Predicting NFL Quarterback Contracts and its Value Using Machine Learning Models
By Ece Melisa Kocak

1 Introduction

The National Football League, or NFL, is a global leader in revenue and viewership among all professional sports. NFL revenue was reported at $326 million in 1979 ("N.F.L. Net in 1970 Put At," 1981), $1.3 billion in 1990 (Associated Press, 1992), and exceeded $23 billion in 2024 (Peters, 2025). According to CNBC's Official NFL Team Valuations 2025, the average NFL franchise is worth $7.65 billion (Ozanian, 2025). The league's revenue extends far beyond ticket sales. Merchandise, sports betting and broadcasting rights have made the NFL billions of dollars. Operating at this scale makes every decision more significant, and every year teams commit hundreds of millions of dollars directly to player contracts. Contract decisions are incredibly impactful for team success and popularity, but there is no gold standard for how to evaluate players, and teams cannot afford to be wrong.

Among all positions, the quarterback plays the most financially and tactically significant role. Quarterbacks touch the ball on essentially every offensive snap, and are trusted to rapidly execute precise mechanics in the face of pressure. No other position has as much influence on the outcome of a game, season, or franchise. Nearly every year, another quarterback resets the contract market, and all of the highest annual value contracts belong to quarterbacks. There are around 2000 players in the NFL with another 81,000 in the college ranks (Football by the numbers, 2023). Despite the 83,000 football players at or near the professional level, roughly 15-25 of those players are widely considered NFL-caliber starting quarterbacks. So, quarterbacks are the most important position, they make more money than any other, and there are not enough high-quality options for every team to have a reliable starter year over year.

This leaves teams investing heavily into scouting at the college level. Although the draft allows teams to acquire their choice of player, scouting relies on projection rather than replicable science. Among the shortcomings in scouting, according to researcher Bailey Durkin, "teams, on average, overvalue their picks in the early rounds compared to picks in later ones. This phenomenon causes teams to pay higher salaries for players drafted earlier who produce less results than players drafted in later rounds with lower contract costs" (Durkin, 2020). This reflects an inefficiency of traditional scouting. As revenue and team staffs have ballooned, data-driven decision making has become more and more prevalent in the NFL.

Integrating data science and professional sports was popularized by the Oakland A's of the MLB. Their successful implementation of sabermetrics to produce efficient contract decisions has influenced analysts across professional sports. The A's' strategy was inextricably linked MLB salary systems where spending power is not limited. Operating one of the most cash-poor organizations incentivized a strategy built around undervalued metrics and players. The NFL employs a salary cap. All teams are limited to a set annual spending on contracts. Under a salary cap, player development and high-yield drafting are the incentivized strategies. Rookie contracts lock players into cheap deals that allow for outsized impact relative to spending. Competing in the NFL relies on staff scouting, drafting, and retaining one only around fifteen starter-quality quarterbacks. This clearly positions quarterback evaluation and contract negotiation as paramount for organizational success.

The goal of this project is to answer the research questions: "Can cumulative/per-season statistics predict if a quarterback will get a second contract after four years, and can they predict where that contract will rank in positional value?". This analysis will also attempt to identify which statistics can predict whether a quarterback will receive an above market contract as measured by total value. For this project, three machine learning models, Logistic Regression, Random Forest and Support Vector Machine (SVM) were trained on the NFL quarterback data using passing stats for each players' first four seasons in order to predict a second contract. Additionally, feature importance was used to interpret different statistical metrics' predictiveness of the outcome and relative contract value, measured by normalized total contract value. Understanding these factors provides insights into which performance metrics impact contract decisions, a proxy for the organizational value.

2 Data

2.1 Data Source

For the first research question, I retrieved an NFL QB Statistics dataset spanning 1970-2022 from Kaggle. The original data consists of 3177 rows and 17 columns. The features include players, passing yards, yards per attempt, attempts, completions, completion percentage, touchdowns, interceptions, QB rating, first downs, first downs percentage, 20+ yard completions, 40+ yard completions, longest single throw, number of sacks and total lost sack yardage. For the second research question, I retrieved another dataset from Kaggle containing NFL contracts and draft pick data for players between 2000-2023. The dataset includes draft year, round, pick number, team, player, number of games played in their career, search key, year signed, signing team, total US dollar value of their contract at signing and the total US dollar guaranteed value of their contract at signing. The dataset contained normalized value and guaranteed value to the NFL salary cap for that year which was crucial for comparison of contracts across years. The original NFL contracts and draft dataset included 12629 rows and 15 columns.

