	12.24	8.86	12.00
	R^2	0.90	0.90	0.91	0.91	0.92	0.86	0.93	0.86

XGBoost Models		1	2	3
Offensive Line - Average	Simple XGBoost Model (no-tuning)	X		
	XGBoost with Hypertuning RandomizedSearchCV		X	
	XGBoost with Hypertuning Manual Random Search using xgb.train() and Dmatrix.			X
	MAE	4.52	3.44	3.77
	RMSE	7.88	5.42	5.52
	R^2	0.94	0.97	0.97

Stacking Models		1	2
Offensive Line - Average	Simple Averaging Ensemble	X	
	Stacked Ensemble		X
	MAE	6.49	3.58
	RMSE	9.77	5.19
	R^2	0.91	0.97

Table 5.7: Five Machine Learning Model Results for the Outside Linebacker Position. Each table presents a different learning model, with columns representing configuration variations. An "X" indicates inclusion of a specific preprocessing step or method. Yellow shading highlights the configuration that achieved the best performance.

KNN Regressor Models		1	2	3	4	5	6	7	8
Outside Linebacker - Average	Missing Values Median	X	X	X	X				
	Missing Values Fill 0					X	X	X	X
	Scaling/Standardize Features		X	X	X		X	X	X
	No Scaling	X				X			
	GridSearchCV			X				X	
	RandomizedSearchCV				X				X
	MAE	10.04	8.97	8.97	8.60	9.80	8.95	8.95	8.83
	RMSE	16.33	15.01	15.01	14.52	16.12	15.01	15.01	14.99
	R^2	-0.05	0.11	0.11	0.17	-0.02	0.11	0.11	0.12

Random Forest Models		1	2	3	4	5	6	7	8
Outside Linebacker - Average	Missing Values Median	X	X			X		X	
	Missing Values Fill 0			X	X		X		X
	Scaling/Standardize Features		X		X			X	X
	No Scaling	X		X		X	X		
	GridSearchCV					X	X	X	X
	MAE	6.18	6.18	6.16	6.16	6.72	6.71	6.72	6.71
	RMSE	11.51	11.51	11.60	11.60	12.91	12.98	12.91	12.98
	R^2	0.48	0.48	0.47	0.47	0.34	0.34	0.34	0.34

Gradient Boosting Models		1	2	3	4	5	6	7	8
Outside Linebacker - Average	Missing Values Median	X	X			X		X	
	Missing Values Fill 0			X	X		X		X
	Scaling/Standardize Features		X		X			X	X
	No Scaling	X		X		X	X		
	GridSearchCV					X	X	X	X
	MAE	5.47	5.47	5.52	5.52	9.26	9.27	9.26	9.27
	RMSE	9.05	9.05	9.12	9.12	16.05	16.05	16.05	16.05
	R^2	0.68	0.68	0.67	0.67	-0.01	-0.02	-0.01	-0.02

XGBoost Models		1	2	3
Outside Linebacker - Average	Simple XGBoost Model (no-tuning)	X		
	XGBoost with Hypertuning RandomizedSearchCV		X	
	XGBoost with Hypertuning Manual Random Search using xgb.train() and Dmatrix.			X
	MAE	5.75	6.83	5.29
	RMSE	10.32	13.10	9.49
	R^2	0.58	0.32	0.65

Stacking Models		1	2
Outside Linebacker - Average	Simple Averaging Ensemble	X	
	Stacked Ensemble		X
	MAE	5.76	3.29
	RMSE	7.96	4.52
	R^2	0.94	0.98

error is high for this position across all model iterations. We again see the worst optimization with the KNN Regressor models (R^2 0.16; RMSE 232.32; MAE 95.37).

5.1.9 Running Back

As found with the Defensive tackle, Defensive End, and Defensive back, our models are best optimized for the Running Back position with the XGBoost model with hypertuning manual random search using `xgb.train()` and `DMatrix`. We can see these results in Table 5.9. The performance metrics are all again optimized in this model iteration (R^2 0.83; RMSE 18.04; MAE 5.76). We again see the KNN Regressor models perform the poorest with the lowest R^2 value (0.26) and highest error (RMSE 37.31 and MAE 17.69). This phenomenon is again shown in evaluating model performance for the Wide Receiver position.

5.1.10 Wide Receiver

Table 5.10 shows that KNN Regressor models deliver the lowest R^2 value (0.16) and highest error (RMSE 37.46 and MAE 23.82) for Wide Receivers. The highest R^2 value (0.81) and lowest RMSE value (17.76) for this position is found with the XGBoost model using hypertuning manual random search with `xgb.train()` and `DMatrix`. The lowest MAE for this position (9.33) is found with the Gradient Boosting model filling missing average price sold values with zero, not using scaling, and applying `GridSearchCV`.

5.1.11 Tight End

As found with the Outside Linebacker position, the Gradient Boosting models using the median average price sold for missing values yields the best optimization for the Tight End Position. Again, values are identical in the first and second iterations (R^2 0.93; RMSE 10.11; MAE 6.31), regardless of scaling. We find that the KNN Regressor models again produce the lowest R^2 values (0.54) and highest error values (RMSE 26.77; MAE 14.16). These results can be found in Table 5.11.

Table 5.8: Five Machine Learning Model Results for the Quarterback Position. Each table presents a different learning model, with columns representing configuration variations. An "X" indicates inclusion of a specific preprocessing step or method. Yellow shading highlights the configuration that achieved the best performance.

KNN Regressor Models		1	2	3	4	5	6	7	8
Quarterback - Average	Missing Values Median	X	X	X	X				
	Missing Values Fill 0					X	X	X	X
	Scaling/Standardize Features		X	X	X		X	X	X
	No Scaling	X				X			
	GridSearchCV			X				X	
	RandomizedSearchCV				X				X
	MAE	95.37	74.64	65.98	73.17	103.50	74.64	65.98	73.17
	RMSE	225.33	163.76	140.45	158.25	232.32	163.76	140.45	158.25
	R^2	0.21	0.58	0.69	0.61	0.16	0.58	0.69	0.61

Random Forest Models		1	2	3	4	5	6	7	8
Quarterback - Average	Missing Values Median	X	X			X		X	
	Missing Values Fill 0			X	X		X		X
	Scaling/Standardize Features		X		X			X	X
	No Scaling	X		X		X	X		
	GridSearchCV					X	X	X	X
	MAE	43.29	43.12	44.11	44.10	51.09	51.38	50.74	51.17
	RMSE	113.54	112.92	114.33	114.48	140.50	141.62	140.23	141.29
	R^2	0.80	0.80	0.80	0.80	0.69	0.69	0.69	0.69

Gradient Boosting Models		1	2	3	4	5	6	7	8
Quarterback - Average	Missing Values Median	X	X			X		X	
	Missing Values Fill 0			X	X		X		X
	Scaling/Standardize Features		X		X			X	X
	No Scaling	X		X		X	X		
	GridSearchCV					X	X	X	X
	MAE	58.85	56.93	58.82	56.46	58.78	55.13	57.16	53.89
	RMSE	180.15	172.97	179.50	171.32	172.68	177.56	167.33	171.19
	R^2	0.49	0.53	0.50	0.54	0.54	0.51	0.56	0.54

XGBoost Models		1	2	3
Quarterback - Average	Simple XGBoost Model (no-tuning)	X		
	XGBoost with Hypertuning RandomizedSearchCV		X	
	XGBoost with Hypertuning Manual Random Search using xgb.train() and Dmatrix.			X
	MAE	59.44	50.99	39.84
	RMSE	175.72	150.19	122.96
	R^2	0.52	0.65	0.76

Stacking Models		1	2
Quarterback - Average	Simple Averaging Ensemble	X	
	Stacked Ensemble		X
	MAE	45.85	49.06
	RMSE	123.10	93.19
	R^2	0.76	0.86

Table 5.9: Five Machine Learning Model Results for the Running Back Position. Each table presents a different learning model, with columns representing configuration variations. An "X" indicates inclusion of a specific preprocessing step or method. Yellow shading highlights the configuration that achieved the best performance.

KNN Regressor Models		1	2	3	4	5	6	7	8
Running Back - Average	Missing Values Median	X	X	X	X				
	Missing Values Fill 0					X	X	X	X
	Scaling/Standardize Features		X	X	X		X	X	X
	No Scaling	X				X			
	GridSearchCV			X				X	
	RandomizedSearchCV				X				X
	MAE	17.65	14.17	14.02	14.25	17.69	14.17	14.01	14.25
	RMSE	37.30	28.89	29.62	29.44	37.31	28.89	29.63	29.44
	R^2	0.26	0.56	0.54	0.54	0.26	0.56	0.54	0.54

Random Forest Models		1	2	3	4	5	6	7	8
Running Back - Average	Missing Values Median	X	X			X		X	
	Missing Values Fill 0			X	X		X		X
	Scaling/Standardize Features		X		X			X	X
	No Scaling	X		X		X	X		
	GridSearchCV					X	X	X	X
	MAE	7.13	7.08	7.21	7.21	7.32	7.17	7.36	7.21
	RMSE	22.87	22.86	23.91	23.90	24.01	23.69	24.55	23.55
	R^2	0.72	0.72	0.70	0.70	0.69	0.70	0.68	0.71

Gradient Boosting Models		1	2	3	4	5	6	7	8
Running Back - Average	Missing Values Median	X	X			X		X	
	Missing Values Fill 0			X	X		X		X
	Scaling/Standardize Features		X		X			X	X
	No Scaling	X		X		X	X		
	GridSearchCV					X	X	X	X
	MAE	7.12	7.11	6.89	6.89	5.91	5.99	6.79	6.85
	RMSE	21.11	21.11	20.32	20.32	19.09	20.32	20.18	20.23
	R^2	0.76	0.76	0.78	0.78	0.81	0.78	0.78	0.78

XGBoost Models		1	2	3
Running Back - Average	Simple XGBoost Model (no-tuning)	X		
	XGBoost with Hypertuning RandomizedSearchCV		X	
	XGBoost with Hypertuning Manual Random Search using xgb.train() and Dmatrix.			X
	MAE	6.75	6.75	5.76
	RMSE	20.74	18.29	18.04
	R^2	0.77	0.82	0.83

Stacking Models		1	2
Running Back - Average	Simple Averaging Ensemble	X	
	Stacked Ensemble		X
	MAE	7.42	8.51
	RMSE	22.55	13.90
	R^2	0.73	0.90

Table 5.10: Five Machine Learning Model Results for the Wide Receiver Position. Each table presents a different learning model, with columns representing configuration variations. An "X" indicates inclusion of a specific preprocessing step or method. Yellow shading highlights the configuration that achieved the best performance.

