**Data Expectations**

1. You are expected to use **at least three different sources of data.** Ideally, you should look at **current data** (typically data that you scrape or connect to using an API since datasets on Kaggle and GitHub are typically snapshots of old data).
2. The code to extract and merge the data should be hosted on a GitHub Repository. That code should also ensure the merged dataset is **tidy, technically correct, and consistent**. In addition, you should also provide a data validation table for the merged dataset, where you show any observations that violated the data.

**PDF Report**

1. *Introduces the team*
2. *Presents a detailed description of the problem you are trying to solve and any gathered data. This should include the following:*
   1. **What is the problem you are trying to examine? Why should we care about this problem?**

The Wonderlic test was an IQ examination that all NFL hopeful players took at the NFL Combine from 1970 to 2022. To see if the discontinuation of the Wonderlic was justified, our team set out to examine how NFL players’ Wonderlic test scores correlate with performance. We should care about this problem because the Wonderlic had been used to test players' mental ability for 52 years. Wonderlic scores had the potential to help or tank NFL players’ draft stock. Whether the Wonderlic had any impact on how players actually performed in the NFL is questionable. So, it was important to our team to find out for ourselves if players’ Wonderlic scores actually impacted their performance in the NFL.

* 1. **Relationship of your planned project to the current state of knowledge (i.e., what has been done with this data before? What questions/hypotheses are you examining in this area? Are the questions you are asking new?)**

People have often wondered whether or not NFL on-field performance correlated Wonderlic test scores. Studies have been done in the past to try and determine some sort of correlation. One of the more prominent studies was done in 2009 by Dr. Brian Hoffman and Brian D. Lyons in collaboration with California State University (Fresno) and Towson University. In the study, NFL performance was found to have a strong correlation with the Wonderlic test among only two positions: tight ends and defensive backs. Also, these 2 position groups showed negative correlation. There have been many studies completed in many different ways. Another study focussed on only quarterbacks and found that when they removed quarterbacks who never really played in the NFL (1000 passing yards or less), there was a strong positive correlation between NFL performance and Wonderlic scores. There are many different possible routes to take when examining the correlation between Wonderlic scores and performance. Our team has chosen to

<https://bleacherreport.com/articles/2431408-what-does-the-wonderlic-test-really-mean#:~:text=%22The%20issue%20isn't%20whether,Wonderlic%20results%20and%20future%20performance>.

<https://www.playerprofiler.com/article/wonderlic-test-nfl-draft-advanced-stats-metrics-analytics-profile/>

* 1. *To answer the questions you are interested in, you are expected to use at least three different data sources (see Data Expectations).*
     + **What is your approach to getting the data? In addition, you should reference the code link to get the data in your description.**

We are scraping the data from the webpages. Also, tidying them up. Putting them into tables. Importing the csv into tableau. Creating calculated fields. Adding colorblind friendly colors.

* + - **Describe the data that you have obtained. In addition, you should also describe the data quality and present descriptive statistics and visuals (possibly from autoViz in Python or DataExplorer in R) to summarize this data. The links for code to describe the data should also be referenced in this subsection.**
      1. *Wonderlic Scores* - This data provides 406 observations throughout the history of the NFL that gives a glimpse into where certain players and positions Wonderlic scores landed. This allows us to see where position groups ranked on average with their Wonderlic scores. What is flawed about this data is how there are only 406 observations that comprise
      2. *Quarterback Leaders Ranked* - This data, pulled from Pro Football Reference, gives 250 observations about past and current NFL quarterbacks. The data is composed of 7 columns: Rank, Player, Yards, Team(s), Year Start, Year End, and Hall of Famer. These metrics allow us to see the best quarterbacks, their overall yards, and whether they are in the Hall of Fame. When this data was pulled, it needed some cleaning due to multiple metrics sharing the player column. That is where the Hall of Fame column comes from. We use this data to compare Wonderlic scores of different quarterbacks as well as to other positions on the field, seeing if performance is an indicator of Wonderlic scores.
      3. *Fantasy Rankings* - This data pulls the top 262 individuals in football based on their Fantasy Rankings. Fantasy.NFL.com has players separated by position then ranked based on other metrics that are not included. This data gives those final rankings with the player, position, and team for 2023. For cleaning purposes, the news article links were eliminated and the name, position, and team were separated. This table can help us compare different positions and their Wonderlic scores while also analyzing the top versus bottom ranked players.
      4. *Player’s Salary Data* -
      5. *AV Player Scores* -