After filtering, the datasets for this project will be the passing statistics of the first four seasons of NFL quarterbacks drafted between 2000 and 2022 along with their contracts when applicable. Any player appearing in fewer than four seasons was excluded from the data. A four year sampling window was selected because drafted quarterbacks typically receive a contract of four years, after which point second contracts would be awarded. There are many exceptions to this rule, especially before the 2011 collective bargaining agreement that standardized contract length based on draft position. Current evaluation uses a four year timescale, and a uniform number of seasons is key for comparisons across players.

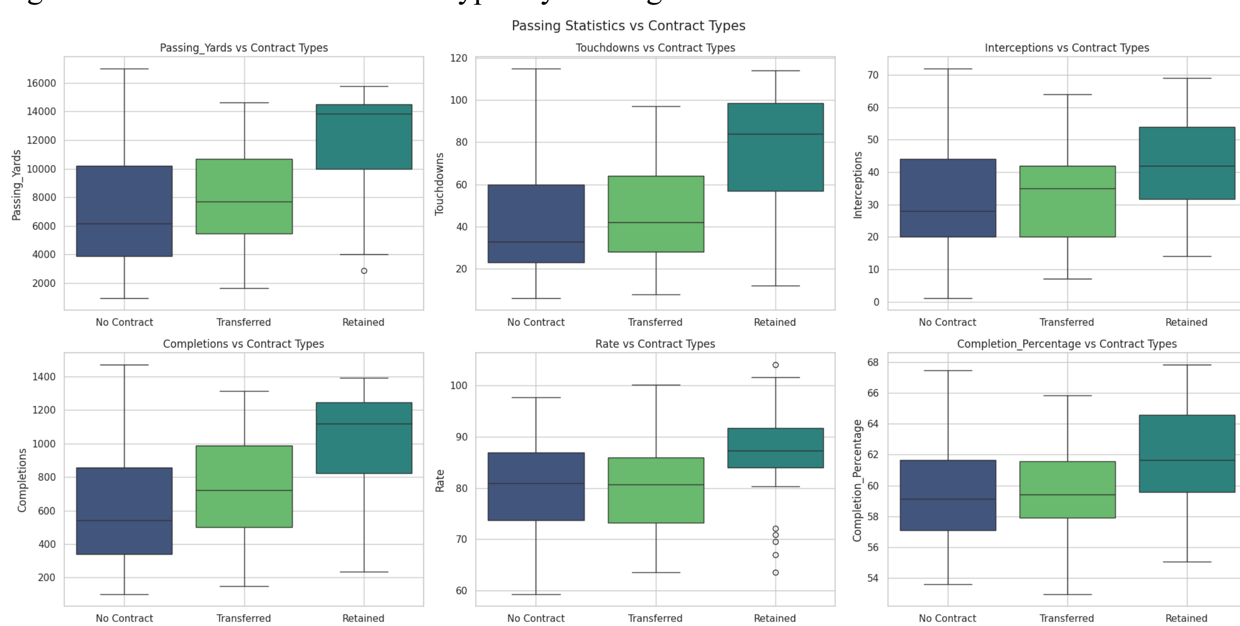
Despite access to data from previous seasons, the observation window was limited to 2000-2022 draftees as a reflection of the dynamics of the style of play in the NFL. Over time, statistics and quarterback standards have fluctuated with the dynamics and play style and officiating. Additionally, the total number of eligible entries is limited as a function of the number of teams in the NFL. With only 32 teams, there are limited roles available, and incentive for teams to retain their player limits turnover at the position. Because of the dynamic playstyle, career timelines, and set number of teams, 144 quarterbacks meet all listed criteria.

2.2 Exploratory Data Analysis

With the dataset of passing statistics, I calculated z-scores for the raw cumulative stats to identify whether quarterback performance was considered underperforming, average or elite. Elite and underperforming were defined by exceeding one standard deviation from the mean for that feature. Players were categorized into one of three groups for all statistics which identified strengths and weaknesses. This allowed for group-based comparisons like touchdowns vs completions (Figure 1). I also compared passing yards and interceptions to identify error prone/conservative quarterbacks. A third comparison between passing yards and completion percentage identified efficiency compared to volume (Figure 2).