KNN Regressor Models		1	2	3	4	5	6	7	8
Wide Receiver - Average	Missing Values Median	X	X	X	X				
	Missing Values Fill 0					X	X	X	X
	Scaling/Standardize Features		X	X	X		X	X	X
	No Scaling	X				X			
	GridSearchCV			X				X	
	RandomizedSearchCV				X				X
	MAE	23.61	20.41	19.40	19.25	23.82	20.41	19.41	19.26
	RMSE	37.46	35.40	32.65	32.36	37.61	35.40	32.65	32.36
	R^2	0.16	0.25	0.36	0.37	0.16	0.25	0.36	0.37

Random Forest Models		1	2	3	4	5	6	7	8
Wide Receiver - Average	Missing Values Median	X	X			X		X	
	Missing Values Fill 0			X	X		X		X
	Scaling/Standardize Features		X		X			X	X
	No Scaling	X		X		X	X		
	GridSearchCV					X	X	X	X
	MAE	10.02	10.03	9.70	9.68	10.02	10.20	10.03	10.16
	RMSE	20.29	20.36	19.39	19.39	20.29	21.95	20.36	21.94
	R^2	0.75	0.75	0.78	0.78	0.75	0.71	0.75	0.71

Gradient Boosting Models		1	2	3	4	5	6	7	8
Wide Receiver - Average	Missing Values Median	X	X			X		X	
	Missing Values Fill 0			X	X		X		X
	Scaling/Standardize Features		X		X			X	X
	No Scaling	X		X		X	X		
	GridSearchCV					X	X	X	X
	MAE	11.38	11.38	11.01	11.01	9.89	9.33	10.01	12.06
	RMSE	31.35	31.35	30.46	30.46	25.73	24.45	25.35	34.65
	R^2	0.41	0.41	0.45	0.45	0.60	0.64	0.62	0.28

XGBoost Models		1	2	3
Wide Receiver - Average	Simple XGBoost Model (no-tuning)	X		
	XGBoost with Hypertuning RandomizedSearchCV		X	
	XGBoost with Hypertuning Manual Random Search using xgb.train() and Dmatrix.			X
	MAE	11.13	9.81	10.56
	RMSE	23.71	20.59	17.76
	R^2	0.66	0.75	0.81

Stacking Models		1	2
Wide Receiver - Average	Simple Averaging Ensemble	X	
	Stacked Ensemble		X
	MAE	10.01	8.07
	RMSE	20.60	13.21
	R^2	0.75	0.90

Table 5.11: Five Machine Learning Model Results for the Tight End Position. Each table presents a different learning model, with columns representing configuration variations. An "X" indicates inclusion of a specific preprocessing step or method. Yellow shading highlights the configuration that achieved the best performance.

KNN Regressor Models		1	2	3	4	5	6	7	8
Tight End - Average	Missing Values Median	X	X	X	X				
	Missing Values Fill 0					X	X	X	X
	Scaling/Standardize Features		X	X	X		X	X	X
	No Scaling	X				X			
	GridSearchCV			X				X	
	RandomizedSearchCV				X				X
	MAE	13.62	11.80	9.98	11.60	14.16	11.91	9.85	11.64
	RMSE	26.54	19.29	20.94	20.21	26.77	19.35	20.91	20.22
	R^2	0.54	0.75	0.71	0.73	0.53	0.75	0.71	0.73

Random Forest Models		1	2	3	4	5	6	7	8
Tight End - Average	Missing Values Median	X	X			X		X	
	Missing Values Fill 0			X	X		X		X
	Scaling/Standardize Features		X		X			X	X
	No Scaling	X		X		X	X		
	GridSearchCV					X	X	X	X
	MAE	7.17	7.18	7.13	7.16	6.90	7.03	6.91	7.07
	RMSE	15.22	15.23	15.68	15.70	15.30	15.30	15.36	15.32
	R^2	0.85	0.85	0.84	0.84	0.85	0.85	0.84	0.85

Gradient Boosting Models		1	2	3	4	5	6	7	8
Tight End - Average	Missing Values Median	X	X			X		X	
	Missing Values Fill 0			X	X		X		X
	Scaling/Standardize Features		X		X			X	X
	No Scaling	X		X		X	X		
	GridSearchCV					X	X	X	X
	MAE	6.31	6.31	6.31	6.30	8.38	7.37	8.39	7.43
	RMSE	10.11	10.11	10.38	10.38	20.22	15.28	20.22	15.29
	R^2	0.93	0.93	0.93	0.93	0.73	0.85	0.73	0.85

XGBoost Models		1	2	3
Tight End - Average	Simple XGBoost Model (no-tuning)	X		
	XGBoost with Hypertuning RandomizedSearchCV		X	
	XGBoost with Hypertuning Manual Random Search using xgb.train() and Dmatrix.			X
	MAE	6.50	6.42	6.59
	RMSE	12.54	14.16	12.30
	R^2	0.90	0.87	0.90

Stacking Models		1	2
Tight End - Average	Simple Averaging Ensemble	X	
	Stacked Ensemble		X
	MAE	5.94	6.01
	RMSE	12.48	8.56
	R^2	0.90	0.95

5.1.12 Safety

Of interest, we only find model optimization using KNN Regressor models for the Safety position, as shown in Table 5.12. The highest R^2 value (0.66) and lowest error (RMSE 8.28 and MAE 6.33) are all produced by the KNN Regressor model that fills missing average price values with the median and that apply scaling. Gradient Boosting models yield the lowest R^2 values (0.08) and highest RMSE (13.52) for the Safety position. We find the highest MAE with the KNN Regressor models (9.57) when filling missing average selling price with zero and when scaling is not applied.

5.1.13 Comparison of Models Across Positions

R^2 was optimized for the most positions with the XGBoost models, including Defensive Back (0.89, which is the same in the Random Forest model), Defensive End (0.78), Defensive Tackle (0.59), Linebacker (0.82), Offensive Line (0.97, which is the same in Random Forest modeling for this position), Running Back (0.83), and Wide Receivers (0.81). Random Forest models produced the best R^2 values for Cornerbacks (0.70), Defensive Backs (0.89, matching the XGBoost model), Offensive Linemen (0.97, matching the XGBoost), and Quarterbacks (0.80). Gradient Boost models optimized R^2 for Outside Linebackers (0.68) and Tight Ends (0.93). R^2 is only optimized for Safeties (0.66) in the KNN model.

XGBoost also gave the lowest error for most positions, including Defensive Backs (5.57 RMSE), Defensive Ends (3.55 MAE; 7.61 RMSE); Defensive Tackles (5.10 MAE; 12.80 RMSE), Linebackers (6.90 MAE; 27.71 RMSE), Offensive Linemen (3.44 MAE; 519 RMSE), Outside Linebackers (5.29 MAE), Quarterbacks (39.84 MAE); Running Backs (5.76 MAE; 18.04 RMSE), and Wide Receivers (17.76 RMSE). Gradient Boost models gave error minimization for Cornerbacks (6.26 MAE), Outside Linebackers (9.05 RMSE), Wide Receivers (9.33 MAE), and Tight Ends (6.31 MAE; 10.11 RMSE). Random Forest models optimized error for Cornerbacks (11.29 RMSE), Defensive Backs (3.30 MAE), and Quarterbacks (112.92 RMSE). KNN modeling only optimized error for Safeties, as with the R^2 value for this model

Table 5.12: Five Machine Learning Model Results for the Safety Position. Each table presents a different learning model, with columns representing configuration variations. An "X" indicates inclusion of a specific preprocessing step or method. Yellow shading highlights the configuration that achieved the best performance.

