**An Analysis of Draft Positioning and Career Success for NFL Quarterbacks Based on Pre-Draft Athletic Testing**

**Background to the NFL Draft**

The National Football League (NFL) is the professional American football league in the NFL. It is made of 32 teams who compete annually for the Super Bowl. Each year, in April, the teams take turns to sign the best players from college football via the NFL Draft. Each team takes it in turns to pick a new player, with the order being determined by each team's record in the previous season. Teams pick in reverse order such that the latest Super Bowl winner will pick last, and the team with the worst record will pick first. Once every team has picked a player (or traded their pick to another team) then one round of the draft is complete and the worst team picks again. The draft is complete after seven rounds. After the NFL Draft is complete, teams are able to sign any remaining players they are interested in as undrafted free agents.

Each year, the best players from college football are invited to the NFL's Scouting Combine. Here, they compete in a series of standardised athletic tests. Traditionally, these tests are thought to be informative for NFL teams in judging how well a college football player will perform in the NFL, and therefore whether they would make a good draft pick or not. Different positions in American football require different combinations of traits such as strength and speed, and so often certain tests are emphasised as important for different positions. For the purpose of this analysis, we will focus on the quarterback, i.e., the player who throws the ball. The list of athletic tests in the combine are provided below:

**40-Yard Dash** - Player is timed to complete a sprint of 40 yards (36.576 metres) to assess speed.

**Vertical Jump** - From a standing position, a player crouches down and jumps as high as he can, used to assess explosiveness.

**Bench Rep** - Number of reps completed of a 225lb bench press, used to assess strength and stamina.

**Broad Jump** - From a standing stance, the player explodes forward as far as he can and must land without moving. Used to measure explosiveness.

**3-Cone Drill** - Player is timed to run between 3 cones, changing direction each time, which is used to assess agility.

**20-Yard Shuttle** - Player is timed to run 5-yards, 10-yards, and 5-yards, changing direction each team, used to assess speed and agility.

**Motivating Research Question - Are Teams Drafting Efficiently?**

1) From the features available pre-draft - e.g., results from Scouting Combine - which are most informative in predicting the draft pick in which a player is taken.

2) From the features available pre-draft - e.g., results from NFL Scouting Combine - which are most informative in predicting the success that a player will have in their NFL career? Do these features vary depending upon the American football position played? Career success can be measured in many ways...

**All-Pro Selections** - the consensus best two players in the league for each position, awarded once per year

**Pro-Bowl Selections** - players selected to play in the end of season Pro-Bowl Game, effectively the top 6 players for each position, awarded once per year.

**First-team Selection** - player is considered a first team starter for their NFL franchise, effectively one of the top 32 players for each position, awarded once per year.

We would expect, given the resources available to the scouting departments of teams in the NFL, that they have refined the scouting process such that they have good knowledge of which pre-draft features are most predictive of later on-field success. Therefore, we would expect the feature predictions for draft position and career success to be similar, as better players should be chosen at the start of the draft. If, instead there are differences in the features that predict career success and draft position, it suggests that the drafting process is inefficient. Identifying these market inefficiencies, if they exists, will allow teams in the NFL to draft talent more effectively than their rivals, and so provide a competitive advantage.

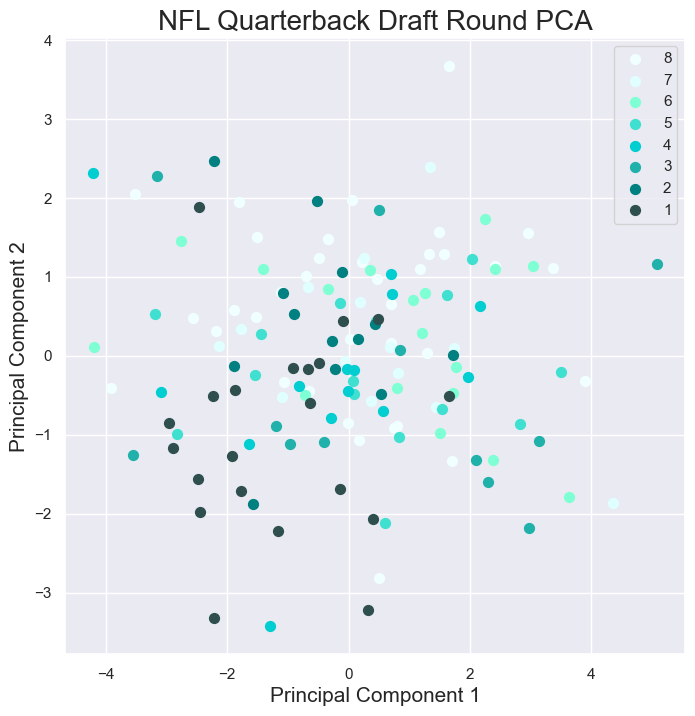
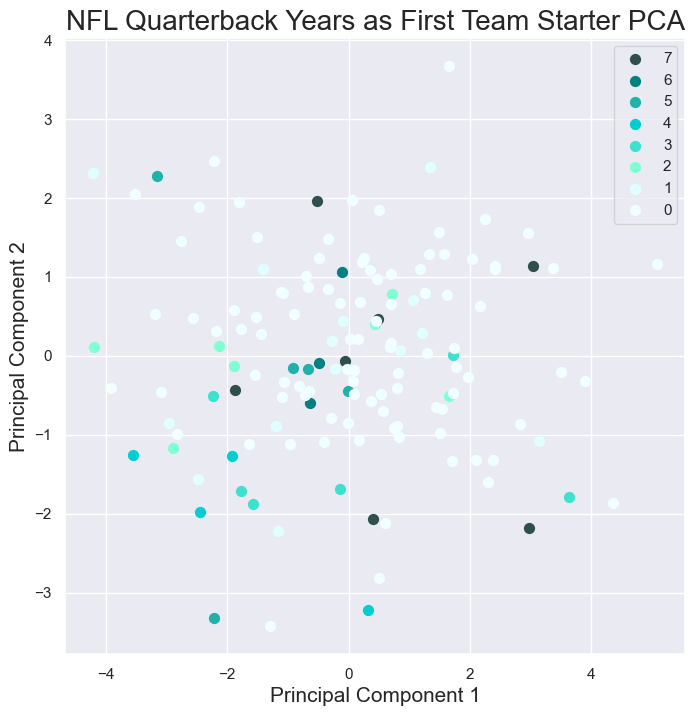
**Methods**