With the second dataset, I grouped players by contract outcome: retained, transferred, or no contract. Retained and transferred were separate categories because many players are forced to become backup quarterbacks and accept a smaller role and contract, and this outcome is markedly different than retention for a starting role. Merging contract types with passing statistics, along with their z-score-based groups helped answer the first research question (Figure 4). If a quarterback did not appear in the contract data, they were placed in the no contract group. I also analyzed the distribution of the 3 contract types (No contract, transferred and retained) within the passing statistics which can be seen below.

Figure 3: Distribution of Contract Types by Passing Statistics



Important trends to notice include the repeating staircase pattern where players who do not receive a contract are outperformed by those who transfer who are further outperformed by those who are retained. This pattern holds for all statistics except for passer rating ("Rate") and completion percentage. In these cases, retained players exceed their peers, but those without a second contract are on par or superior to those who transferred. This abnormality is primarily explainable from the perspective of volume. The other statistics are impacted by the number of opportunities, and players who do not receive a second contract are likely receiving fewer opportunities within their first four seasons, and therefore have lower total volumes. Furthermore, passer rating and completion percentage fail to tell the whole story. Passer rating specifically is a highly imperfect metric that relies on preset weights for different statistics and receives frequent criticism for an inability to disambiguate performances of clear differences in quality. The formula for passer rating along with the included features can be seen below.

$$Passer\ Rating = \left(\frac{\left(\frac{COMP}{ATT} - 0.3\right) * 5 + \left(\frac{YARDS}{ATT} - 3\right) * 0.25 + \left(\frac{TD}{ATT}\right) * 20 + 2.375 - \left(\frac{INT}{ATT} * 25\right)}{6}\right) * 100$$

Furthermore, completion percentage is filtered by organization to some extent. No quarterback completing fewer than 50% of throws would be given a long term role, and defenses are too capable to allow performance over 70% reliably. This leaves completion percentage squarely between 55-70% for all quarterbacks.

Another important trend to identify is the pattern of retained quarterbacks within the touchdown and passing yards features. Each of these boxplots shows that the 75th percentile transferred quarterbacks often perform worse than the 25th percentile retained quarterbacks. This trend is more dramatic than in other features such as interceptions. This phenomena occurs because passing yards will accumulate with opportunities, whereas interceptions result from poor decision making and are less linearly related to volume of attempts. The same is true for touchdowns in that not every completion contributes to total touchdowns whereas they all contribute to yards.

To summarize these boxplots, the retained quarterbacks clearly outplay all others, and there are certain specific metrics that capture group belonging better than others. Specifically, relatively high total touchdowns and yards are related to belonging to the retained group whereas passer rating and completion percentage struggle to differentiate between transferred players and those who do not receive a second contract.

Figure 4: Group and Per Season Stats

	Passing_Yards_Group	Touchdown_Group	Interceptions_Group	Attempts_Group	Completion_Group	Completion_Percentage_Group	Rate_Group
Average	78	87	87	79	78	93	95
Underperforming	33	25	30	31	32	25	29
Elite	33	32	27	34	34	26	20

	Passing_Yards	Touchdowns	Interceptions	Attempts	Completions
Second_Contract					
0	1790.893836	10.613014	8.102740	261.438356	156.126712
1	2408.880282	14.757042	9.193662	344.169014	210.179577

For further analysis, the per-season averages of quarterbacks with or without second contracts were analyzed. The table shows the expected trend that quarterbacks with second contracts tend to have higher per-season averages. Similar data exploration was done for the second research question, where the value of retained and transferred contracts were compared (Figure 5).

3 Models Training & Testing

3.1 Models

For this project, I used supervised classification models to predict whether a quarterback receives a second contract based on their first four seasons in the NFL. The models selected were Logistic Regression, Random Forest Classifier (RFC), and Support Vector Machine (SVM). All methods were implemented using scikit-learn (sklearn) machine learning library in Python. After merging the aggregated passing statistics with contract outcomes, the final modeling dataset was split into 80% training and 20% testing. For logistic regression and SVM, I applied StandardScaler to train and test the data. Additionally, for handling imbalance, I trained the classification models with class_weight = "balanced." Models were evaluated using fixed hyperparameters.

3.1.1 Logistic Regression

Logistic regression was used as it is a simple model to interpret binary classifications, which helped predict contract outcomes from the early career statistics. It provided interpretable coefficients to identify the various passing statistics and to predict earning a second contract. After scaling, the logistic model fit the training data and generated probabilities using the sigmoid function. It is important to note that logistic regression assumes that the log odds of the outcome are a linear function of the input features.