KNN Regressor Models		1	2	3	4	5	6	7	8
Safety - Average	Missing Values Median	X	X	X	X				
	Missing Values Fill 0					X	X	X	X
	Scaling/Standardize Features		X	X	X		X	X	X
	No Scaling	X				X			
	GridSearchCV			X				X	
	RandomizedSearchCV				X				X
	MAE	9.37	6.33	8.99	9.24	9.57	6.33	8.99	8.85
	RMSE	12.41	8.28	12.71	12.76	12.40	8.29	12.82	12.75
	R^2	0.23	0.66	0.19	0.18	0.23	0.66	0.18	0.18

Random Forest Models		1	2	3	4	5	6	7	8
Safety - Average	Missing Values Median	X	X			X		X	
	Missing Values Fill 0			X	X		X		X
	Scaling/Standardize Features		X		X			X	X
	No Scaling	X		X		X	X		
	GridSearchCV					X	X	X	X
	MAE	6.43	6.42	6.69	6.66	6.80	6.93	6.79	6.90
	RMSE	10.63	10.65	10.75	10.77	11.04	11.08	11.05	11.07
	R^2	0.43	0.43	0.42	0.42	0.39	0.38	0.39	0.39

Gradient Boosting Models		1	2	3	4	5	6	7	8
Safety - Average	Missing Values Median	X	X			X		X	
	Missing Values Fill 0			X	X		X		X
	Scaling/Standardize Features		X		X			X	X
	No Scaling	X		X		X	X		
	GridSearchCV					X	X	X	X
	MAE	8.18	8.18	8.43	8.41	7.46	7.42	7.51	7.52
	RMSE	12.92	12.92	13.51	13.51	11.82	12.35	12.40	12.16
	R^2	0.16	0.16	0.08	0.08	0.30	0.24	0.23	0.26

XGBoost Models		1	2	3
Safety - Average	Simple XGBoost Model (no-tuning)	X		
	XGBoost with Hypertuning RandomizedSearchCV		X	
	XGBoost with Hypertuning Manual Random Search using xgb.train() and Dmatrix.			X
	MAE	7.88	6.82	6.90
	RMSE	12.65	10.91	10.39
	R^2	0.20	0.40	0.46

Stacking Models		1	2
Safety - Average	Simple Averaging Ensemble	X	
	Stacked Ensemble		X
	MAE	6.20	5.71
	RMSE	9.90	8.67
	R^2	0.51	0.62

(6.33 MAE; 8.28 RMSE). Error was highest and R^2 lowest for all positions, except for Safety, with the KNN modeling.

Overall, XGBoost optimized error and R^2 performance for most positions before stacking. KNN models gave the worst overall performance. In general, error was the highest by far for the Quarterback position. Future work should explore this relationship to determine if this can be reduced, but it likely reflects the highest card values overall as this position in often the most collected. Higher average selling prices will inherently give larger ranges and variance, but a deeper look into this will only benefit future predictive modeling.

5.2 Results of Stacking Models

Seeking overall improvement with model performance, stacking is applied as an ensemble learning technique that combines the predictions of multiple base models to generate an optimized final prediction. Predictions from the same training dataset in base models are passed to a second-level model that makes the final predictions in the optimized final model. By learning how to weigh and combine the strengths of base model predictions, the final second-level model should demonstrate higher performance metrics than observed with the base models.

Except for the Safety position, stacking models optimized models for all positions. The optimized model for the Safety position was the only position where KNN Regression yielded the best performance for a role. With the Safety position, R^2 is optimized by KNN modeling (0.66 versus 0.62). Error is also optimized for this position (RMSE of 8.28 versus 8.67; MAE 6.33 versus 5.71). Otherwise, we see improvement for all models using stacked ensemble models over simple averaging ensemble models. For the Cornerback position, stacking produced an increase in R^2 (0.69 to 0.77) and error reduction in RMSE (11.45 to 9.94) and MAE (6.26 to 6.29). For the Defensive Back position, R^2 increased from 0.89 to 0.91. We also find the lowest RMSE value (4.98) and MAE value 3.13). R^2 values for the Defensive End position increases to 0.84 while MAE values remain lowest in the XGBoost models. RMSE is optimized with model stacking (6.37). We see higher R^2 performance (0.79 versus 0.59) and improved RMSE scores (9.30 from 12.80) with the Defensive Tackle position, but MAE is optimized with XGBoost models (5.10 versus 6.91 with stacking). Ensemble modeling for Linebackers produces a large jump in R^2 (0.82 increases to 0.98) and a sizable reduction in RMSE (27.71 reduces to 9.67). MAE for Linebackers remains optimal with the XGBoost models (6.90 versus 7.03).

We see a consistent R^2 value (0.97 before and after stacking) and a modest RMSE decrease (5.42 reduces to 5.19) for Offensive Linemen, but MAE has the best performance with the XGBoost models (3.44 versus 3.58). For Outside Linebackers, all metrics are optimized with stacking models (R^2 increases from 0.68 to 0.98; RMSE decreases from 9.05 to 4.52; MAE reduces from 5.29 to 3.30). R^2 and RMSE performance metrics for Quarterbacks are also enhanced with ensemble models (R^2 value increases from 0.80 to 0.87; RMSE decreases from 112.92 to 93.19), but MAE is optimized with XGBoost modeling (39.84 over 49.06 with stacking). Similarly, for the Running Backs, ensemble modeling boosts R^2 (from 0.83 to 0.90) and RMSE (from 18.04 to 13.90), but MAE is optimized in the XGBoost models (5.76 versus 8.51). Predictive performance for Wide Receivers is enhanced across all metrics with stacking (R^2 increases from 0.81 to 0.90; RMSE reduces from 17.76 to 13.21; MAE reduces from 9.33 to 8.07). We also see improvement with all metrics for the Tight End position (R^2 increases from 0.93 to 0.95; RMSE reduces from 10.11 to 8.56; MAE reduces from 6.31 to 6.01).

Overall, we get the best predictive ability utilizing ensemble learning for the Tight End (0.95), Outside Linebacker (0.97, but this is unchanged with stacking), Offensive Line (0.97), and Linebacker (0.98) positions. It appears that our ability to explain variance in the average price a card will sell for based on draft and season performance features is best for these roles. All of these positions, except for Tight End, have three performance variables. There are seven performance variables for the Tight End position. We get the least predictive explanation for the Safety (0.66), Defensive End (0.79), and Cornerback (0.77). The biggest improvement in predictive performance by introducing ensemble learning is seen with the Defensive End (increases of 0.19), Linebacker (increases of 0.16), and Outside Linebacker (increases of 0.30).

5.3 Evaluating Machine Learning Models: Top Feature Performance in Optimized Models

Features utilized for model optimization varied by position. This is largely due to the nature of varying performance metrics respective to evaluation by role, as discussed earlier. It is important to understand how different features lead to different model optimization and which features are most important for our learning models if we are to be able to accurately predict trading card values. It is important to note that it will not be possible to make such an evaluation for the KNN models as it does not build a global model or learn explicit parameters linking features to predictions. Instead, it calculates distances between instances using all features, treating all features equally unless otherwise modified. This does not allow for an intuitive means of measuring or ranking the importance of individual features.

The other models under consideration permit an assessment of feature importance by evaluating the contribution a feature makes to reducing uncertainty or error when making predictions. Decision tree models compute feature importance by splitting nodes with decision trees, for which the algorithm selects a feature and a corresponding threshold to split the data into subsets. By minimizing the mix of classes in classification or reducing the variance in regression, the algorithm seeks to make these subsets as pure as possible. Gini impurity or entropy are metrics used for classification trees and reduction in variance or mean squared error for regression trees. When a split is made, the algorithm calculates the decrease in impurity (i.e., error) that results from the split. This decrease is a measure of how well the feature, when used at that split separates the data. Reduction in impurity is attributed to the feature used for that split – the more a feature contributes to decreasing such impurity, the more important it is considered.

Each time a feature is used in a tree, its contribution is recorded, and these contributions are summed across all splits in that tree. In ensemble models like Random Forests, Gradient Boosting, or XGBoost, this process is repeated across many trees. The importance of a feature is then typically averaged or summed over all trees, providing a robust measure of overall importance to the model. Features are scaled so that they sum to 100% or 1, making it easier for comparison. Evaluation of feature importance helps our understanding of which features the model relies on most for predictions. This insight helps with model interpretability, diagnostics, and for informing domain-specific decisions. Features with low importance may be removed from a model in the pursuit of model simplification without experiencing a significant loss in performance. Because feature importance scores are aggregated over many trees, they provide a stable measure even if individual trees differ in splits, helping to highlight features that consistently contribute to reducing prediction error. I will look at feature importance through each tree-based model within position.

Offensive positions (Offensive Lineman, Quarterback, Running Back, Tight End, Wide Receiver) share several top 20 features: card grading, autographing, weighted career value, transaction count, and Pro Bowl selection (on the list for each of the five positions). Weighted career approximate value was on the list for all positions but Tight Ends. 1st team all pro, solo tackles, weight, age, games played, years played, years as primary starter, rushing attempts, receiving yards, and receiving receptions were on the list of three offensive positions. Height, draft year, 40-yard dash, vertical jump, draft pick, bench, passes completed, rushing touchdowns, yards gained by passing, and receiving touchdowns appeared in the top 20 list for two of the five offensive positions. Interestingly, there were several performance features that did not make the top 20 list for any of these positions: 3 cone, sacks, and broad jump. Draft round was only on the list for Tight Ends, indicating that draft round may not be important to predicting success for other positions. Future research should explore this relationship to help uncover important trends.

Defensive positions, which include the Cornerback, Defensive Back, Defensive End, Defensive Tackle, Linebacker, Outside Linebacker, and Safety positions, share four top 20 features: card grading, autographing, weighted career value, and transaction count, which were also on all of our offensive position lists. Age, games played, Silver card variation, and draft pick were in six of the seven position lists. Weighted accumulated value to drafting team, Pro Bowl selections, solo tackles, height, weight, sacks, 40-yard, and bench were on five of the seven position lists. 1st team all pro, broad jump, draft year, vertical, and years as a primary starter were on four of the seven position lists. Draft round, years played, Disco variation, and shuttle were on three of the seven position lists. Passes intercepted on defense and 3 cone were on two of the seven positions lists.