1. Provides a link to your published dashboard;
   1. if Tableau, you are expected to make it publicly available via Tableau Public (not Tableau Online).
2. A 10-minute in-class presentation during the scheduled time for the final exam.

# 

# Our Team

* ***Jaiden Knecht*** ⇒ Junior Business Analytics | Marketing Co-Major | Working in the Audit Data Analytics Co-Op for Fifth-Third Bank next summer in Cincinnati, OH
* ***Sam Peeler*** ⇒ Junior Business Analytics Major | Currently working as a Business Analytics Intern for the Cincinnati Reds
* ***Samuel W Rogers*** ⇒ Senior Business Analytics Minor | Finance Major | Entrepreneurship Co-Major | Working as a Financial Analyst at Filtration Group in Oak Brook, IL Upon Graduation
* ***Ashley Stratton*** - Junior Business Analytics Major | Atlanta, GA | Mergers and Acquisitions Consulting Intern at West Monroe next summer in New York City

# Introduction

**Problem Background**

The Wonderlic test was an IQ examination that all NFL hopeful players took at the NFL Combine from 1970 to 2022. Originating in 1932 by E.F. Wonderlic, the test measures general cognitive ability for retention, math, vocabulary, and reasoning. The test was adapted by companies in the 1940s to help employee selection, then moved to the US Army and National Football League in the 1970s. Ever since, the Wonderlic test has been a component of the NFL combine, which is the event before the NFL draft that measures potential players’ physical and mental capabilities. The Wonderlic test had been used to examine players' mental ability for 52 years and had the potential to help or tank NFL players’ draft stock.

**Significance of the Problem**

However, the NFL discontinued the Wonderlic score at the combine because of its irrelevance to players real-time performance. Determining if this was the correct decision, our team has set out to examine if and how NFL players’ Wonderlic test scores correlate with their performance in the NFL. This test was important for many years, and we are hoping to understand whether it was ever actually impactful and predictive or not. Because it was important to the NFL for so long, it is important to actually examine if the scores impacted players’ performance in the NFL.

We will be looking at salary, rankings, other scoring metrics, yards, and more to see how they all represent overall performance. This will hopefully help our team understand different measures of performance for players in the NFL, and we will get a better understanding of how Wonderlic fits in as a whole.

**Current State of Knowledge**

People have often wondered whether or not NFL on-field performance correlated Wonderlic test scores. Studies have been done in the past to try and determine some sort of correlation. One of the more prominent studies was done in 2009 by Dr. Brian Hoffman and Brian D. Lyons in collaboration with California State University (Fresno) and Towson University. In the study, NFL performance was found to have a strong correlation with the Wonderlic test among only two positions: tight ends and defensive backs. Also, these 2 position groups showed negative correlation. There have been many studies completed in many different ways. Another study focussed on only quarterbacks and found that when they removed quarterbacks who never really played in the NFL (1000 passing yards or less), there was a strong positive correlation between NFL performance and Wonderlic scores. There are many different possible routes to take when examining the correlation between Wonderlic scores and performance. Our team has chosen to …

# Data Gathering & Quality Assessment

**Sources of Data & Approach to Gathering Data**

Our first step to collecting data was finding credible sources that are highly respected in the sports world. We hoped to find objective data by NFL experts rather than opinions on rankings by sports fans. After determining which sites were credible, such as Pro Football Reference, Fantasy NFL, ESPN, and more, we had to determine which ones were able to scrape into RStudio. We used robots txt on all of our sources, determining which data we could pull effectively. With those results, we started to get a direction for the data. Focusing mainly on player salary, fantasy rankings, and certain position rankings, we tried to get a wide variety of different performance metrics to compare to Wonderlic data.