3.1.2 RFC

RFC was selected as it can capture the nonlinear relationships between key predictors in passing statistics. It reduces the risk of overfitting as it builds many decision trees and aggregates predictions. With RFC, feature importance was analyzed, and touchdowns and completions were among the important predictors for second contracts.

3.1.3 SVM

SVM was included as a third model because it is an effective algorithm to identify a decision boundary when the relationships between features are complex and not linear. The RBF kernel was used by default for SVM, which allows for nonlinear decision boundaries.

3.1.4 Evaluation Metrics

To evaluate model performance, I used multiple classification metrics. For each model, I computed the accuracy, precision, recall, F1 score, confusion matrix, ROC curve and AUC. Accuracy reflects the overall percentage of correct predictions, whereas precision evaluates whether the model is reliable with the positive predictions. Recall indicates if the positive cases were identified correctly. The F1 score combines precision and recall into a single metric which reflects the model's performance. The confusion matrix shows the distribution of true positives, false positives, true negatives and false negatives. Lastly, the ROC curve and AUC measured each model's ability to differentiate between quarterbacks who received a second contract and did not. To obtain these evaluation metrics, I applied 5 fold cross validation using KFold with accuracy, precision, recall and F1 score.

3.2 Results

After training the models, performance was evaluated with a 20% test set and 5 fold cross validation with KFold to assess the stability of evaluation metrics. Random Forest Classifier showed the strongest performance. Specifically, RFC had 69% accuracy and an AUC value of 74% (Figure 6). RFC distinguished well between the binary outcomes of receiving a second contract or not. Logistic regression had 55% accuracy whereas SVM was slightly better with 58% accuracy. Both models' AUC values were 65%. RFC's strong performance is based on its proficiency in capturing non linear relationships. While RFC performed best, logistic regression is less computationally expensive. The confusion matrices demonstrate that all models were more accurate when predicting that a quarterback would not receive a second contract. This reflects the imbalance in the dataset. RFC achieved a more even tradeoff between precision and recall compared to the other models, as a result of identifying second contract quarterbacks with more consistency. With the cross validation results, RFC shows stable performance across folds whereas logistic regression and SVM have higher variance. RFC's feature importance shows that passing yards, touchdowns and completions were the strongest predictors (Figure 7). This aligns with the intuition that passing volume and efficiency over a quarterback's first four seasons both reflect and predict future contract decisions.

The first research question asked whether cumulative and or per-season statistics could predict whether a quarterback received a second contract. To answer this question, I split the quarterbacks into stratified performance groups for each feature. At this point, quarterbacks were identified as underperforming, average, or overperforming. The models then used this fact in tandem with the ultimate contract outcome to weigh the predictiveness of each statistic. Reviewing the boxplots based on contract outcome as well as the logistic regression revealed the limited predictive nature of these statistics. The decision to sign a quarterback to a new contract is impacted by factors that outnumber the field based statistics. Furthermore, the NFL is a fast-paced high stakes environment. Quarterbacks who fail early are given up on quickly, quarterbacks who succeed are surrounded with talent and given organizational support immediately. These external pressures forced statistics most directly related to volume of playtime to be the most predictive of whether a quarterback would or would not receive a second contract after four years of statistics. A rookie year can be deceiving or opaque in terms of the career trajectory of a player, but by their sophomore season, teams have already begun long term planning with or without that player. With all this in mind, passing yards was considered the most predictive of group belonging with touchdowns and completions as the two next most predictive. Overall, these results suggest that the use of four years' data is an imperfect lens for predicting contract outcomes as the foundations for these contract decisions are often laid within the first year or two. The future contract outcome casts a shadow on the passing statistics of these quarterbacks long before a pen ever gets the chance to sign paper.

Research question two dealt only with the quarterbacks from the data who received a second contract. For this portion of the project, the question was whether passing statistics could predict the relative contract value given to the quarterback. The dataset for contracts included a normalized value for the total contract which was crucial for comparisons between eras. Even with the normalized value, I further standardized quarterback contracts by segmenting them into five year periods. This was to stabilize comparison between contracts prior to z-scoring the data. Z-scores identified above market contracts as greater than 0.5 standard deviations from the mean and below market as more than 0.5 below the mean. With this step finished, all quarterbacks' second contracts were classified as average, below market, or above market contract.

