The Right Quarterback for Your NFL Franchise

Using Clustering to Identify Patterns for the Most Valuable Position in American Football

1. Motivation and Business Need

"Sure, luck means a lot in football. Not having a good quarterback is bad luck." - Don Shula

Why you shouldn't make a thoughtless decision about your quarterback:

1. **High-Stakes Position:**
   Quarterbacks can make or break a franchise - poor decisions on QB contracts or trades could hold back on-field success, break your team, and cost you millions of dollars.

2. **Overwhelming Metrics:**
   Modern football produces a flood of stats: Successful Passes, Touchdowns, Interceptions, and many other that can lead to subjective choices if not analysed systematically.

3. **Data, Data and more Data:**
   We have over 20 years of detailed QB passing records - enough to uncover meaningful patterns and trends.

4. **Clustering Opportunity:**
   By grouping QBs with similar profiles, we reduce guesswork and boost confidence in finding or evaluating the right quarterback for your team.

2. The Dataset

"No data is clean, but most is useful." - Dean Abbott

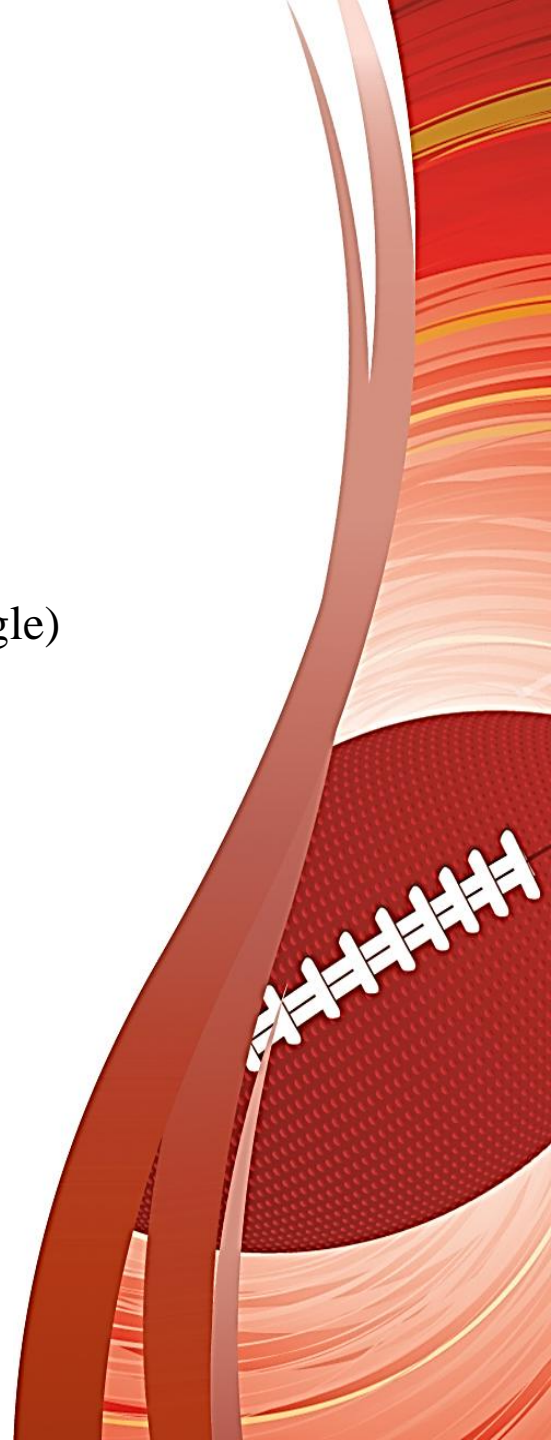**What do we have?**

1. **Over Two Decades of Clean Data:**
   - NFL passing records from 2001–2023, source: Pro Football Reference (obtained via Kaggle)
   - Each entry is a passer with his stats in one specific season/year, so a player can occur multiple times in the dataset when he played more than one season
   - More than 2,000 entries and no missing data, making the analysis reliable

2. **Rich Feature Set:**
   - 27 columns in total, 20 of them are numeric
   - Mainly QB performance metrics - ideal for in-depth cluster analysis

3. **Why NFL Passer Rating and not ESPN's Total Quarterback Ranking?:**
   - The NFL passer rating is officially established by the NFL in 1973
   - Alternative QBR rating was introduced by the sports media channel ESPN in 2011
   - QBR contains additional running stats, but it's unofficial and lacks data for earlier years

3. Preliminary Analysis Plan

"Statistics and numbers are no good unless you have good people to analyse and then interpret their meaning and importance." - Brandon Rodgers

How do I approach it?

