

October 30th, 2022

From: Neil Sweigard*,* Graduate Student Consultant, Penn State University

To: Howard Roseman, Executive Vice President and General Manager, Philadelphia Eagles

PROJECT TITLE:

# EXECUTIVE SUMMARY

The executive summary should be able to stand alone from the rest of the report. It should include a very brief introduction to the problem, with a concise summary of the results and conclusions.

Perhaps this is what upper-level management would read, only.

# 1.0 - PROJECT DESCRIPTION

NFL analysts expect the Eagles to have a strong roster in the 2022-2023 season, but there are uncertain expectations around how the starting quarterback Jalen Hurts should perform. The Eagles personnel department is carefully evaluating Hurts’ performance this year because his current rookie contract extends through 2023. The department will need to determine if he is a “Franchise Quarterback” that has a future with the team and is worth extending. If he is not the long-term solution, then the team will need to consider replacing him. The team engaged the Consultant for an objective assessment of Hurts’ statistics and to predict his potential as the long-term solution.

The Consultant leveraged the nflfastR R library that offers NFL play-by-play data from the 1999 until current games. The analysis filtered the data to players who have been the primary passer, or the player that threw the most passes for their team within a game, for at least as many games as Hurts. At the time of publishing this analysis, Hurts was the primary passer in 26 games within the data. The project also required NFL draft data from the nflreadr R library to identify quarterbacks who have started their career in the year 1999 or later. Team-based statistics such as win ratio prior to the QB’s first primary start required additional schedule data combined from Pro Football Reference and nflreadr. Finally, the Consultant included contract data from nflreadr to support the research question around Hurts’ potential contract value.

Preparation of the data enabled an observational study of 90 quarterbacks including Hurts. The study considered 16 individual and team-based performance variables, outlined in 1.2 – VARIABLES OF INTEREST. The data preparation also normalized the variables to reflect each quarterback’s performance as of their 26th game and only included the primary passer from each game to provide a fair comparison to Hurts at this point in his career. Several algorithms selected which predictive variables were most significant in predicting three response variables: one categorical variable to label a quarterback as a Franchise QB vs. not, and two continuous variables for contract value.

# 1.1 - RESEARCH QUESTIONS

Question 1: What is the probability that Jalen Hurts will become a Franchise QB?

Question 2: What would the annual contract value be for a quarterback of Jalen Hurts’ ability if they were to extend?

Question 3: How many years would a contract for a quarterback of Jalen Hurts’ ability extend?

# 1.2 – VARIABLES OF INTEREST

Table 1. Predictor Variables

|  |  |  |
| --- | --- | --- |
| **Variable Name** | **Category** | **Description** |
| Days to Hurts Game | Individual Performance | The number of calendar days between the quarterback’s first game as primary passer until their 26th game. |
| EPA per Play | Advanced Individual Performance | The mean expected points added (EPA) per play. QB EPA is the difference in the expected points at the start and end of a play. This does not penalize QBs for mistakes they are not at fault for, such as a receiver fumbling. |
| Fumbles per Attempt | Individual Performance | The number of fumbles the QB was responsible for divided by the number of plays that the quarterback was the passer or rusher. |
| Interceptions per attempt | Individual Performance | The number of interceptions the QB threw divided by the number of plays that the quarterback was the passer. |
| Mean Completion Percentage Over Expected (CPOE) | Advanced Individual Performance | The Actual Completion Percentage minus the Expected Completion Percentage. Only available for the 2006 season and after. Explanation behind Expected Completion Percentage available in Appendix A: Predictors Included in Expected Completion %. |
| Net Point Differential Change | Team Performance | The mean point differential (a team’s final score minus the opposing team’s final score for a given game) for the QB subtracted by the mean point differential for the 26 games prior to the QB’s first game as a primary passer. |
| Net Win Percentage Change | Team Performance | The number of games the quarterback’s team won divided by 26, subtracted by the number of games the team won in the 26 games prior to the QB’s first game as primary passer divided by 26. |
| Passing Completion Percentage | Individual Performance | The number of passes successfully caught by a receiver on the same team divided by the number of passes the QB attempted. |
| Passing Yards per Attempt | Individual Performance | The total passing yards divided by the number of passes the QB attempted. |
| Primary Passing TDs per Game | Individual Performance | The total passing touchdowns divided by the number of games the QB played as primary passer. |
| Primary Rushing TDs per game | Individual Performance | The total rushing touchdowns divided by the number of games the QB played as primary passer. |
| Rushing Yards per Attempt | Individual Performance | The total rushing yards divided by the number of rushes the QB attempted. |
| Sacks per Play | Individual Performance | The number of times the opposing team sacked the QB divided by the number of snaps the QB played. |
| Turnovers per Attempt | Individual Performance | The number of fumbles and interceptions the QB was responsible for divided by the number of snaps the QB played. |