The final step to answer this research question was to compare the two datasets and identify whether performance within the passing statistics was predictive of relative contract value. I created two sets of stacked bar charts, one displayed the elite performers from the passing features, and the other showed the underperformers (Figures 8 and 9). Within each bar, there were segments representing the relative contract value groups. More than half of quarterbacks classified as elite in passing yards, completions, touchdowns, and passer rating received an above market value contract. Furthermore, underperformers in all features received a below market contract more than half of the time (with the exception of interceptions). These underperformers with second contracts are most likely quarterbacks who are signing as a backup.

As for interceptions, there is an inverse relationship between volume and quality of performance. It is detrimental to throw an interception, and those who throw a high volume are considered underperforming in this data. There are very few "overperformers" within this category. The only "overperformer" received a below market contract likely because they played minimally which prevented them from throwing many interceptions. On the other hand, the "underperforming" quarterbacks who threw many interceptions have an exceeding number of average and above market contracts. This is most likely due to the longer leash given to starting quarterbacks in the league. Players will accumulate interceptions with more games played, some faster than others, so having many may be a sign of being a starter. With this in mind, the atypical presentation makes sense. Overall, to answer my second research question, yes, passing statistics can predict whether quarterbacks who receive a second contract will receive an average, below, or above market contract. Elite performance in touchdowns, yards, passer rating, completion percentage, and total completions all predict an above market contract, and underperformance in those same categories predicts a below market contract.

4. Final Comments

Among the lessons learned completing this project, the first and most glaring was the deficit of publicly available statistics for quarterbacks in a downloadable format. Although there are innumerable websites where you can freely review the statistics of most NFL players, many of those sites offer either very little or altogether no options in terms of downloading as a csv across large samples. Instead, fans are either required to pay to access larger databases or rely on the work of other data science interested fans. In my case, the first database I used was user generated and publicly available via Kaggle. Had it not been for that source, the prospect of making comparisons across twenty-two years of NFL data would have become quite labor intensive. There are paid options available that fit the needs of organizations, but it is clear that there is work to be done in unifying the scattered data into downloadable formats for public use.

Beyond access issues, having two separate sources of data impacted by real world constraints provided another challenge and room for learning. At the outset, I wanted to ensure that there were enough total quarterbacks included in the dataset to make a valid comparison without including so many so as to compare the performance of quarterbacks in the ‘70s to today. Setting the sampling window with respect to the changes in the NFL required filtering out a great number of quarterbacks. Beyond that, the external limitations of cap space, number of teams, etc all shaped the data in ways that I had to be flexible and work with.

The two primary lessons that a reader can draw from this work are the importance of adaptability in the face of real world data, and that the counting statistics generated by quarterbacks are limited in their predictiveness of contract outcome. In terms of adaptability, I had to spend an unusual amount of time in the search for my data set and the process of merging the two sources while maintaining the relationship between my analysis and ground truth. On the predictiveness of these statistics, it seems that the predictive link is strongest in the volume statistics that are most associated with total opportunities as opposed to necessarily good or bad performance. This means that when high in volume, these statistics point to a quarterback with organizational support and therefore a quarterback who is more likely to receive a second contract. Ultimately, there is no single statistic that can be reviewed to determine the likelihood of a quarterback receiving a second contract after their first four seasons, but there are trends that both help to predict who is retained versus transferred, but also who is more likely to receive an above market contract upon signing.

References

(1981, July 21). *N.F.L. Net In 1970 Put At*. New York Times. https://www.nytimes.com/1981/07/21/sports/nf-l-net-in-1970-put-at.html

Associated Press. (1992, July 2). *Discrepancies In N.F.L. Revenue.* New York Times. https://www.nytimes.com/1992/07/02/sports/discrepancies-in-nfl-revenue.html

Peters, A. (2025, April 10). *NFL revenue reportedly passes $23B in latest fiscal year, teams to receive $416M*. Bleacher Report. https://bleacherreport.com/articles/25183332-nfl-revenue-reportedly-passes-23b-latest-fiscal-year-teams-receive-416m

Ozanian, M. (2025, September 4). *Cowboys now worth a record $12.5B - 10 other NFL teams valued at more than $8B*. CNBC. https://www.cnbc.com/2025/09/04/cnbcs-official-nfl-team-valuations-2025.html#:~:text=The%20National%20Football%20League%20continues,not%20authorized%20to%20speak%20publicly

(2023, July 25). *Football by the numbers*. National Football Foundation. https://footballfoundation.org/news/2023/7/25/football-by-the-numbers.aspx#:~:text=1869),Average%20NFL%20Player%27s%20Career%20(Statista)

Durkin, B. (2020). *The importance of the NFL combine in determining draft position and future performance*[Dissertation].