Overall, grading, autographs, weighted career value, and transaction count are the most important features across all positions and were on the top 20 feature lists for all positions. Pro Bowl selection, Silver card variations, weighted accumulated value to drafting teams, 1st team all pro selection, solo tackles, weight, age, games played, years as a primary starter, height, 40-yard, draft pick, and bench appeared on at least seven of 12 position top 20 lists. These features should be considered when trying to predict NFL trading card values and can provide important insight into such predictions.

5.3.1 Cornerback

The top 20 features of importance for the Cornerback position are shown in Table 5.13. When we average all models, we see that card grading and being a 1st team all pro selection are of tremendous importance. However, these features vary in their importance for different models. Card grading is of more importance to the Random Forest and Gradient Boosting models than with the XGBoost model. 1st team selection is of greater importance to the XGBoost model than the other models. Several features have varying significance for different models, but we only find a few at or approximating a 0.10 or greater level of importance, including weighted accumulated approximate value for drafting team (Random Forest model), autograph (Gradient Boosting), weighted career approximate value (XGBoost). Transaction count (Random Forest and Gradient Boosting), passes intercepted on defense (XGBoost), draft round (XGBoost), pro bowl selection (Random Forest), broad jump (XGBoost), solo tackles (XGBoost), height (Gradient Boosting), and draft year (XGBoost) were also important features for this position. Playing professionally for Houston is an important feature, suggesting that playing for this team will drive average card prices for this position. Position-specific features in the top 20 for Cornerbacks include 1st team all pro selection, approximate career value, approximate value for drafting team, intercepted passes, draft round, Pro Bowl selection, broad jump, solo tackles, sacks, games played, and 40-yard dash time.

5.3.2 Defensive Back

Card grading shows great importance for the Defensive Back. As shown in Table 5.14, this feature is of greatest importance to all models. Feature importance is spread more evenly than with the Cornerback position. Weighted career value (XGBoost), weight (XGBoost), 1st team selection (XGBoost), and round drafted (XGBoost) all approximate or surpass 0.01. We see moderate importance among autograph, pro bowl selection, Silver card variation, age of player, and height for the XGBoost model. Weighted career value, weight, transaction count, and weighted approximate value to drafting team were moderately important for the Random Forest model. Of moderate importance to the Gradient Boosting model are weighted career value, weight, 1st team selection, transaction count, autograph, draft year, 3 cone, and solo tackles. Passes intercepted on defense, autograph, pro bowl selection, Silver card variation, age, and height were important features in the XGBoost model. Position-specific performance metrics in the top 20 features include weighted career value, 1st team all pro selection, draft round, intercepted passes, Pro Bowl selection, 3 cone, approximate value to the drafting team, solo tackles, vertical jump, number of games played, draft pick, and bench strength.

Table 5.13: CB Top 20 Features by Importance Across Tree-based Models. Features at the top of the list are of greater average, across all tree-based models, than those lower on the list. Darker blue coloration represents higher feature importance while lighter blue coloration corresponds with lower feature importance.

CB Top 20 Features by Importance Across Tree-based Models

Feature	RandomForest	GradientBoosting	XGBoost	Average
card grading	0.4519	0.3403	0.0709	0.2877
1st Team All Pro Selections	0.1388	0.0608	0.2914	0.1637
Team_HOU	0.0000	0.2218	0.0119	0.0779
autograph	0.0265	0.0960	0.0492	0.0572
Weighted Career Approx Value	0.0191	0.0204	0.1101	0.0498
Weighted Accumulated Approx Value for Team That Drafted	0.0855	0.0491	0.0138	0.0495
trans count	0.0729	0.0584	0.0082	0.0465
Passes Intercepted on Defense	0.0405	0.0285	0.0664	0.0451
Round Drafted	0.0186	0.0267	0.0678	0.0377
Pro Bowl Selections	0.0470	0.0207	0.0257	0.0311
Broad Jump	0.0025	0.0062	0.0733	0.0274
Solo Tackles	0.0016	0.0025	0.0723	0.0255
Ht (in)	0.0140	0.0245	0.0347	0.0244
Draft Year	0.0190	0.0001	0.0389	0.0193
Wt (lbs)	0.0202	0.0081	0.0010	0.0098
Sacks	0.0033	0.0000	0.0185	0.0073
Age	0.0027	0.0038	0.0150	0.0072
Games Played	0.0043	0.0056	0.0044	0.0048
40yd	0.0109	0.0014	0.0015	0.0046
card version_Silver	0.0025	0.0039	0.0038	0.0034

Table 5.14: DB Top 20 Features by Importance Across Tree-based Models. Features at the top of the list are of greater average, across all tree-based models, than those lower on the list. Darker blue coloration represents higher feature importance while lighter blue coloration corresponds with lower feature importance.

DB Top 20 Features by Importance Across Tree-based Models

Feature	RandomForest	GradientBoosting	XGBoost	Average
card grading	0.6380	0.5682	0.1441	0.4501
Weighted Career Approx Value	0.0222	0.0332	0.1462	0.0672
Wt (lbs)	0.0371	0.0218	0.0922	0.0504
1st Team All Pro Selections	0.0044	0.0257	0.0838	0.0380
Round Drafted	0.0148	0.0103	0.0883	0.0378
Passes Intercepted on Defense	0.0227	0.0024	0.0841	0.0364
trans count	0.0530	0.0423	0.0105	0.0352
autograph	0.0265	0.0306	0.0367	0.0313
Draft Year	0.0053	0.0846	0.0030	0.0310
Pro Bowl Selections	0.0109	0.0072	0.0616	0.0266
card version_Silver	0.0049	0.0238	0.0360	0.0216
3Cone	0.0024	0.0466	0.0000	0.0163
Age	0.0061	0.0000	0.0419	0.0160
Weighted Accumulated Approx Value for Team That Drafted	0.0288	0.0036	0.0100	0.0141
Solo Tackles	0.0127	0.0252	0.0044	0.0141
Vertical	0.0100	0.0141	0.0095	0.0112
Games Played	0.0164	0.0138	0.0031	0.0111
Ht (in)	0.0052	0.0045	0.0228	0.0108
Draft Pick	0.0108	0.0088	0.0091	0.0096
Bench	0.0259	0.0002	0.0020	0.0094

5.3.3 Defensive End

Feature importance for the Defensive End position is shown in Table 5.15. We see that Pro Bowl selection (Random Forest, Gradient Boosting, and XGBoost) is a stronger feature. Card grading (Random Forest, Gradient Boosting, and to a lesser extent, XGBoost), and transaction count (Random Forest and Gradient Boosting) are of the next greatest importance. Draft pick position is important (0.09) to Random Forest and Gradient Boost modeling (0.07). Draft year and years played are the most important features after Pro Bowl selection for XGBoost. We also see that certain team participation may contribute to increases in average price sold – playing at Ohio State and Oregon in college may reflect tougher defensive programs that produce players who have rookie cards sold at higher prices. We see a similar trend for playing professional football for players on the Washington Commanders, Green Bay Packers, and New York Giants. Position-specific performance metrics that demonstrate top 20 importance for the Defensive End position include Pro Bowl selection, draft pick, solo tackles, weighted career approximate value, years as a starter, sacks, and bench strength.

5.3.4 Defensive Tackle

Card grading is of strong importance for the Defensive Tackle position, as shown in Table 5.16. Among the most important features for this position are autograph, shuttle, and 40-yard dash times. Playing professionally for Philadelphia is important for all models. Playing college ball for the University of Georgia is also of moderate importance to all models. Transaction count is moderately important for the Random Forest model. Sacks, vertical jump, age, broad jump, weight, height, draft pick, draft round, and bench performance were important features for the XGBoost model. Position-specific performance metrics among the top 20 features include shuttle, 40-yard dash, games played, sacks, vertical jump, weighted career value, broad jump, draft pick, draft round, years as a primary starter, and bench.

Table 5.15: DE Top 20 Features by Importance Across Tree-based Models. Features at the top of the list are of greater average, across all tree-based models, than those lower on the list. Darker blue coloration represents higher feature importance while lighter blue coloration corresponds with lower feature importance.

DE Top 20 Features by Importance Across Tree-based Models

Feature	RandomForest	GradientBoosting	XGBoost	Average
Pro Bowl Selections	0.2643	0.2803	0.4282	0.3243
card grading	0.2800	0.2518	0.0500	0.1939
trans count	0.1633	0.1698	0.0207	0.1179
Draft Pick	0.0922	0.0714	0.0185	0.0607
Draft Year	0.0234	0.0171	0.0930	0.0445
Years Played	0.0074	0.0212	0.1029	0.0439
autograph	0.0404	0.0484	0.0137	0.0342
College/Univ_Ohio St.	0.0036	0.0215	0.0291	0.0181
College/Univ_Oregon	0.0026	0.0018	0.0436	0.0160
Team_WAS	0.0108	0.0171	0.0184	0.0154
card version_Silver	0.0190	0.0204	0.0048	0.0147
Team_GNB	0.0024	0.0042	0.0309	0.0125
Age	0.0098	0.0123	0.0130	0.0117
Solo Tackles	0.0093	0.0171	0.0082	0.0116
Weighted Career Approx Value	0.0056	0.0104	0.0165	0.0109
Team_NYG	0.0027	0.0022	0.0239	0.0096
Years as Primary Starter	0.0016	0.0003	0.0151	0.0057
card version_Disco	0.0041	0.0083	0.0023	0.0049
Sacks	0.0053	0.0005	0.0080	0.0046
Bench	0.0031	0.0001	0.0105	0.0046

Table 5.16: DT Top 20 Features by Importance Across Tree-based Models. Features at the top of the list are of greater average, across all tree-based models, than those lower on the list. Darker blue coloration represents higher feature importance while lighter blue coloration corresponds with lower feature importance.