Table 2. Response Variables

|  |  |  |
| --- | --- | --- |
| **Variable Name** | **Category** | **Description** |
| Franchise QB | Categorical | Labeled as 1 if the QB played at or above the median number of games as primary passer for the 90 QBs (71 games). Otherwise, 0. |
| Contract Average per Year (Average per year) | Continuous | The total dollar value of the contract adjusted for inflation in 2022 divided by the number of years the contract extends. |
| Contract Years | Continuous | The number of years the contract extends. |

# 2.0 - EXPLORATORY DATA ANALYSIS (EDA)

Table 3. Numerical Summary of Franchise QB Predictor Variables

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Variable** | **Missing Values** | **Mean** | **Standard Deviation** | **Min** | **Max** | **Range** |
| EPA Per Play | 0 | 0.03 | 0.09 | -0.18 | 0.32 | 0.51 |
| Fumbles Per Attempt | 0 | 0.01 | 0.00 | 0.00 | 0.01 | 0.01 |
| Interceptions Per Attempt | 0 | 0.03 | 0.01 | 0.01 | 0.04 | 0.03 |
| Mean CPOE | **20** | -0.41 | 3.22 | -9.48 | 8.58 | 18.06 |
| Net Point Differential Change | 0 | 52.04 | 144.94 | -386.00 | 365.00 | 751.00 |
| Net Win Percentage Change | 0 | 0.08 | 0.19 | -0.42 | 0.50 | 0.92 |
| Passing Completion Percentage | 0 | 0.60 | 0.03 | 0.53 | 0.67 | 0.14 |
| Passing Yards Per Attempt | 0 | 6.92 | 0.69 | 5.42 | 8.79 | 3.37 |
| Primary Passing Tds Per Game | 0 | 1.31 | 0.36 | 0.65 | 2.69 | 2.04 |
| Primary Rushing Tds Per Game | 0 | 0.15 | 0.15 | 0.00 | 0.73 | 0.73 |
| Rushing Yards Per Attempt | 0 | 3.91 | 1.51 | 1.00 | 6.98 | 5.98 |
| Sacks Per Play | 0 | 0.06 | 0.02 | 0.02 | 0.11 | 0.09 |
| Turnovers Per Attempt | 0 | 0.03 | 0.01 | 0.01 | 0.05 | 0.03 |

The analysis to support Research Question 1 removed 27 quarterbacks who did not play their first season in 2018 or prior because they have not been in the league long enough to play the median 71 games, and therefore should not disqualify as a Franchise QB. Table 3. Numerical Summary of Predictor Variables reflects the numerical summary of the predictor variables. Per the nflfastR documentation, Mean CPOE is only available for the 2006 season and later and causes the 20 missing values for QBs who started their careers earlier in the data. Otherwise, Appendix B: Histograms of Franchise QB Predictor Variables shows most variables have an approximately normal distribution. Skewness exists in Primary Rushing TDs per Game and arguably in Net Point Differential Change, but neither are concerning enough to warrant transformation of variables.

Table 4. Numerical Summary of Contract Predictor Variables

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Variable** | **Missing Values** | **Mean** | **Standard Deviation** | **Min** | **Max** | **Range** |
| Days To Hurts Game | 0 | 860.19 | 547.72 | 407.00 | 2,958.00 | 2,551.00 |

The Contract analyses considered the Days To Hurts Game variable not included in the Franchise QB analysis. The study removed quarterbacks who were missing Contract data and Jalen Hurts because his data should not influence the train or test datasets. Numerical summaries for predictor variables already contained in the Franchise QB analysis but filtered for the Contract analyses are available in Appendix C: Predictor Variables for Contract Analyses. Appendix D: Histograms of QB Contract Predictor Variables shows most variables have an approximately normal distribution. Skewness exists in the Days to Hurts Game variable because some veteran QBs have spent many years as a backup. The analysis will remove QBs who took 2,000 days or later to reach their 26th game as primary passer because we want to narrow the study to QBs who received their first extension past their rookie contract.