1. **Data Preparation & Preprocessing:**
   - Entries that contain a player with less than 224 pass attempts in this year are filtered out
   - Reason: 14 passes per game over a 16-game season are required to occur on the NFL passer leaderboard for that year
   - This also ensures that only quarterbacks are evaluated

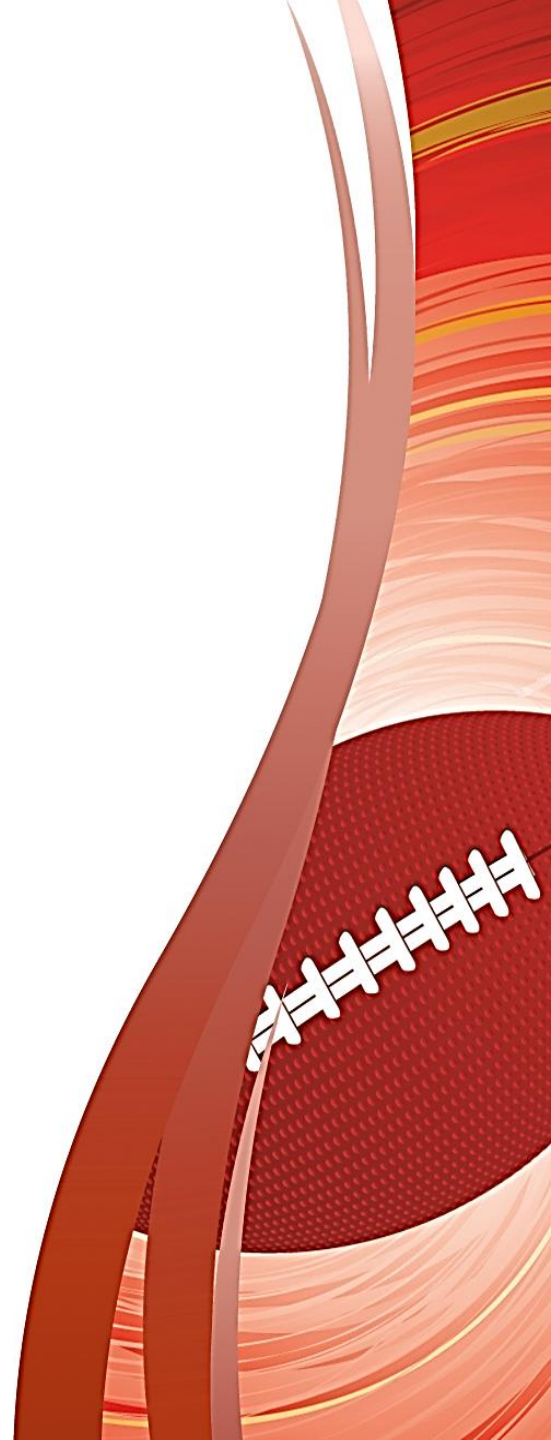2. **A First Look - Statistical Properties:**
   - Using EDA to compute different statistics, distributions, and feature relations

3. **Dimensionality Reduction:**
   - High-dimensional data (27 columns) can be reduced to key principal components
   - Highlights core performance factors, e.g. efficiency vs. volume

4. **Clustering:**
   - Group players with similar traits via methods like k-Means, DBSCAN, and HAC

4. Goals & Value Proposition

"The goal is to transform data into information, and information into insight." - Carly Fiorina

What can you expect?

1. **Identify Distinct QB Archetypes:**
   - Find and Group Quarterbacks by similar styles
   - Comparing your QB's performance with these groups to see in which one he falls

2. **Support Scouting & Contract Decisions:**
   - A data-driven approach to reduce the risk of costly multi-million-dollar contracts on QBs who may not perform as wanted

3. **Enhance Strategic Planning:**
   - Insights into how different archetypes align with your offensive playing style
   - Aiding coaches and general managers in roster building and play design

4. **Future Opportunities:**
   - Integrate additional metrics like ESPN's QBR (from 2011 onward) for deeper analysis
   - Use and extend methodology to other positions