DT Top 20 Features by Importance Across Tree-based Models

Feature	RandomForest	GradientBoosting	XGBoost	Average
card grading	0.5583	0.5733	0.1716	0.4344
autograph	0.1119	0.1397	0.0612	0.1043
Team_PHI	0.1044	0.0778	0.1085	0.0969
Shuttle	0.0308	0.0693	0.0649	0.0550
40yd	0.0232	0.0025	0.0943	0.0400
Games Played	0.0118	0.0115	0.0898	0.0377
College/Univ_Georgia	0.0137	0.0377	0.0607	0.0373
trans count	0.0366	0.0134	0.0155	0.0218
Sacks	0.0122	0.0168	0.0363	0.0217
Vertical	0.0038	0.0000	0.0399	0.0145
Age	0.0122	0.0013	0.0247	0.0127
Weighted Career Approx Value	0.0061	0.0049	0.0216	0.0109
Broad Jump	0.0031	0.0008	0.0279	0.0106
Wt (lbs)	0.0053	0.0024	0.0240	0.0106
Ht (in)	0.0107	0.0003	0.0207	0.0105
card version_Silver	0.0095	0.0054	0.0110	0.0087
Draft Pick	0.0028	0.0004	0.0212	0.0082
Round Drafted	0.0004	0.0000	0.0234	0.0079
Bench	0.0014	0.0004	0.0193	0.0070
Years as Primary Starter	0.0041	0.0115	0.0042	0.0066

5.3.5 Linebacker

See Table 5.17 for the top 20 features for the Linebacker position. Card grading, autograph, and transaction count are the most important features for the Random Forest and Gradient Boosting models. 1st team all pro selection and years played are most important for the XGBoost model. Card grading, autograph, Pro Bowl selection, games played, playing professionally for Dallas, playing college ball for the University of Texas, and weighted approximate career value were the other most important features for the XGBoost model. Pro Bowl selection, sacks, games played, approximate value for drafting team, playing college ball for Penn State, and career approximate value are the next more important features for the Random Forest model. Pro Bowl selection, sacks, games played, approximate value to drafting team, and playing for Penn State in college were the next most important features for the Gradient Boosting model. Position-specific performance features in the top 20 for Linebackers include 1st team all pro selection, years played, Pro Bowl selection, sacks, games played, approximate value for drafting teams, approximate career value, shuttle, draft pick, 3 cone, years as a primary starter, solo tackles, and 40-yard.

5.3.6 Offensive Line

Card grading is the most important feature for the Offensive Line position across all models, as seen in Table 5.18. This single feature comprises over 60% of feature importance for the Random Forest and Gradient Boosting. In the figure below, we can see that it is relatively important to the XGBoost model, but there are other features over the 0.10 threshold for this model (weighted career value and 1st team all pro selection). Other important features for the XGBoost model (0.03-0.07) include autograph, approximate value to drafting team, Pro Bowl selection, years as a primary starter, transaction count, vertical jump, and age. Position-specific features of importance for tree-based models include career value, value to drafting team, 1st team all pro selection, Pro Bowl selection, years as a primary starter, vertical jump, bench, games played, 40-yard dash, shuttle, years played, and draft pick.

Table 5.17: LB Top 20 Features by Importance Across Tree-based Models. Features at the top of the list are of greater average, across all tree-based models, than those lower on the list. Darker blue coloration represents higher feature importance while lighter blue coloration corresponds with lower feature importance.

LB Top 20 Features by Importance Across Tree-based Models

Feature	RandomForest	GradientBoosting	XGBoost	Average
card grading	0.2998	0.2382	0.0501	0.1960
autograph	0.1658	0.2329	0.0908	0.1631
trans count	0.1347	0.1886	0.0106	0.1113
1st Team All Pro Selections	0.0182	0.0475	0.2144	0.0933
Years Played	0.0052	0.0026	0.1650	0.0576
Pro Bowl Selections	0.0270	0.0608	0.0385	0.0421
Sacks	0.0643	0.0359	0.0107	0.0370
Games Played	0.0331	0.0374	0.0226	0.0310
Weighted Accumulated Approx Value for Team That Drafted	0.0320	0.0478	0.0071	0.0290
Team_DAL	0.0137	0.0056	0.0655	0.0283
College/Univ_Texas	0.0108	0.0000	0.0651	0.0253
College/Univ_Penn St.	0.0293	0.0339	0.0000	0.0210
Weighted Career Approx Value	0.0275	0.0041	0.0286	0.0201
Shuttle	0.0165	0.0118	0.0103	0.0128
Draft Pick	0.0092	0.0102	0.0174	0.0122
3Cone	0.0092	0.0050	0.0181	0.0108
Years as Primary Starter	0.0029	0.0032	0.0191	0.0084
Solo Tackles	0.0061	0.0064	0.0091	0.0072
card version_Silver	0.0178	0.0006	0.0023	0.0069
40yd	0.0102	0.0016	0.0087	0.0068

Table 5.18: OL Top 20 Features by Importance Across Tree-based Models. Features at the top of the list are of greater average, across all tree-based models, than those lower on the list. Darker blue coloration represents higher feature importance while lighter blue coloration corresponds with lower feature importance.

OL Top 20 Features by Importance Across Tree-based Models

Feature	RandomForest	GradientBoosting	XGBoost	Average
card grading	0.6172	0.6296	0.2919	0.5129
Weighted Career Approx Value	0.0382	0.0461	0.1118	0.0654
autograph	0.0580	0.0422	0.0403	0.0468
Weighted Accumulated Approx Value for Team That Drafted	0.0309	0.0531	0.0468	0.0436
1st Team All Pro Selections	0.0043	0.0024	0.1128	0.0398
Pro Bowl Selections	0.0104	0.0154	0.0699	0.0319
Years as Primary Starter	0.0129	0.0142	0.0509	0.0260
trans count	0.0318	0.0070	0.0336	0.0241
Vertical	0.0140	0.0028	0.0505	0.0224
Age	0.0057	0.0024	0.0550	0.0210
Bench	0.0261	0.0280	0.0029	0.0190
Games Played	0.0282	0.0260	0.0010	0.0184
Draft Year	0.0111	0.0407	0.0006	0.0175
Ht (in)	0.0053	0.0058	0.0308	0.0140
card version_Silver	0.0080	0.0058	0.0231	0.0123
card version_Disco	0.0070	0.0106	0.0175	0.0117
40yd	0.0067	0.0053	0.0224	0.0115
Shuttle	0.0113	0.0127	0.0037	0.0092
Years Played	0.0124	0.0145	0.0000	0.0090
Draft Pick	0.0065	0.0045	0.0115	0.0075

5.3.7 Outside Linebacker

See Table 5.19 for feature importance to the Offensive Linebacker position. Similar to the finding with Offensive Linemen, card grading is the single most important feature for all models, contributing over 60% to the Random Forest and Gradient Boosting models and almost 50% with the XGBoost model. Features that were also of elevated importance to the Random Forest model are autograph, career value, bench, transaction count, and value to drafting team. For the Gradient Boosting model, autograph, Pro Bowl selection, career value, age, and height were of heightened importance. Pro Bowl selection, career value, bench, value to drafting team, years as a starter, shuttle, games played, draft pick position, vertical jump, years played, 40-yard, and broad jump were position-specific features in the top 20 features.

5.3.8 Quarterback

As seen in Table 5.20, autographs were the single most important feature for the Random Forest and Gradient Boosting models. For the Random Forest model, interceptions thrown, transaction count, passes completed, card grading, passes attempted, passing touchdowns, Silver card variation, passing yards, age, and rushing yards were the most important features. Interceptions thrown, transaction count, completed passes, rushing touchdowns, passing touchdowns, Silver card variation, age, and Pro Bowl selection were the most important features for the Gradient Boosting model. Interceptions thrown and passes completed were the most important features for the XGBoost model by far, with autograph, transaction count, card grading, rushing attempts, rushing touchdowns, weight, and Pro Bowl selections also being of importance to a much lesser extent. Interceptions thrown, passes completed, rushing attempts, passes attempted, rushing touchdowns, passing touchdowns, yards gained by passing, Pro Bowl selection, rushing yards gained, career value, solo tackles, value to drafting team, and years as a primary starter were all top 20 features for this position. Except for value to drafting team, performance metrics of importance were all from time in

Table 5.19: OLB Top 20 Features by Importance Across Tree-based Models. Features at the top of the list are of greater average, across all tree-based models, than those lower on the list. Darker blue coloration represents higher feature importance while lighter blue coloration corresponds with lower feature importance.