Appendix

Figure 1: Completions vs Touchdowns

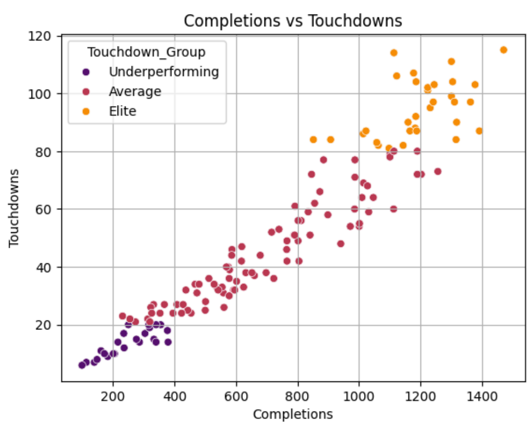


Figure 2: Passing Yards vs Completion Percentage

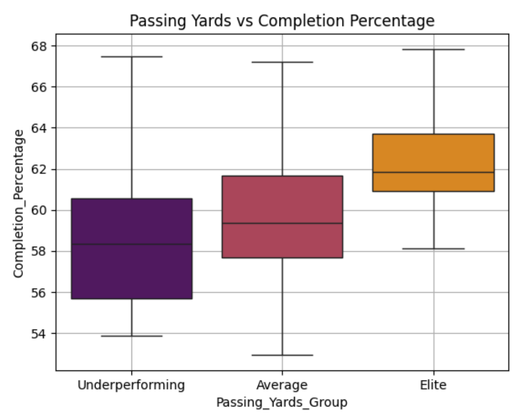


Figure 5: Second Contract Value by Contract Type

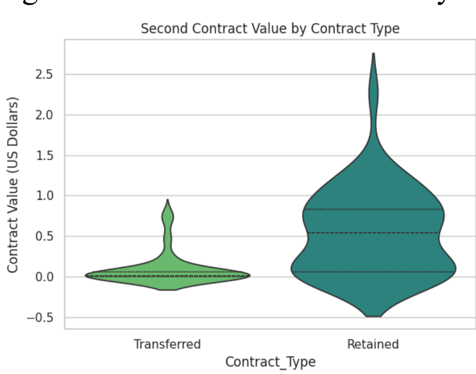


Figure 6: RFC ROC Curve

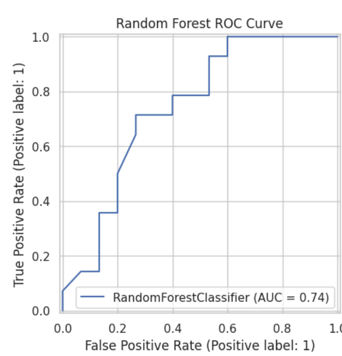


Figure 7: RFC Feature Importance

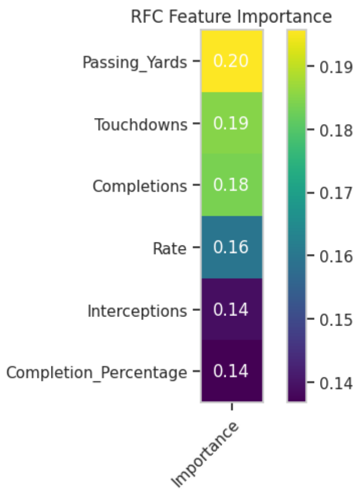


Figure 8: Underperforming QBs by Contract Type

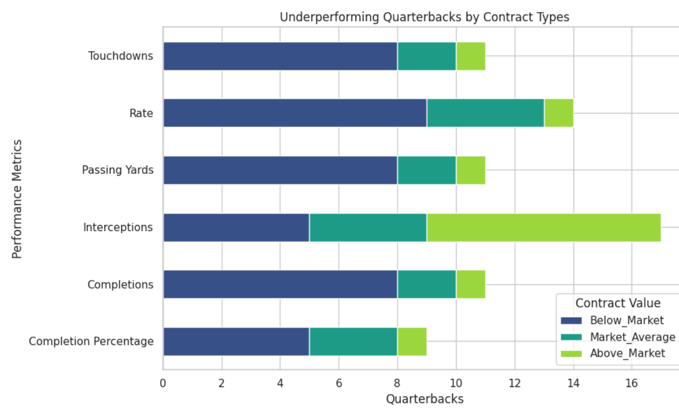


Figure 9: Elite QBs by Contract Type