Relevant data on pre-draft athletic testing and post-draft performance in the NFL was downloaded for all players who were drafted between 2000 and 2015 using the online databases from Pro Football Reference (<https://www.pro-football-reference.com/>). This range was chosen to reflect the tactical trends of the modern NFL, i.e., a preference for a high passing offense, and to give sufficient time for the drafted players to become established in the league and eligible for honours such as going to the Pro Bowl. The data was filtered to focus exclusively on quarterbacks, resulting in a final dataset of 206 individuals.

This dataset was analysed using a principal components analysis combined with a K-nearest neighbour cluster analysis, subject to hyper-parameterisation, to understand which of the data available pre-draft are most influential in determining the round in which a quarterback will be drafted in.The same pre-draft dataset was also used to predict measures of success taken from each quarterback’s post-draft NFL career, such as the number of Pro-Bowl and All-Pro selections. This analysis was all completed using Python, with additional use of the pandas, seaborn and sklearn modules. The Jupyter Notebook for this work can be found on GitHub at <https://github.com/KingsleyHunt/SWBio_Project.git>.

**Results and Discussion**

The results from our analysis show that it is difficult to predict what round of the NFL Draft a quarterback will be drafted in, when based on NFL Scouting Combine data alone. Even when compressing the high dimensionality of the scouting data into two principal components, our machine learning model has an accuracy score of only 0.28 (out of a maximum of 1). One possible trend that emerges when exploring the principal components visually is that high round quarterbacks, i.e., the most favoured prospects, tend to cluster together with low principal component 1 scores and low principal component 2 scores (Figure 1). When combined with the loadings of the traits that map onto these components, it suggests that high athleticism (quicker Cone, Forty and Shuttle times combined with higher Vertical and Broad Jump performance) and large size (measured by height and weight) lead to higher draft positioning. This makes intuitive sense in what we understand about the quarterback position whereby athleticism to throw powerful deep balls or make quick running plays, and height to stand tall in the pocket, are all important characteristics.



**Figure 1**: Scatter plots showing how Principal Components 1 and 2 can be used to predict the NFL draft round a quarterback prospect will be taken in (right) and the number of years they will spend as a first year starter (left). The loadings for Principal Component 1 negatively correlate with high speed, agility and explosiveness, collectively termed athleticism. The loadings for Principal Component 2 negatively correlate with height and weight, collectively termed size. From these plots we suggest that more athletic and larger quarterbacks are generally more successful, both initially in the NFL draft, and subsequent in the longevity of their career as a starter in the NFL.

When applying our machine learning model to try and predict the longevity that each quarterback prospect will have as a first team starter in the NFL, we get slightly better performance. In this context, our K-nearest neighbour model receives an accuracy score of 0.68 (also measured out of a maximum score of 1). This is likely driven by the fact that most quarterback prospects do not have a successful career in the NFL, as there are only 32 NFL starting QB vacancies in any one year, and many more prospects than this. Examining the plot, we can see that for most of the parameter space of Principal Component 1 and Principal Component 2, our model predicts 0 years as a starter. The exception, is once again, in the bottom left where the most athletic and tallest prospects are predicted to make it in the NFL. The similarity between the predictions from this model and the model for predicting the NFL draft round suggests that teams do draft efficiently and prioritise the right traits in prospects that are likely to transfer to on-field success.