OLB Top 20 Features by Importance Across Tree-based Models

Feature	RandomForest	GradientBoosting	XGBoost	Average
card grading	0.6188	0.6152	0.4696	0.5679
autograph	0.0596	0.0576	0.0835	0.0669
Pro Bowl Selections	0.0106	0.0333	0.1279	0.0573
Weighted Career Approx Value	0.0354	0.0537	0.0582	0.0491
Bench	0.0261	0.0767	0.0028	0.0352
Age	0.0026	0.0005	0.0812	0.0281
Ht (in)	0.0041	0.0032	0.0734	0.0269
trans count	0.0291	0.0232	0.0148	0.0224
Weighted Accumulated Approx Value for Team That Drafted	0.0415	0.0141	0.0000	0.0185
Years as Primary Starter	0.0120	0.0348	0.0000	0.0156
Shuttle	0.0157	0.0089	0.0209	0.0152
Games Played	0.0194	0.0149	0.0045	0.0129
Draft Pick	0.0135	0.0079	0.0138	0.0117
card version_Disco	0.0070	0.0109	0.0091	0.0090
Vertical	0.0081	0.0141	0.0033	0.0085
Years Played	0.0130	0.0019	0.0103	0.0084
40yd	0.0089	0.0042	0.0088	0.0073
Draft Year	0.0081	0.0066	0.0019	0.0055
Wt (lbs)	0.0098	0.0031	0.0028	0.0052
Broad Jump	0.0066	0.0045	0.0032	0.0048

the NFL. Combine performance and draft metrics did not make it into the top 20 features as with other performance, possibly suggesting that NFL performance is uniquely key to card values for this position. Of note, we also see that playing for San Francisco may help boost card values. Cards for this position likely sell for higher values when autographed, graded, or when it is a Silver variant.

Table 5.20: QB Top 20 Features by Importance Across Tree-based Models. Features at the top of the list are of greater average, across all tree-based models, than those lower on the list. Darker blue coloration represents higher feature importance while lighter blue coloration corresponds with lower feature importance.

QB Top 20 Features by Importance Across Tree-based Models

Feature	RandomForest	GradientBoosting	XGBoost	Average
autograph	0.2049	0.2407	0.1148	0.1868
Interceptions Thrown	0.1324	0.1492	0.2441	0.1752
trans count	0.1752	0.2061	0.0221	0.1344
Passes Completed	0.0430	0.0216	0.2898	0.1181
card grading	0.1213	0.1506	0.0213	0.0978
Rushing Attempts (sacks not included)	0.0187	0.0317	0.0694	0.0400
Passes Attempted	0.0272	0.0726	0.0013	0.0337
Rushing Touchdowns	0.0219	0.0258	0.0485	0.0321
Passing Touchdowns	0.0496	0.0348	0.0026	0.0290
card version_Silver	0.0480	0.0222	0.0131	0.0278
Yards Gained By Passing	0.0378	0.0049	0.0075	0.0168
Age	0.0182	0.0198	0.0108	0.0163
Wt (lbs)	0.0059	0.0002	0.0305	0.0122
Pro Bowl Selections	0.0052	0.0105	0.0190	0.0115
Rushing Yards Gained (sacks not included)	0.0194	0.0002	0.0020	0.0072
Weighted Career Approx Value	0.0073	0.0002	0.0122	0.0066
Solo Tackles	0.0018	0.0000	0.0127	0.0049
Team_SFO	0.0005	0.0004	0.0126	0.0045
Weighted Accumulated Approx Value for Team That Drafted	0.0098	0.0003	0.0032	0.0044
Years as Primary Starter	0.0056	0.0002	0.0064	0.0041

5.3.9 Running Back

As shown in Table 5.21 for Running Backs, card grading, autographs, rushing touchdowns, Pro Bowl selection, transaction count, receiving yards, value to drafting team, and playing professionally for Atlanta are the most important features in the top 20 features for the Random Forest and Gradient Boosting models. Card grading and autograph are the most important features to both models by far. Rushing touchdowns and receiving yards are the most important features in the XGBoost above 0.10 importance, with card grading, autographs, Pro Bowl selection, value for drafting team, playing professionally for Atlanta, playing college ball for the University of Texas, age, games played, vertical jump, 1st team all pro selection, receiving touchdowns, receiving receptions, and the Red White Blue variant also being of importance. As with the Quarterback position, we see several NFL performance metrics falling into the top 20 features: rushing touchdowns, Pro Bowl selection, receiving yards, value to drafting team, games played, career value, vertical jump, 1st team all pro selection, receiving touchdowns, receiving receptions, years, played, and rushing attempts.

5.3.10 Safety

Card grading is the most important feature for Safeties in the Random Forest and Gradient Boosting models, making up approximately 40%. See Table 5.22 for results. Value to drafting team, autograph, sacks, weight, and transaction count were also of increased importance to both models, with vertical jump and solo tackles also making the list for the Gradient Boosting model. Card grading and value to drafting team are the most important features for the XGBoost model, with autographs, sacks, weight, vertical jump, solo tackles, Disco card variation, Lazer card variation, 40-yard, 1st team all pro selection, and age also being of increased importance. Value for drafting team, sacks, vertical jump, solo tackles, broad jump, games played, career value, bench, 40-yard, and 1st team all pro selection are the important performance features related to this position.

Table 5.21: RB Top 20 Features by Importance Across Tree-based Models. Features at the top of the list are of greater average, across all tree-based models, than those lower on the list. Darker blue coloration represents higher feature importance while lighter blue coloration corresponds with lower feature importance.

RB Top 20 Features by Importance Across Tree-based Models

Feature	RandomForest	GradientBoosting	XGBoost	Average
card grading	0.2085	0.1898	0.0329	0.1437
autograph	0.1792	0.1840	0.0539	0.1390
Rushing Touchdowns	0.0804	0.0689	0.1554	0.1016
Pro Bowl Selections	0.0502	0.0710	0.0762	0.0658
trans count	0.0823	0.0936	0.0097	0.0619
Receiving Yards	0.0292	0.0203	0.1253	0.0582
Weighted Accumulated Approx Value for Team That Drafted	0.0631	0.0576	0.0455	0.0554
Team_ATL	0.0476	0.0503	0.0392	0.0457
College/Univ_Texas	0.0193	0.0213	0.0713	0.0373
Age	0.0091	0.0131	0.0696	0.0306
Games Played	0.0067	0.0252	0.0481	0.0267
card version_Silver	0.0249	0.0273	0.0120	0.0214
Weighted Career Approx Value	0.0223	0.0342	0.0032	0.0199
Vertical	0.0137	0.0051	0.0405	0.0197
1st Team All Pro Selections	0.0099	0.0067	0.0305	0.0157
Receiving Touchdowns	0.0037	0.0016	0.0361	0.0138
Receiving Receptions	0.0079	0.0047	0.0274	0.0133
Years Played	0.0107	0.0146	0.0100	0.0118
Rushing Attempts (sacks not included)	0.0107	0.0095	0.0068	0.0090
card version_Red White Blue	0.0036	0.0026	0.0189	0.0084

Table 5.22: S Top 20 Features by Importance Across Tree-based Models. Features at the top of the list are of greater average, across all tree-based models, than those lower on the list. Darker blue coloration represents higher feature importance while lighter blue coloration corresponds with lower feature importance.

S Top 20 Features by Importance Across Tree-based Models

Feature	RandomForest	GradientBoosting	XGBoost	Average
card grading	0.4046	0.3987	0.2001	0.3345
Weighted Accumulated Approx Value for Team That Drafted	0.0853	0.0920	0.2364	0.1379
autograph	0.1350	0.1662	0.0737	0.1250
Sacks	0.0301	0.0435	0.1377	0.0704
Wt (lbs)	0.0623	0.0409	0.0617	0.0550
trans count	0.0745	0.0446	0.0152	0.0448
Vertical	0.0132	0.0545	0.0510	0.0396
Solo Tackles	0.0182	0.0520	0.0326	0.0343
card version_Silver	0.0246	0.0160	0.0170	0.0192
Broad Jump	0.0137	0.0178	0.0207	0.0174
Games Played	0.0179	0.0183	0.0024	0.0129
Weighted Career Approx Value	0.0145	0.0152	0.0073	0.0123
Ht (in)	0.0169	0.0069	0.0104	0.0114
Bench	0.0152	0.0070	0.0086	0.0103
card version_Disco	0.0061	0.0024	0.0200	0.0095
Draft Pick	0.0130	0.0064	0.0040	0.0078
card version_Lazer	0.0042	0.0000	0.0184	0.0075
40yd	0.0036	0.0000	0.0163	0.0066
1st Team All Pro Selections	0.0009	0.0000	0.0188	0.0066
Age	0.0020	0.0009	0.0140	0.0056

5.3.11 Tight End

As we can see in Table 5.23, card grading and autographs are the most important Tight End features for the Random Forest and Gradient Boosting models, with other features sharing reduced importance for these models: playing professionally for Detroit, transaction count, playing college ball at the University of Iowa, receiving receptions, weight, years as a starter, years played, and Silver card variation (Random Forest model only). Card grading, autographs, playing professionally for Detroit, Pro Bowl selection, career value, draft year, receiving receptions, weight, years as a starter, round drafted, 40-yard, solo tackles, playing in college for Colorado State, bench, and playing professionally for Buffalo were the most important features for the XGBoost model. Pro Bowl selections, receiving receptions, years as a starter, years played, receiving yards, 40-yard, solo tackles, and bench are the position-based performance metrics most important to the Tight End position.