**Data Quality Assessment**

Once scraped and in R, the next step had to do with cleaning. Pulling the data from a website, the data will obviously not be tidy or cleaned, which was one of our biggest hurdles. Before we could do any analysis, we had to go through and make sure the data was usable. There are many steps to cleaning data, but we first wanted to make sure the data was tidy. This included a lot of strsplit() for different columns. In our data, we had many occurrences where the player, team, and position were all in the same column. It seems simple to fix, but many names are not just two words, which made tidying the data difficult. On top of that, not all the data was formatted the same because of missing values, which complicated the cleaning process.

Also, when pulling from websites, most of the positions were separated by tabs on the same website. This made the scraping process difficult, causing some processes to be repeated many different times before making a full dataframe…

* Merging the data so it is useful

Once our data was tidy, technically correct, and consistent, we then moved on to data analysis and visualizing our data for the audience.

**Descriptive Statistics and Visuals**

# Link to Storyboard

# Data Explanation & Analysis

**Data Sources**

*Wonderlic Scores*

This data provides 406 observations throughout the history of the NFL that gives a glimpse into where certain players and positions Wonderlic scores landed. This allows us to see where position groups ranked on average with their Wonderlic scores. We were able to pull from two different sources of historic Wonderlic data so that we could increase our sample size. What was limiting about our dataset was how many different generations it covered and how few were sampled from today’s athletes. This data was at the center of a lot of our analysis and it was important that we pull from both data sources.

*Quarterback Leaders Ranked*

This data, pulled from Pro Football Reference, gives 250 observations about past and current NFL quarterbacks. The data is composed of 7 columns: Rank, Player, Yards, Team(s), Year Start, Year End, and Hall of Famer. These metrics allow us to see the best quarterbacks, their overall yards, and whether they are in the Hall of Fame. When this data was pulled, it needed some cleaning due to multiple metrics sharing the player column. That is where the Hall of Fame column comes from. Also, for the Rank, that is a metric that was created by Pro Football Reference, which may have some subjectivity to the quarterback rankings. We use this data for many different functions, comparing Wonderlic scores to hall of famers, yards, ranking, and years in the NFL.

*Fantasy Rankings*

This data pulls the top 262 individuals in football based on their Fantasy Rankings. Fantasy.NFL.com has players separated by position then ranked based on other metrics that are not included. This data gives those final rankings with the player, position, and team for 2023. For cleaning purposes, the news article links were eliminated and the name, position, and team were separated. This table can help us compare different positions and their Wonderlic scores while also analyzing the top versus bottom ranked players.

*Player’s Salary Data* -

*AV Player Scores*

This data pulls 22,123 NFL players from ProFootballReference.com based on Approximate Value scores. Approximate Value (AV) scores are a single number that measures the value of any given NFL player’s career.

**Data Analysis**

# Learning Outcomes

# Sources

* [Huge List of NFL Wonderlic Scores by Position](https://wonderlictestpractice.com/nfl-wonderlic-scores/)
* [NFL Passing Yards Career Leaders - Pro Football Reference](https://www.pro-football-reference.com/leaders/pass_yds_career.htm)
* [Player Rankings - NFL Fantasy](https://fantasy.nfl.com/research/rankings?leagueId=0&position=QB&statSeason=2023&statType=seasonStats&week=10)
* [The Story of Wonderlic](https://wonderlic.com/about/history/)
* [What Does the Wonderlic Test Really Mean? - Bleacher Repor](https://bleacherreport.com/articles/2431408-what-does-the-wonderlic-test-really-mean#:~:text=%22The%20issue%20isn%27t%20whether,Wonderlic%20results%20and%20future%20performance)t
* [What You Need To Know About the Wonderlic Test and the NFL Draft](https://www.playerprofiler.com/article/wonderlic-test-nfl-draft-advanced-stats-metrics-analytics-profile/)
* [Wonderlic Score Database](https://footballiqscore.com/wonderlic-score-database)
* [AV Score Description](https://www.pro-football-reference.com/about/approximate_value.htm)