# 3.0 –STATISTICAL ANALYSIS

**3.1 Predicting Franchise QBs**

The study first developed a predictive model to determine how likely a quarterback is to become a Franchise Quarterback in their career. The consultant first determined an 80% vs. 20% split of the data into training and test data respectively. The R script randomly assigned quarterbacks into each. The consultant optimized Logistic Regression, Random Forest, and XGBoost algorithms against the same training dataset across the candidate predictive models. Within Logistic Regression, the study compared several variable selection techniques including Best Subsets, LASSO, Ridge Regression, and Elastic Net Regression. Each model then captured their predictions from the test dataset into a confusion matrix that separates true positives, false positives, true negatives, and false negatives. The consultant selected the model that had the largest testing overall accuracy, measured as:

Ultimately, the Logistic Regression with LASSO and Elastic Net variable selections both held the highest testing overall accuracy:

Table 5. Overall Accuracy by Predictive Model in descending order

|  |  |
| --- | --- |
| **Model** | **Overall Accuracy** |
| Logistic Regression with LASSO | 70.6% |
| Logistic Regression with Elastic Net | 70.6% |
| XGBoost | 58.8% |
| Logistic Regression with Ridge Regression | 52.9% |
| Random Forest | 50.0% |

Describe statistical analysis in terms that your client will be able to understand and perhaps use in presenting their own research. This section should summarize all relevant analyses that lead to your final conclusions/decisions/recommendations.

* Model assumptions & how your verified them (details and residual plots can be placed in the Appendix)
* Interpretation of estimates/statistics/p-values IN CONTEXT
* Enough detail that the model/analysis could be reproduced

Do not include code or lengthy output in the main body of your report (as that level of detail belongs in the Appendix).

4.0 – RECOMMENDATIONS

Succinct response to each question laid out in Section 1.2. Provide a succinct (1 to 3 sentence) answer based on your analysis. This a much shorter version of section 3, and focuses on conclusions rather than the analyses.

Question 1:

Question 2:

Etc.

(Follow a parallel order as sections above.)

5.0 - RESOURCES

List resources that your client might find useful. This section may be small or larger depending on the client’s needs.

6.0 - CONSIDERATIONS

For example:

* Limitations to the recommendations
* Concerns you may have about the study
* Technical comments

At the end of the report, you may want to thank your client for the opportunity to work on this project.

Detailed information and a copy of the program used should be referred to in an Appendix. Additional graphs and supporting figures may also be placed in the Appendix.

# Appendix A: Predictors Included in Expected Completion %

As found in <https://www.the33rdteam.com/breakdowns/cpoe-explained/>:

* Field position
* Down
* Yards to go
* Air yards
* Distance to sticks (air yards – yards to go)
* Whether possession team is at home
* Whether the game is played indoors
* Era, broken down into 2006-2013, 2014-2017, 2018 and beyond
* Pass location (binary: middle or not middle)
* Whether quarterback was hit on the play

# Appendix B: Histograms of Franchise QB Predictor Variables

Diagram

Description automatically generated

Figure 1. Histograms of Franchise QB Predictor Variables

# Appendix C: Predictor Variables for Contract Analyses

Table 6. Numerical Summary of Contract Predictor Variables after filtering.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Variable** | **Missing Values** | **Mean** | **Standard Deviation** | **Min** | **Max** | **Range** |
| EPA Per Play | 0 | 0.04 | 0.09 | -0.18 | 0.32 | 0.51 |
| Fumbles Per Attempt | 0 | 0.01 | 0.00 | 0.00 | 0.01 | 0.01 |
| Interceptions Per Attempt | 0 | 0.03 | 0.01 | 0.01 | 0.04 | 0.03 |
| Mean CPOE | 9 | -0.51 | 3.26 | -9.48 | 8.58 | 18.06 |
| Net Point Differential Change | 0 | 69.93 | 142.34 | -300.00 | 365.00 | 665.00 |
| Net Win Percentage Change | 0 | 0.08 | 0.19 | -0.35 | 0.50 | 0.85 |
| Passing Completion Percentage | 0 | 0.60 | 0.03 | 0.53 | 0.67 | 0.14 |
| Passing Yards Per Attempt | 0 | 6.96 | 0.68 | 5.54 | 8.79 | 3.25 |
| Primary Passing Tds Per Game | 0 | 1.33 | 0.35 | 0.65 | 2.69 | 2.04 |
| Primary Rushing Tds Per Game | 0 | 0.15 | 0.15 | 0.00 | 0.73 | 0.73 |
| Rushing Yards Per Attempt | 0 | 3.96 | 1.57 | 1.00 | 6.98 | 5.98 |
| Sacks Per Play | 0 | 0.06 | 0.02 | 0.04 | 0.11 | 0.07 |
| Turnovers Per Attempt | 0 | 0.03 | 0.01 | 0.01 | 0.05 | 0.03 |

# Appendix D: Histograms of QB Contract Predictor Variables

Diagram

Description automatically generated