5.3.12 Wide Receiver

As shown in Table 5.24, autographs, grading, transaction count, value to drafting team, receiving receptions, and receiving yards were the most important features for the Wide Receiver position in the Random Forest and Gradient Boosting models. Also of high importance to the Random Forest model were Pro Bowl selection, receiving touchdowns, career value, passes completed, Silver card variation, rushing attempts, and 1st team selection. Yards gained by passing, weight, playing in college for Bringham Young University (BYU), and solo tackles were of additional importance for the Gradient Boosting model. Pro Bowl selection was the most important feature for Wide Receivers in the XGBoost model, with autograph, grading, transaction count, value to drafting team, Pro Bowl selection, receiving touchdowns, career value, passes completed, weight, playing for BYU in college, draft pick, and rushing attempts also being important features. Position-related performance metrics of importance for this role include value to drafting team, Pro Bowl selection, receiving receptions, receiving touchdowns, receiving yards, career value, yards gained by passing, passes

Table 5.23: TE Top 20 Features by Importance Across Tree-based Models. Features at the top of the list are of greater average, across all tree-based models, than those lower on the list. Darker blue coloration represents higher feature importance while lighter blue coloration corresponds with lower feature importance.

TE Top 20 Features by Importance Across Tree-based Models

Feature	RandomForest	GradientBoosting	XGBoost	Average
card grading	0.2746	0.2712	0.0730	0.2063
autograph	0.1948	0.2307	0.1048	0.1768
Team_DET	0.0490	0.0582	0.1778	0.0950
trans count	0.1052	0.1049	0.0154	0.0751
Pro Bowl Selections	0.0194	0.0137	0.1006	0.0446
Weighted Career Approx Value	0.0184	0.0041	0.1052	0.0426
Draft Year	0.0119	0.0150	0.0838	0.0369
College/Univ_Iowa	0.0474	0.0616	0.0000	0.0363
Receiving Receptions	0.0334	0.0252	0.0494	0.0360
Wt (lbs)	0.0222	0.0170	0.0395	0.0262
Years as Primary Starter	0.0116	0.0140	0.0506	0.0254
Years Played	0.0192	0.0505	0.0046	0.0248
Receiving Yards	0.0168	0.0559	0.0000	0.0242
card version_Silver	0.0320	0.0119	0.0074	0.0171
Round Drafted	0.0108	0.0111	0.0277	0.0166
40yd	0.0155	0.0087	0.0193	0.0145
Solo Tackles	0.0018	0.0001	0.0354	0.0124
College/Univ_Colorado St.	0.0129	0.0074	0.0151	0.0118
Bench	0.0022	0.0057	0.0249	0.0109
Team_BUF	0.0038	0.0020	0.0209	0.0089

completed, draft pick, rushing attempts, solo tackles, 1st team all pro selection, and games played.

Table 5.24: WR Top 20 Features by Importance Across Tree-based Models. Features at the top of the list are of greater average, across all tree-based models, than those lower on the list. Darker blue coloration represents higher feature importance while lighter blue coloration corresponds with lower feature importance.

WR Top 20 Features by Importance Across Tree-based Models

Feature	RandomForest	GradientBoosting	XGBoost	Average
autograph	0.2028	0.1826	0.0696	0.1517
card grading	0.1553	0.1361	0.0220	0.1045
trans count	0.1113	0.1540	0.0312	0.0988
Weighted Accumulated Approx Value for Team That Drafted	0.0878	0.1013	0.0902	0.0931
Pro Bowl Selections	0.0398	0.0107	0.2231	0.0912
Receiving Receptions	0.0370	0.0998	0.0150	0.0506
Receiving Touchdowns	0.0650	0.0277	0.0489	0.0472
Receiving Yards	0.0348	0.0799	0.0117	0.0421
Weighted Career Approx Value	0.0254	0.0115	0.0732	0.0367
Yards Gained By Passing	0.0092	0.0764	0.0036	0.0297
Passes Completed	0.0156	0.0004	0.0644	0.0268
Wt (lbs)	0.0099	0.0183	0.0503	0.0262
card version_Silver	0.0387	0.0206	0.0069	0.0221
College/Univ_BYU	0.0140	0.0217	0.0268	0.0208
Draft Pick	0.0088	0.0033	0.0481	0.0201
Rushing Attempts (sacks not included)	0.0177	0.0012	0.0372	0.0187
Solo Tackles	0.0033	0.0228	0.0122	0.0128
1st Team All Pro Selections	0.0105	0.0018	0.0109	0.0077
Games Played	0.0077	0.0045	0.0106	0.0076
Ht (in)	0.0059	0.0000	0.0160	0.0073

5.4 Discussion

Moving towards a precise pricing model for NFL trading cards will require additional efforts that seek to incorporate more longitudinal data, more trading card lines, and more features that can capture additional measures of player and sport popularity. This project makes strides in the direction of developing optimized pricing algorithms that begin to facilitate greater understanding of how certain performance metrics, player health and physical attributes, and card condition help us to explain card value variation. Overall, I find that there are correlates with average card values that bolster stronger learning models and that certain models optimize performance differently for different positions. This project highlights several consistent findings with prior work. Johnson [2023] and Jackson [2023] found, as I also find, that graded cards sell for higher amounts and player performance should ultimately inform card values. Primm et al. [2010] similarly posit that a player's performance and Pro Bowl participation should influence card values – support for that argument is shown here. Barbatsis et al. [2024] show that touchdowns, interceptions, air yards per attempt, and yards per attempt contributed to the success of a model and this study lends support to their work. This paper shows strong positive relationships between passing touchdowns, passing yards, passing completion, passing attempts, passing interceptions, weighted value for drafting team, approximate weighted career value, draft pick, sacks, and Pro Bowl selection ($R^2 > 0.60$ for all features) with average selling price of a trading card. These features, in turn, are top 20 features for many of the optimized models by position.

While Brower [2020] found support for over-reliance on the 40-yard dash to predict NFL success, this study finds that it strengthens model performance for a number of positions, While NFL performance is not the same as trading card values, the two are likely intertwined and while this study cannot explicitly state such causality, it is clear that a number of NFL performance metrics are tied to model performance. Hendricks et al. [2003] and Berri and Simmons [2011] argue that collegiate performance at dominant colleges and universities and over-representation of players at popular events such as Pro Bowls can lead to over-evaluation of a player for drafting teams, but this paper lends support for Pro Bowl and draft event performance as important features for model performance. Reynolds et al. [2015] find that higher draft pick values correlate with playoff performance – both were found to be important features for models in this study. Pitts and Evans [2019] argue that drafting teams undervalue performance on draft metrics such as receiving yards for wide receivers. I find that this is likely the case with receiving yards landing in the top 20 important features to the tree-based models for the wide receiver position.

In evaluation of model performance across positions, I find varying degrees of optimization across the different types of models. Gradient Boosting provided the best optimization for the Tight End position ($R^2 = 0.93$) and for Outside Linebackers (0.68). Random Forest models gave the best R^2 performance for Quarterbacks (0.80) and Cornerbacks (0.70). XGBoost delivered optimized model performance of R^2 values for the most positions: Wide Receivers (0.81), Running Backs (0.83), Linebackers (0.82), Defensive Tackles (0.59), Defensive Ends (0.78), and Defensive Backs (0.89). Overall, KNN Regressor modeling was not a strong performer – only the Safety position model was optimized with this type of model (R^2 of 0.66 versus 0.62 with ensemble modeling). Before stacking, we are best able to explain variation in the average selling price of a trading card iteration for Tight Ends (0.93 achieved with Gradient Boosting) and Defensive Backs (0.89 achieved with XGBoost. Model performance for R^2 was best optimized for each position with stacking (except for the Safety position, where KNN Regressor gave the best performance): Tight Ends (0.95), Wide Receivers (0.90), Running Backs (0.90), Quarterbacks (0.86), Outside Linebackers (0.98), Linebackers (0.98), Defensive Tackles (0.79), Defensive End (0.84), Defensive Back (0.91), and Cornerback (0.77).

When examining the top 20 features of importance across tree-based models, we see that several features are important across positions. Card grading, autographs, and transaction counts made the top of the list for all positions: Cornerbacks, Defensive Backs, Defensive Ends, Defensive Tackles, Linebackers, Offensive Linemen, Outside Linebackers, Quarterbacks, Running Backs, Safeties, Tight Ends, and Wide Receivers. While not necessarily at the top of the top 20 features, draft pick was on the list for all positions but Quarterbacks, Running Backs, and Tight Ends (round drafted was in the top 20 for this position). Pro

Bowl selection was important for Wide Receivers, Tight Ends, Running Backs, Quarterbacks, Outside Linebackers, Offensive Linemen, Linebackers, and Defensive Ends. Overall, I find NFL and Combine performance features in the top 20 list for each position. It is important that future work continues to assess how these features differently inform model performance across positions. It is interesting to note that either weighted career value or approximate value to drafting teams (or both) are on the top feature list for every position. As suggested by Lewis [2003], Yurko et al. [2019], and Citrone and Ventura [2017], I find support for approximate valuations of career and team contributions as important features in determining what will drive average card price.

Future work should examine the fit of these models on longitudinal pricing data. While outside the scope of this project, it is likely that model performance will improve from such work. Additionally, approximate career value and value to drafting team should be further evaluated for model optimization. I believe that these features are key to understanding our ability to predict player performance as these have proven to be key features for every position under review (either both or one is included in top features for every position). This suggests that we have made progress in understanding how to predict player performance in the NFL and the average selling price of a player's rookie card. I also find support for physical attributes and card condition playing a role in card values. Primm et al (2010) suggest that a player's off-field behavior and their perceived level of skill can play a role in what trading cards will sell for – I was unable to find support for this within the scope of this project, but future work should be able to highlight some of the social mechanisms at play in the relationship between card values and player popularity. Efforts towards this end will only continue to remove "noise" from our ability to accurately predict trading card values.

Chapter 6

Conclusion and Future Work

This thesis demonstrates that NFL trading card prices can be effectively modeled using a combination of player performance metrics, card attributes, and contextual features like draft pick and Pro Bowl selections. Through the application of tree-based ensemble learning methods, most notably XGBoost and Gradient Boosting, predictive models were optimized for nearly every NFL position, with stacking models providing the highest R^2 scores overall. Also, player features such as career value, value to the drafting team, card grading, and Pro Bowl selections consistently emerged as top predictors across models. The key takeaway from this work is that NFL on-field performance is intricately connected to trading card value, even if causality cannot be directly established. By integrating diverse data sources and tailoring models to position-specific attributes, this study offers a strong foundation for predictive pricing in the sports card market. It also lends support to existing literature asserting the influence of performance and media visibility on card values.

Future work should consider expanding the dataset to include a greater number of players at each position, which could reduce overfitting, improve model stability, and offer more accurate position-specific predictions. Also, future research could focus on developing models based solely on collegiate performance data to explore the feasibility of predicting NFL success – and, by extension, card value – prior to a player's professional debut. This approach could offer valuable insights for early-stage investment or scouting strategies. Lastly, I briefly explored sentiment analysis through web scraping data from the social platform X (formerly Twitter), but experienced limited access to the platform's Application Programming Interface (API) due to subscription status. Incorporating social sentiment, whether positive or negative, could provide an important feature for player popularity and media presence, enabling a more holistic modeling approach that accounts for both performance-based and perception-based drivers of card value.

Overall, this work makes a meaningful contribution to both the fields of sports analytics and collectible valuation by offering a data-driven framework for modeling trading card prices and further understanding player performance contributions. It provides a practical blueprint for future research at the intersection of analytics, consumer behavior, and financial forecasting within the collectibles market. Moreover, the results demonstrate the importance of position-specific modeling, which highlights the value of tailoring predictive approaches to the unique position-specific performance indicators. These findings have potential applications for draft selection by NFL teams as it provides insight into potential features that can help optimize a player's contribution to a given team. It is also reasonable to assume that these same contributing factors could help develop more accurate drafting in Fantasy Football leagues.

BIBLIOGRAPHY

Liz Aldrich and Ashley. McKilips. “Study: These factors can help you predict sports card prices.”. https://moneymade.io/learn/article/sports-card-price-study, 2022.

Alex Barbatsis, Isaac Rhee, and Jimmy Zhang. “Using machine learning and coillege profiles to predict nfl success.”. *Northwestern Sports Analytics Group*, 2024.

Casey Barney, Anthony Caravella, Michael Cullen, Gary Jackson, and Craig E. Wills. “Evaluating talent acquisition via the nfl draft. technical report wpi-cs-tr-13-01.”. Technical report, Worchester, MA: Computer Science Department, Worcester Polytechnic Institute., 2013. URL https://web.cs.wpi.edu/~cew/papers/nflupdate14.pdf.

Ryan. Barone. “Which panini football cards are the best?”. https://ballcardgenius.com/blog/which-panini-cards-are-the-best/, 2023.

Beckett. Beckett.com. https://beckett.com, 1984.

David J. Berri and Rob Simmons. Catching a draft: on the process of selecting quarterbacks in the national football league amateur draft. *Journal of Productivity Analysis*, 35(1):37–49, 2011. ISSN 0895562X, 15730441. URL http://www.jstor.org/stable/23883795.

Owen Brower. *Applying Analytics to the NFL Draft: Athletic Performance Measures as a Predictor of Future Success*. Student Research and Creative Works. University of Puget Sound, JSTOR, 2020. URL https://jstor.org/stable/community.36513983.

LLC Card Hedge. Card hedger. https://www.cardhedger.com/price-guide/football, 2020.

Robert Nicholas Citrone and Sam. Ventura. “A statistical analysis of the nfl draft: Valuing draft picks and predicting future player success.”. In *Baltimore JSM - Statistics: It's Essential.*, 2017.

Jason. Clarke. “The best football card boxes and packs to buy (2025).”. https://sportscardspecialist.com/best-football-card-boxes-to-buy/, 2025.

Eric Eager and George. Chahrouri. Pff war: Modeling player value in american football. In *MIT Sloan Sports Analytics Conference*, 2020. URL https://www.sloansportsconference.com/conference/2020-conference#research-papers.

eBay. ebay.com. https://www.ebay.com, 1995.

GiantSportsCards. The oldest football card: What was the first football card ever made? https://giantsportscards.com/blogs/blog/the-oldest-football-card-what-was-the-first-football-card-ever-made?srsltid=AfmBOooAqSQlYNGg2bfRW2bbbfq_3erocWtCLRmkdt3YhNZ5CsIia154, 2024.

Jahn K. Hakes and Raymond D. Sauer. An economic evaluation of the moneyball hypothesis. *Journal of Economic Perspectives*, 20(3):173–186, September 2006. doi: 10.1257/jep.20.3.173. URL https://www.aeaweb.org/articles?id=10.1257/jep.20.3.173.

Wallace Hendricks, Lawrence DeBrock, and Roger Koenker. Uncertainty, hiring, and subsequent performance: The nfl draft. *Journal of Labor Economics*, 21(4):857–886, 2003. URL https://EconPapers.repec.org/RePEc:ucp:jlabec:v:21:y:2003:i:4:p:857-886.

Paul. Hickey. No offseason.com. https://nooffseason.com/about/, 2020.

Leon. Jackson. “How to tell if football cards are worth money: Unlock their value now.”. https://huffsports.com/football/how-to-tell-if-football-cards-are-worth-money/, 2023.

Matt. Johnson. “Are football cards worth money? collecting football cards.”. https://gradedcardinvestor.com/collecting-football-cards, 2023.

Michael. Lewis. *Moneyball*. W.W. Norton and Company, 2003.

MavinWorks LLC. Card mavin. https://cardmavin.com/category/football, 2016.

Cade Massey and Richard H. Thaler. ”Overconfidence vs. market efficiency in the national football league (april 2005)”. *NBER Working Paper No. w11270*, 2005. URL https://www.nber.org/system/files/working_papers/w11270/w11270.pdf.

Laine Pearce. 130 point. https://130point.com/about/, 2016.

Joshua Pitts and Brent Evans. Evidence on the importance of cognitive ability tests for nfl quarterbacks: what are the relationships among wonderlic scores, draft positions and nfl performance outcomes? *Applied Economics*, 50(27):2957–2966, 2018. URL https://EconPapers.repec.org/RePEc:taf:applec:v:50:y:2018:i:27:p:2957-2966.

Joshua D. Pitts and Brent Evans. Drafting for success: How good are nfl teams at identifying future productivity at offensive-skill positions in the draft? *The American Economist*, 64 (1):102–122, 2019.

Darry. Port. “Best football cards to buy in 2023 from brady to mahomes.”. https://moneymade.io/learn/article/best-football-cards-to-buy-2021, 2022.

Eric Primm, Robert M. Regoli, and John D. Hewitt. Where are we and how did we get here? assessing race, performance, and position in topps football card placement. *Journal of African American Studies*, 12(1):73–84, 2008. doi: 10.1007/s12111-007-9032-3. URL https://doi.org/10.1007/s12111-007-9032-3.

Eric Primm, Nicole Leeper Piquero, Robert M. Regoli, and Alex R. Piquero. The role of race in football card prices. *Social Science Quarterly*, 91(1):129–142, 2010. doi: https:

//doi.org/10.1111/j.1540-6237.2010.00685.x. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1540-6237.2010.00685.x.

PSA. Psa card authentication. https://www.psacard.com/info/about-psa, 1991.

Zachary Reynolds, Traci Bonds, Steele Thompson, and Carrie W. LeCrom. "Deconstructing the draft: An evaluation of the nfl draft as a predictor of team success.". *Journal of Applied Sport Management: Vol. 7 : Iss. 3.*, 2015. URL https://trace.tennessee.edu/jasm/vol7/iss3/12.

LLC Sports Reference. Pro football reference. https://www.pro-football-reference.com, 2000.

SportsCardsPro. Sportscardspro.com. https://www.sportscardspro.com, 2020.

Ronald Yurko, Samuel Ventura, and Maksim Horowitz. nflWAR: a reproducible method for offensive player evaluation in football. *Journal of Quantitative Analysis in Sports*, 15(3): 163–183, September 2019. doi: 10.1515/jqas-2018-0010. URL https://ideas.repec.org/a/bpj/jqsprt/v15y2019i3p163-183n1.html.

Display Zone. "Investing in the future: Predicting the next big thing in valuable football trading cards". https://displayzoneshop.com/blogs/news/investing-in-the-future-predicting-the-next-big-thing-in-valuable-football-trading-cards?srsltid=AfmBOorSAsLmd59KrTWAvqQObyJdnVl_CqiI6Ax_f6hyibvwcuDbIUCp, 2023.